

학술 웹 데이터베이스의 품질 비교 평가 : NDSL, PubMed, Scopus와 Web of Science를 중심으로

A Study on the Quality Evaluation of Scholarly Web Databases Focused on NDSL, PubMed, Scopus, and Web of Science

김 상 준*

Sang-Jun Kim

차 례

- | | |
|-----------------------|-----------------------|
| 1. 서 론 | 5. Web DB 품질의 비교평가 결과 |
| 2. DB 품질평가 선행연구 분석 | 6. 결 론 |
| 3. 비교대상 Web DB 현황 조사 | • 참고문헌 |
| 4. Web DB 품질평가 기준과 방법 | |

초 록

NDSL과 PubMed와 같이 연구현장에서 무료로 서비스되는 DB가 있는 주제분야에 Scopus와 WoS DB와 같은 상용 DB를 도입하기 위한 검토과정에 DB의 품질평가를 위한 연구가 필요하다. 본 연구는 선행연구 분석을 통해 품질평가기준을 추출하여 짧은 시간에 객관적으로 측정하기 쉬운 평가항목을 중심으로 발전적·실무적 비교평가를 실시한 것이다. 그 결과 NDSL은 최신성과 저널 및 주제의 포괄성과 망라성에서 우수하며, PubMed는 최신성과 MeSH 색인에서 전통을 자랑하는 의학 및 생명과학 전문 DB인 반면, Scopus는 사회 및 자연과학 분야에서 현재 1996년 이후의 인용정보를 이용자 중심 인터페이스로 쉽게 이용할 수 있고, WoS는 인용색인을 탄생시킨 전통과 명성을 자랑하지만 구독료가 고가인 관계로 폭넓은 접근이 어렵다. 따라서 인용정보가 포함된 상용 Web DB 도입의 타당성 검토와 이용자교육 및 도서관정보 서비스는 이러한 Web DB의 품질평가 결과를 바탕으로 이루어져야 보다 효율적이고 생산적인 정보 서비스가 가능하다.

키 워 드

데이터베이스, 데이터베이스 품질, 비교평가, 정보 시스템, 품질비교, 품질평가

* 한국생명공학연구원 도서관 사서, 충남대학교 박사과정
(Librarian, Korea Res. Inst. of Biosci. & Biotech Library, Graduate School of Library & Information Science, Chung Nam Univ., sjkim@kribb.re.kr)
• 논문접수일자: 2005년 8월 10일
• 게재확정일자: 2005년 9월 15일

ABSTRACT

This study is focused on the quality of the Web databases which has been produced in science. For the quality evaluation of NDSL, PubMed, Scopus, and WoS, 10 evaluating criteria are developed on the basis of literature review. The evaluation results show that NDSL and PubMed are superior in the currentness and cost. Scopus and WoS are superior in the information of citing and the analysis tool. It is needed for purchasing, user training, and library service based on the above evaluation results.

KEYWORDS

DB, Database, DB Evaluation, DB Quality, Medline, NDSL, PubMed, Quality Evaluation, SCI Expanded, Scopus, Web DB, Web of Science

1. 서 론

1.1 연구의 필요성과 목적

한국데이터베이스진흥센터(2003, 15 28)의 조사에 의하면, 전체 DB 중 학술 DB 이용자는 13.6%이지만 점유율은 1.2%에 불과하다고 한다. 또한 국내에서 기관 종사자들의 상용 온라인 DB 이용은 1998년을 정점으로 감소 추세이며, 2002년에는 57.5%의 이용률과 1주일간 평균 7.9시간에 39.7건을 사용하고 있다고 한다. 따라서 우리나라에서는 DB 이용자들이 상용 온라인 학술 DB의 이용료에 대한 부담을 크게 느끼는 것으로 판단된다. 그러므로 상용 DB의 경우 용도에 적합한 DB를 선정하여 효율적으로 이용해야 할 필요성이 높다.

인터넷과 정보통신기술의 발달로 최근 학술

정보검색 시스템과 DB의 구축 및 정보 서비스가 증가하고 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 정보검색 시스템이나 DB에 대한 올바른 품질평가가 전제되어야만 가능한 일이다. 그러나 정보검색 시스템이나 DB에 대한 정립된 품질평가기준과 방법이 많지 않은 관계로 연구자들은 저마다 다양한 견해와 연구를 시도하고 있는 실정이다.

학술활동에 있어 생산성은 여러 요인에 의해 영향을 받지만, 연구자를 둘러싼 정보 시스템의 품질도 중요한 요인 중의 하나이다. 이러한 정보 시스템의 가장 핵심적인 요소가 DB이며, DB의 품질은 정보 시스템의 유용성에 결정적인 영향을 미친다(이재환 2002, 62). 국내에

서도 '정보화'가 국정의 주요 목표로 등장하면서 다양한 형태의 학술 정보검색 시스템이 구축되고, 이제는 구축된 DB의 효율적인 서비스를 위해 정보검색 시스템 및 DB의 품질평가가 중요해지고, 그에 따른 연구도 늘어나고 있다. 급속히 증가하는 Web DB의 정보수요를 충족시키기 위해 양적인 DB 구축도 중요하지만, 이제는 질적인 DB 구축에도 더 힘을 기울일 필요가 있다.

한국교육학술정보원(2003)의 조사에 의하면, 2003년도 국내 대학도서관의 연간 자료구입비 중 해외 전자정보의 비중이 15.5%이며, 해외 전자정보 중 Web DB(49.1%로 평균 6.5종)와 전자 저널(50.9%로 평균 9종)이 비슷하게 지출되고 있다. 그러나 Web DB로 대표되는 해외 학술 DB 구매의 애로사항 중 가장 큰 애로사항은 해외 학술 DB에 대한 정보부족이 62.8%로 가장 높다(홍현진 외 2005, 92 재인용). 늘어나는 DB 구축과 DB 검색빈도의 증가에 따라 학술 DB 구입의 애로사항을 해결하기 위해서 우수한 DB 선정을 위해 DB 품질평가 모델의 표준정립과 표준 품질평가 모델에 따른 DB 품질평가가 필요한 실정이다. 특히 외국의 고가 상용 DB 도입에는 웹 정보자원의 선정과 평가의 측면에서 심도 깊은 DB의 평가와 검토를 요한다.

그렇지만 국내 현실은 국내 DB 구독대행사의 영세성, 문헌정보학자의 무관심, 도서관 실무자의 연구활동 부족(김상준 2002) 등으로 정확한 DB 품질평가 과정 없이 학술 DB 영업회

사의 권유와 대학교 평가에 따른 경쟁심에 의해 너무도 쉽게 상용 DB가 도입되고 있는 실정이다. 상용 DB 컨소시엄 구독이 늘어나 DB 품질평가의 중요성이 더 높아지는 실정을 감안하면, 정확한 DB 품질평가를 거쳐 올바른 DB 이용자교육의 방향과 DB를 활용한 정보 서비스의 초점을 명확히 할 필요가 있다.

학술 DB에 대한 정보부족의 핵심인 DB의 품질문제는 컴퓨터가 탄생한 이래 줄곧 이어져 온 고질적인 문제로 DB 사용의 보편화·대중화 추세로 더 심각해지면서 일반 이용자들은 불량 데이터의 위험에 속수무책이 되고 있다(이용봉 외 2001, 110). 따라서 정보검색 시스템의 핵심적인 품질이라 할 수 있는 DB의 효과성과 효율성의 측정은 DB, 정보검색 엔진, 이용자 인터페이스, 이용자 만족도, 정보접근·전달 요소 등의 정보검색 시스템 구성요소(유사라 1999, 158) 각각의 측면에서 품질평가가 이루어져야만 한다.

DB가 인터넷의 발달로 대부분 웹으로 제공되면서, 이러한 Web DB의 품질 비교평가를 통해 상용 DB 선정과 도입을 위한 타당성을 쉽게 밝히고, 실무자의 소속 도서관 이용자들에게 적합한 Web DB를 선택하여 최적의 정보를 탐색할 수 있는 토대를 마련하고, 특히 무료 DB와 상용 DB의 수록대상 자료가 많이 겹치는 학술 Web DB의 품질 비교평가를 통해 분석된 문제점에 따른 개선방안을 모색하여 DB의 품질개선과 정보 서비스 활성화에 보탬이 되고자 하는 것이 본 연구의 주요한 목적이다.

그리고 저자의 선행연구(김상준 2004)에서 시도된 2개 의학 분야 Web DB의 품질평가에 관한 연구를 4개 국내외 학술 Web DB로 확대하여 선행연구의 유용성을 보다 정밀하게 살펴보고자 하는 것도 연구목적의 하나이다. 또한 본 연구를 통해 정립된 DB의 품질평가기준과 방법이 KERIS 컨소시엄에서 시도하는 DB 품질평가 외에 국내 도서관계에서 공개적으로 뚜렷한 DB의 품질평가 없이 이루어지는 상용 DB의 영업 및 도입 타당성 검토과정에서 실무적으로 유용한 역할을 할 수 있게 되기를 기대한다.

1.2 연구의 내용과 방법

본 연구는 Web 환경에서 제공되는 대표적인 무료 학술 DB로 2001년부터 시작된 국내의 국가과학기술전자도서관인 NDSL과 외국의 Medline 중심의 웹 서비스인 PubMed는 물론 2002년에 개발이 시작되어 2004년 출시되어 상용(유료)으로 이용되는 Scopus와 40년 이상의 전통을 자랑하는 Web of Science(본 논문에서는 인문·예술 분야 AHCI와 사회과학 분야 SSCI 및 과학기술 분야의 SCIE를 중심으로 구성된 Web of Science를 WoS로 표기하고, 과학기술 분야만 대상으로 할 때는 SCI Expanded로 표기) DB의 품질 비교평가를 주요한 내용으로 한다.

본 연구에서는 이러한 국내의 환경과 현실을 감안하여 DB 품질평가에 관한 국내외 선행 연구를 종합적으로 조사·분석하여 검토한 뒤

실무자 중심의 품질평가 표준 모델로 사용할만한 DB 품질평가기준을 설정한 저자의 선행연구(김상준 2004)를 기반으로 DB의 품질평가를 시행하고자 한다. 즉, 선행연구의 DB 품질평가기준에 따라 국내 학술분야 연구현장에 서지정보 Web DB 형태로 널리 활용되고 있는 무료 NDSL과 PubMed는 물론, 상용 Scopus와 WoS DB의 품질평가를 비교하여 무료 DB가 이용되는 상황에서 상용 DB 도입의 타당성 검토를 지원할 DB 품질평가기준의 타당성을 확인하고자 한다.

무료 및 상용 학술 Web DB 품질의 비교평가를 위한 본 연구의 연구방법은 다음과 같이 순차적으로 실시되었다.

첫째, 정보 시스템 및 DB 품질평가에 관한 국내외 선행연구와 비교평가 대상 DB의 품질 비교연구에 대한 문헌분석을 간략히 실시한다.

둘째, 저자가 수행한 선행연구의 DB 품질평가에 관한 기준에 따라 객관적으로 조사가 가능한 평가요소를 위주로 무료 및 상용 학술 Web DB의 양적 및 실무적 품질평가를 실시한다.

셋째, 품질평가를 수행하여 밝혀진 사실을 토대로 상용 DB의 도입과 이용자교육 및 도서관정보 서비스를 효율적으로 하기 위한 제안을 결론에서 제시한다.

본 연구는 상용 Web DB의 품질평가에 대한 연구가 척박한 국내 현실에서 DB 수록대상 저널의 상당수가 중복되는 상용 DB와 무료 DB의 품질평가기준을 확보하고, 품질을 비교평가하여 상용 DB 도입 타당성의 근거를 확보하고,

평가결과에 따라 DB가 도입된 이후 DB 검색 서비스와 이용자교육 및 DB의 활발한 활용을 위한 방안모색에 실무적 뒷받침을 할 수 있다는데 그 가치가 있다고 하겠다. 그러나 본 연구에서는 상용 DB의 도입을 위한 검토를 염두에 두고 품질평가를 실무적으로 하는 상황을 설정한 관례로 조사시점에 따른 결과의 상이성을 확인하지 못하였고, DB 검색 후 이용자의 검색결과와 만족도 등에 의한 보다 객관적인 평가결과가 반영되지 못한 한계점이 있다.

2. DB 품질평가 선행연구 분석

DB의 품질평가와 관련된 국내외 선행연구는 평가의 목적과 대상 및 주체에 따라 다양한 평가기준과 절차 및 방법이 개발되어 적용되고 있다. 그렇지만, 크게 보아 불특정 다수 DB를 대상으로 객관적 평가기준과 계량적 측정방법을 이용한 '범용 모델'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 '범용 모델'의 개발에 가장 중요한 것은 보편성 있는 평가기준과 합리적인 측정방법, 그리고 객관적인 등급결정 매커니즘이다. 또한 DB의 개발자와 사용자 사이의 품질에 대한 인식의 차이와 함께 모든 사용자가 동일한 관점에서 DB의 품질을 판단하는 것이 아니라는 사실이 '범용 모델'의 개발을 어렵게 만드는 요인이다.

급증하는 상업용 DB의 경우, 사용자의 선택을 돕기 위해 객관적인 품질을 비교평가하기 위한 '범용 모델'의 개발이 이상적일 지 모르지만, 특정기관에 의해 제작된 DB의 정기적인 품질검증을 위해서는 해당 DB의 구축목적과 기능을 고려한 '사례별(혹은 특화된) 모델'이 보다 바람직할 지 모른다(조순영 외 2002, 31-32).

아래에서는 DB의 품질평가와 관련된 국내외 선행연구를 외국과 국내연구로 나누어 발전과정과 동향을 간략히 살펴보고(국내·외 선행연구에 대한 자세한 내용은 김상준 2004 참조), 연구대상 DB의 품질 비교연구에 대해서도 자세히 알아보려고 한다.

2.1 외국의 DB 품질평가 선행연구

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

텍스트 DB의 품질에 관한 연구는 초기의 인쇄본 2차자료를 대상으로 시작되었다. 60년대에는 기계가독형 DB 출현으로 정확성을 평가하는 연구가 발생하였다. 70년대에는 DB의 비교평가가 활발하였다.

또한 80년대에는 CD-ROM DB의 평가가 나타남과 동시에 DB에 대한 대규모 평가기준

이 나타났다. 그리고 90년대에는 인터넷을 활용한 탐색이 증가되면서 다양한 평가 모델이 시도되었다. 2000년대에는 웹 자원과 디지털 도서관의 사용성 평가와 인터페이스 중심의 평가가 시도되어 정보 시스템 및 DB 품질평가가 치밀해지고 있다.

2.2 국내의 DB 품질평가 선행연구

국내에서 DB 품질평가에 관한 연구는 한국 데이터베이스진흥센터에서 DB 품질평가에 관한 연구(이국희 등 1995)가 문헌분석 결과 처음 시작된 것으로 판단된다. 그 후 이응봉, 이제환, 김선형, 이명희, 유사라를 중심으로 여러 가지 DB 품질평가에 관한 연구가 이어져 오다가 17개 국내외 선행연구의 비교분석을 통해 DB의 품질평가를 위한 범용 모델의 기준을 마련한 연구(안계성, 함정훈 2000)가 있었다. 또한 국내외 5개 주요 선행연구를 분석하여 공통적인 평가지표 간의 중요성에 따른 가중치를 부여하여 지표별 평가척도를 만든 연구도 있었다(장혜란 2000).

이어서 이응봉, 김석영, 문성빈 등의 연구를 거쳐 KERIS의 해외 학술 DB 선정 및 평가 모델 개발에 국내외 17개 선행연구의 품질평가기준을 분석하고 AHP 기법을 활용하여 평가요소에 대한 중요도 및 가중치를 측정하는 프로그램의 개발(조순영 외 2002)에 관한 연구가 이어져 왔다. KERIS의 평가지표를 프로그램을 이용해 전문가와 이용자들이 같이 평가한다

면 이제까지의 다른 방법보다 더 객관적인 평가와 순위화가 가능할 것이다.

그 후 김성희, 이만수, 이춘열, 한국데이터베이스진흥센터 및 저자의 연구에 이어 가장 최신연구는 웹 기반 DB를 위해 데이터, 서비스 및 효과성이라는 3개 영역에 19개 평가지표와 45개 평가요소로 구성하여 전문가와 이용자를 대상으로 설문조사한 연구가 있다(홍현진 2005).

그 동안 국내의 DB 품질평가 연구는 KORDIC(현재 KISTI), KAIST, KINITI(현재 KISTI), KERIS, 국립중앙도서관과 국회도서관 등 공공기관에서 구축한 DB의 평가와 비교에 치중되었다. 따라서 DB의 도입을 위한 상용 DB의 품질평가에 대한 연구도 거의 없는 실정에서 최근 Web 환경에서 DB 자체의 품질과 함께 DB 운영을 위한 인터페이스 측면에 대한 평가는 물론, 실무적 품질평가에 대한 연구가 등장하고 있다.

2.3 비교대상 DB의 품질평가 선행연구

본 연구에서 품질 비교평가 대상으로 삼은 NDSL, PubMed, Scopus 및 WoS에 관한 선행연구를 살펴보면, 최근에 인터넷의 발달과 정보의 디지털화에 따라 구축을 시작한 국내의 NDSL에 관련된 선행연구는 이용자 정보요구와 이용 행태 분석에 대한 연구(유사라 2002)만 있을 뿐 아직 찾아보기가 힘들다. 그러나 가장 최근에 서비스가 시작된 Scopus에 대한 연

구는 관심이 증가되면서 조금씩 나타나고 있다. 반면 Web 형태로 제공되기 전부터 PubMed와 WoS의 전신으로 볼 수 있는 초록 색인지나 CD ROM 및 통신망 형태의 Medline과 SCI에 대한 선행연구는 외국에서 다수의 연구가 있었다.

외국의 많은 연구자와 도서관 사서들이 Medline DB의 특성에 대해 Dialog 온라인 서비스 등을 활용해 비교평가를 시도하였는데, 주로 Medline과 Embase DB의 중복률과 포괄성 및 두 DB가 검색하기에 더 우수한 주제분야에 관한 사례연구가 많다(자세한 내용은 김상준 2004 및 유동숙 1998 참조). 그러나 출판사들이 데이터를 직접 제공한 PubMed 데이터 중 미국 국립 의학도서관인 NLM의 정책에 따라 선별된 저널의 논문만 Medline에 수록되므로 종전의 Medline과 무료로 Web에서 제공되는 PubMed는 데이터 선정 정책과 데이터의 양에서도 차이가 나고 있어 새로운 평가가 요구된다.

국내 의학 분야 연구업적 평가에 이용가능한 DB의 포괄성에 대해 우리나라 연구자가 쓴 1990년 1995년 사이 국내외 저널의 논문을 검색하여 분석한 결과, 의학 분야 연구업적 평가를 위해 1종의 DB를 사용한다면 SciSearch (SCI Expanded와 동일)가 적합하며, 2종을 사용한다면 SciSearch와 이를 보완 할 Embase가 Medline보다 적합하다는 연구가 있다(이춘실, 이은숙 1998).

최근에 출시된 Scopus와 오랜 전통의 WoS를 중심으로 수록 저널의 질적인 평가연구에

의하면(유재복 2004), 수록 저널의 범위가 넓은 Scopus에서 WoS 수록 저널과 중복된 저널을 제외한 Scopus에만 신규 추가된 저널의 경우 원자력과 문헌정보학 분야에서는 질적인 수준이 크게 떨어지지 않는 것으로 나타났다. 따라서 다른 조건이 동일하다면 Scopus가 논문을 위한 DB의 망라적 검색에 WoS보다 더 유용하리라 추정된다.

또한 과학기술 분야 중심의 주제검색에 강점을 가지고 웹 정보와 특허정보를 함께 검색할 수 있는 Scopus와 인문·예술, 사회 및 자연과학 등의 종합정보를 보유하면서 저자명 검색에서 월등한 WoS를 비교평가하면서 만족도를 최대한으로 높이려면 두 DB를 모두 구독하여 서비스해야 한다는 주장도 있다(LaGuardia 2005).

3. 비교대상 Web DB 현황조사

본 연구에서 비교대상으로 정한 4개 Web DB의 주요한 특성을 조사하여 비교하면 다음 <표 1>과 같이 정리해볼 수 있다. 또한 PubMed와 Medline과 다른 비교대상 DB의 관계를 그림으로 나타내면 <그림 1>과 같다.

아래에서는 각 DB의 현황을 주요 내용 중심으로 간략히 살펴본다.

3.1 NDSL

2001년부터 웹으로 서비스가 시작된 NDSL

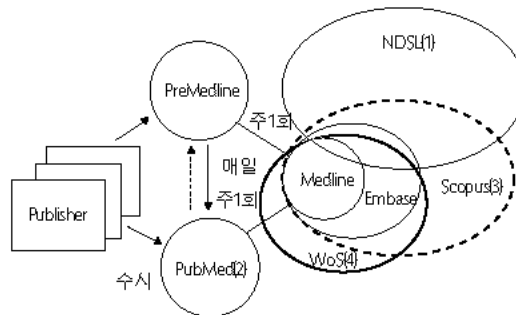
〈표 1〉 비교대상 학술 Web DB의 특성

특성	NDSL	PubMed	Scopus	WoS
사이트	http://ndsl.or.kr/	http://www.ncbi.nlm.nih.gov/entrez/	http://www.scopus.com/	http://isi3.isiknowledge.com/
제작 기관	KAIST	NLM	Elsevier B. V.	Thomson ISI
구독료	무료	무료	전체를 FTE별 3년에 3.8만-10만\$ (추가 할인 가능)	1년에 최소규모의 협의 가격 및 약 3.5만-5만\$
인쇄 색인지	없음(처음부터 디지털화)	Index Medicus, Index to Dental Literature, Internat. Nursing Index	없음(기존 Embase, Compendex, Biobase 등의 콘텐츠를 망라)	SCI(1945-), SSCI(1956-), AHCI(1975-)
검색 엔진	Varity	Entriz	FAST	
갱신 빈도	주 1회	매일	매일	주 1회
자료의 중심 수록 기간	영어자료 보유율은 83.7%, 66년 이전 자료도 보유, 00년 이후 자료의 비중은 40.9%이고, 85년 이후 자료의 비중은 89.3%로 최신 자료 중심	1966년 - 현재 (1953 - 1965 OldMedline 데이터 포함)	1966년 - 현재 (1966년부터 Medline 포함, Medline 유일 1,250여 종은 인용정보 수록하지 않음, 1996이후 인용정보는 STM분야에서 10년분이 85% 인용되기 때문)	1900년 - 현재(Century of Science 출시로 1900년 까지 소급, SCIE, SSCI, AHCI, Index Chem.(1993-), Current Chem. React.(1986-), INPI arch.(1840-1985))
수록 레코드 수	2,291만여건(프로시딩 포함 2,872만여건)	1,551만여건 (52개 자료형태 수록)	2,726만여건 (프로시딩 포함)	3,284만여건
색인 연속 간행물 수	4만7,558종(프로시딩 17만여종은 제외)	8,116종(Medline만 70개국 4,800여종에 북미(47%)와 유럽(43%) 정보에 편중)	1만4,126종(계재심사 거친 저널을 선정기준으로 하며, 미국 이외가 60% 차지)	8,859종(SCI Expanded 6,348종 포함 3개 세부 분야의 중복 제외)
초록 보유	논문 중 29.7%만 보유	1985 이후 69% 이상 보유	1985 이후 80% 저자 초록	1991 이후 초록 보유
수록 및 색인	360개 기관의 소장정보 수록	출판사 통보 후 목차는 5일내(색인은 6개월내) 수록	인쇄저널 도착후 2주내 색인(최대 5주내 수록), 81%가 통제어로 색인	SCI Expanded는 1945년 이후 인용문헌 정보수록
원문 Link	18,870종(KESLI 컨소시엄사 Elsevier의 다수)	KESLI 컨소시엄사 Elsevier 등 Medline만 4,714종 저널 1,730개 기관이 소장정보 표시 위해 LinkOut (219 출판처 4,714 저널) 사용 중	CrossRef 출판사 7,528종 저널 + OA 저널 513종(Open URL과 호환, Link resolver 가능한 이미지 기반 링크의 소장정보 표시(LinkFinder Plus와 SFX) 가능)	알 수 없음(다만 KISTI 컨소시엄 구독 가능하고, Back-up CD 제공 및 Agricola/Genebank/PubMed 통합검색)

<p>색인대상 주제</p>	<p>인문·예술(14%), 사회(28%), 자연(58%) 과학의 종합 분야</p>	<p>의학 및 생명과학 분야 (A I D S L I N E , H e a l t h S T A R , P O P L I N E , T o x l i n e , 간호학, 치의학, 수의학, 약학)</p>	<p>생명과학은 물론 사회 및 자연과학 분야(인용 문헌 정보 포함)</p>	<p>인문·예술, 사회, 자연 과학의 종합 분야(인용 문헌 정보 포함)</p>
<p>저널 (연속간행물) 선정</p>	<p>KESLI 컨소시엄 및 NDSL 운영 센터에서 선정하되 운영위원회 등의 자문을 반영</p>	<p>LSTRC(Literature Selection Technical Review Committee) 위원회가 논문의 목적, 신뢰성, 내용과 출판형태의 질적 수준까지 평가하고 관리</p>	<p>대부분 peer review 과정이 있는 연속간행물을 편집회의에서 심사하여 채택(conference)의 초록을 제외하고는 저널에 수록된 모든 논문을 수록하는 원칙</p>	<p>수록대상 저널 이슈의 모든 논문을 수록하는 원칙</p>
<p>서비스 플랫폼</p>	<p>NDSL에서 서비스 하고 있지만 NOS(NDSL on Site)로 개별 도서관 OPAC을 통한 NDSL 서비스)에서도 서비스 계획 중</p>	<p>BMN(BioMed Net), EP(Biosis Previews), Ebsco, Embase, HighWire, IdDB3, IGM, OCLC First Search, OVID, PML, SciFinder, WoS, Dialog 등에서도 Medline 포함</p>	<p>특허(US, UK, EPO, WIPO)와 웹(Scirus) 정보의 동시검색 가능(GEOBASE Subject Index, EMTREE, MeSH, FLX terms, WTA terms, Regional Index, Species Index, Ei thesaurus 이용해 색인)</p>	<p>ISI Web of Knowledge(BP/CAB Abst./Current Contents Connect/Derwent Innov. Index/ESI/ Inspec/ISI Chem./ISI Proceedings/JCR Web/Web of Nanotech. 등과 동시검색 가능)</p>
<p>기타 특징</p>	<p>인용문헌 검색 불가능, 단기간에 신속하게 구축, Title과 Abstract를 각각 검색 후 검색식 수정을 통해 OR 검색 가능, 저자주소 검색 등 검색필드의 다양성 제한, 초록 보유율이 낮아 키워드 검색 성공률 낮음.</p>	<p>인용문헌 검색 불가능, Single/Batch Citation Matcher, Tutorial, Clinical Query, Journal DB 운영, LinkOut은 현재 179개 출판사 저널 연결 표시 가능, PubMedCentral 등 무료데이터 수록, MeSH Term(2.2만 개념과 12.5만 참조어), Cubby의 Alert</p>	<p>검색결과 분석기능, Boolean, wildcard/truncation, field delimitation, subject/document limits Quick/Advanced Search Profile에 의한 Alert, 85% 색인을 위해 통제어휘(Stem) 사용, Refine 제외한검색이력 50개</p>	<p>검색결과 분석기능, 화학구조식과 cited reference searching 가능, EndNote/Reference Manager/ProCite로 저장, Set up citation alerts 선택 기능, term lists(dictionary), ISI Document Solution, Keyword Plus 검색</p>

은 KAIST에서 추진하는 KESLI 전자 저널 컨소시엄을 기반으로 컨소시엄 참여 출판사의 메타데이터 협조와 영국 국립도서관인 BL의 데

이터를 통합하여 Scopus처럼 짧은 시간에 대규모 서지도구로 성장한 e Gate 형 DB이다. NDSL은 향후에 National Digital Archive



〈그림 1〉 PubMed와 다른 DB의 상호관계

구축, NDSL 웹 서비스 구축 및 보급, Wireless Digital Library 구축을 목표로 운영되고 있으며, 운영 담당 기관이 KAIST에서 KISTI로 변경될 계획이다.

NDSL 운영 센터의 조사(2005. 4. 20. 기준)에 의하면, NDSL의 2005년도 35만5,070건의 레코드 중 논문 32만1,803건은 1993년도 이후부터 수록된 영국 국립도서관인 BL의 목차 데이터가 16만6,397건(51.7%)이고, Elsevier 출판사를 비롯해 13개 KESLI 컨소시엄 참여 출판사의 메타데이터가 15만5,406건(48.3%)의 비중을 차지하고 있다. 출판사 메타데이터 15만5,406건의 47.2%인 7만3,368건을 Elsevier 출판사가 차지해 Elsevier 출판사 데이터의 비중이 다수를 차지하고 있다.

그리고 2004년 12월말 기준 NDSL의 목차 정보(TOC) 수록 저널은 총 4만1,199종이며, 이 중 영국 국립도서관인 BL의 데이터를 수록한 저널이 2만8,131종(68.3%)이고, 나머지 메타데이터 제공 13개 출판사가 총 1만3,068종

(31.7%)을 차지하고 있다. NDSL에는 수록된 4만7,558종의 저널 중 전자원문 링크가 있는 저널이 1만8,870종으로 수록 저널 전체에서 차지하는 비중은 39.7%의 비율로 낮지만 수록 저널 수로만 보면 4개 DB 중 가장 다양하고 포괄적이라 할 수 있다.

3.2 PubMed

PubMed는 미국 국립보건원(NIH : National Institutes of Health) 산하의 NLM(National Library of Medicine)에서 개발되어 NCBI(National Center for Biotechnology Information)의 Entrez 검색 시스템에서 운영되는 가장 대표적인 의학 및 생명과학 분야 서지 DB이다.

1997년 6월부터 Medline을 무료 개방하여 웹 서비스로 제공되는 PubMed에는 Medline 데이터뿐만 아니라 색인하기로 결정하기 전의 저널은 물론, 일부 일반과학 저널과 NLM에 의해 질적인 평가를 거친 PMC(PubMed

Central) 데이터도 수록되고 있다. 이 때문에 Medline에 수록되기 이전의 데이터도 출판사의 적극적인 협조로 PubMed를 통해서도 검색이 가능하다는 중요한 강점을 가진다.

국내 의과대학 도서관, 병원 도서관, 의학 학회 등 146개 의학 관련 기관의 웹사이트를 분석하면 PubMed는 71개 기관(48.6%)에서 소개되어 있고, 미국과 스페인과 일본 및 독일의 99개 의학 관련 기관 중 57개(57.6%) 기관이 PubMed의 링크를 제공하고 있다. PubMed는 국내외 총 245개 웹사이트 중 128개 기관(52.2%)이 링크하고 있으며, 이를 의과대학 도서관으로 한정시키면 평균 82% 이상의 사용가능성을 보이고 있었다(김기오 2003, 32, 66).

외국의 연구에 의하면(김기홍 2000, 5 6 재인용) Medline 이용자의 사용빈도는 64%로 만족도는 66%와 69%로 보고되었다. 또한 캘리포니아대학교(UCB) Melvyl 웹 목록의 이용을 분석한 결과 16개월 동안 740만 건의 DB 선택 중 Medlars가 22%로 32%를 차지한 목록 DB에 이어 2위를 차지하며, 잡지와 정기간행물(18%)은 물론, 다른 과학기술 분야 DB 보다도 Medlars가 월등히 높은 이용률을 보였다(Cooper 2001). 그러나 Medline은 북미지역에서 가장 널리 이용되지만, 유럽에서는 Embase가 자주 사용된다고도 한다(Suarez Almazor et al. 2000, 47). 따라서 Medline의 Web 버전이라고 할 수 있는 PubMed는 전세계적으로 매월 6,400만 건 정도의 검색이 발생

하여 의학 분야의 가장 기초적이면서 대표적인 문헌정보 DB이자 Web DB라고 할 수 있다.

3.3 Scopus

Elsevier 출판사에 의해 개발된 Scopus는 2002년부터 21개 연구기관과 300여명의 과학자들의 의견을 반영하여, 이용자 연구결과를 적극 반영하면서 개발되기 시작했다. 그리하여 2004년 말에 본격적으로 출시된 Scopus는 논문의 인용정보를 수록한 제품으로 45년 이상의 전통을 자랑하는 WoS와 함께 쌍벽을 이루는 사회 및 자연과학 분야의 종합 인용정보 수록 서지 DB로 단시간에 성장하고 있다.

Scopus는 종전의 Embase, Compendex 및 Biobase 등의 Elsevier 서지 DB와 Medline DB를 기반으로 짧은 시간에 대규모 DB로 성장하면서 WoS의 아성에 도전하고 경쟁하는 인용정보 수록 DB 도구로 출시 6개월 만에 500개 구독기관을 확보했다. Scopus는 또한 주제분류가 포괄적이어서 의미가 감소하기는 하지만 DB내 하위 주제분류를 시도하고 있고, 재현율을 정도율에 우선하는 정책과 함께 논문게재 심사를 거치고 영어초록만 보유하고 있다면 비영어권의 다양한 저널도 포괄적으로 수록하려는 정책을 펴고 있다.

NDSL 운영 센터에서 수행한 NDSL과 Scopus 수록 저널 비교조사에 의하면, 조사 당시 수록 저널 4만7,558종의 NDSL과 1만4,013종의 Scopus에서 ISSN이 있어서 ISSN으로

비교한 1만1,740종과 ISSN이 없어서 저널 명으로 비교한 2,273종 등 Scopus 기준 1만 4,013종에서 두 DB에 중복되는 저널은 1만 2,198종이다. 그리고 Scopus에만 수록된 저널은 1,815종(전체의 13%)으로 나타나 저널에 대한 실사를 거치지 않은 점을 고려하더라도 Scopus 기준으로 NDSL과의 저널 중복률이 최소한 87%로 나타나고 있다.

한편 Scopus는 과학기술 선진국이 아닌 국가에서 발행되는 학술 저널을 상대적으로 많이 포함하고 있는데, 인도의 경우 2000년도 기준 자국 저널이 SCI에 10종과 SCI Expanded에 51종에 그쳤지만, Scopus에서는 2004년도 기준 1만3,890종 중 자국 저널이 가장 많은 168종을 수록하고 있다는 보고가 있다(Jain 2005).

3.4 WoS

Eugene Garfield가 1955년에 인용색인 개념을 발표한 후 Science Citation Index의 인용색인 기법이 1961년부터 인쇄물로 생산되다가 웹으로 서비스를 하면서 WoS로 불리어지고 있다. WoS는 인용색인을 정보검색 분야에 도입함으로써 정보검색의 다양성은 물론 참고 문헌이 동일한 유사 연구결과를 손쉽게 찾아주어진 연구결과에 전후 상황을 찾아보기 편리하여 정보학의 지평을 넓힌 독점적 정보기술을 제공하였다. 그러나 40여년의 전통을 가진 WoS만이 아니라 Scopus DB에도 인용정보가

수록되면서 두 DB 사이에 경쟁현상을 불러와 WoS 측의 DB 기능개선을 위한 대응속도도 최근에는 빨라지고 있다.

특히 한국에서는 SCI 또는 SCI Expanded 수록 한국인 논문을 매년 과학기술부나 교육인적자원부에서 조사하여 발표해오고 있다. 이와 같이 국내에서 상용 SCI 또는 SCI Expanded DB에 깊이 의존하여 과학기술계에서 중요시하는 이유는 연구성과에 대한 객관적 평가기준 제정이 쉽지 않고, 그러한 기준에 대한 합의가 매우 어려운 현실 때문이다. 연구성과 평가기준의 객관성에 지나치게 함몰되어 많은 연구기관에서 SCI 또는 SCI Expanded에 지나치게 의존하는 것은 문제가 있다는 일부의 지적이 있기는 하지만, 아직도 그 틀은 깨어지지 않고 있다.

따라서 최근 Scopus의 등장은 DB의 다양성과 함께 WoS와의 경쟁을 불러와 국내의 종합서지 DB에 대한 지형이 변화될 조짐이 엿보인다. 그러나 끊임없는 연구결과에 반영에 의한 Cited Reference, Related Records, Times Cited 및 Keyword Plus 등 다양한 기능을 추가한 WoS의 DB 품질제고 노력은 오랜 전통과 함께 DB 운영의 모범이 되고 있다.

4. Web DB 품질평가기준과 방법

이하에서는 본 연구에서 비교대상으로 삼은 4개 Web DB의 상호 비교를 위한 본 연구의 품질평가기준과 품질평가 방법 설정에 대해 알아

본다.

4.1 DB 품질평가기준을 위한 주요 연구 검토

국내 주요 선행연구의 DB 평가기준과 평가 방법을 종합하여 본 연구에 적용할 품질평가 기준을 설정하기 위해 다음처럼 주요 연구의 결과를 검토하여 정리해 보았다.

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

이용봉 등은 DB 평가에 관한 국내외 17개 선행연구를 종합 정리하여 DB 품질을 위해 데이터 품질과 서비스 품질이라는 품질기준을 추출하고 여러 가지 평가지표를 설정하였다(이용봉 외 2001, 114-115).

한편 이제환은 KERIS 종합목록 DB의 품질평가를 위해 선행연구를 요약하여 품질평가기준을 만들었고, 이 품질평가기준에 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).

최근에 홍현진(2005)의 연구는 웹 기반 DB의 품질평가를 위해 기준에 많은 연구가 있었

던 데이터와 서비스 영역의 품질평가는 물론, 효과성이라는 영역도 중시하여 전문가와 이용자를 대상으로 19개 평가지표와 45개 평가요소로 구성된 평가요소의 신뢰성에 대한 설문조사하여 설정된 품질평가기준의 신뢰성이 높다는 사실을 입증하였다.

국내의 DB 품질평가에 관한 주요 연구 중 안계성과 함정훈(2000, 20)의 한국데이터베이스진흥센터, 조순영(2002, 30) 등의 한국교육학술정보원, 이용봉 등(2001, 115)의 19개 평가기준, 11개 선행연구를 분석한 이제환(2002, 66)과 김성희(2003)의 연구, 4개 선행연구를 분석한 이만수 등(2003, 27), 홍현진(2005)의 연구에 나타난 DB 품질평가기준을 정리해보면 <표 2>와 같다.

그러나 먼저 수행된 한국데이터베이스진흥센터와 나중에 수행된 이용봉과 한국교육학술정보원의 동일한 17개 선행연구를 분석하여 표시한 품질기준 해당 선행연구의 건수가 각각 다르다(김상준 2004). 이러한 사실이 품질측정의 어려움을 단적으로 나타내주고 있다고 하겠다. 또한 품질평가기준에 필요한 항목에는 대체적으로 합의가 이루어졌지만, 세부 측정방안에 대해서는 조금씩 서로 다르다.

4.2 Web DB의 품질평가기준과 항목

본 연구에서는 Web 환경에서 제공되는 학술분야 서지정보 DB의 품질 비교평가를 범용 모델과 유사한 기준에 따라 사례중심으로 발견

〈표 2〉 국내 주요 선행연구의 DB 품질평가기준

구분	품질기준(Criteria)	이응봉 외 연구(2001)	이제환 연구(2002)와 김성희 연구(2003)	이만수 외 연구(2003)	홍현진 연구(2005)
데이터 품질	최신성 Currentness	-데이터 갱신 빈도	-현행성 -신속성 -갱신주기 -타임 래그	-최근 데이터의 갱신 -신속한 갱신	-갱신 주기(갱신 빈도) -최신정보 제공(타임 래그 발생 여부)
	정확성 Accuracy	-레코드 내 오류 데이터의 양 -레코드 작성 기준 유무(완전성에도 포함됨)	-데이터의 신뢰성 -데이터 표현 정확성 -데이터 기술 정확성(철자 오류, 누락 등)	-수록 데이터의 정확성과 일치성 -데이터 표현의 정확성 -철자 오류와 누락	-내용상 정확성(정보의 신뢰성) -문법적/어휘적 정확성(철자, 문법, 기타 오류)
	완전성 Completeness	-레코드 필드 구조의 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)	-데이터 값 누락(누락된 레코드와 빈 필드 값) -구성 필드 적절성(레코드 필드 구조의 완전성)
	포괄성 Coverage 또는 수록범위 Scope	-기타(수록범위)	-수록 자료의 범위 -수록 레코드의 수 -유형과 시기 망라성 -질의에 대한 응답율	-자료 범위/종류 다양 -수록 레코드의 수 -데이터의 유형과 시기 -가치 있는 자료 수록량	-레코드 수(정보의 양) -수록자료의 범위 -수록자료의 형태(유형) -소급 정도
	일관성 Consistency	* 일관성 대신 전문성(Expert) 항목만 별도로 품질기준 표시	-레코드 구조 일관성 -데이터 표현 일관성 -데이터 기술 통일성	-레코드 구조의 일관성 -데이터 표현과 기술의 통일성	-데이터 일관성(표현/기술) -데이터 정의를 레코드의 일치 여부
	(기타) * 홍현진의 연구에만 추가 분야		* 권위(전문성), 유일성, 객관성 등이 아래의 기타로 표시		〈권위〉 : 정보원 출처, 저자 수준, 정보제공 연락체계 〈객관성〉 : 정보제공 목적의 명확성, 정보의 비편향성 〈유일성〉 : 정보의 차별성
	편의성 (사용 용이성) Ease of Use	-메뉴 체계성/다양성 -키워드 용이성/다양성 -디자인 이용자 지향성 -검색결과 표시의 내용/주제별 다양성	-인터페이스의 편의성 -작업처리의 편의성 -접속의 편의성(* 김성희의 연구에는 이항목만 제외)	-인터페이스의 편리성 -접속의 편리성 -결과의 손쉬운 활용	〈디자인〉 : -화면구성(균형 정도) -이미지 사용(비율) -아이콘(가시성과 전달 능력)

서비스 품질	지원성 (사용자 지원성) Customer support	-오류 데이터 수정 체계 -DB 이용안내 제공 -Help 메뉴 제공 -Q & A 제공	-이용자 지원 및 교육 -도큐멘테이션의 유무 -도움 기능의 적절성 -원문제공의 편의성	-사용자 지원 및 교육 -도큐멘테이션, 도움말 제공 -원문접근의 용이성	<상호작용성> -도큐멘테이션, 이용자교육 -피드백 기능 제공 여부 -도움말 기능의 적절성
	검색성 Searchability 또는 접근성 Searching	-시소러스 관리 우수성 -다양한 탐색기법 제공 -복수 DB 검색 가능성 -검색식 1건 평균 소요 시간 -검색결과 출력 다양성	-검색속도 -검색방법의 다양성 -출력기능의 다양성 -시소러스의 유무		-검색방법의 다양성 -검색속도 -탐색의 편리성(메뉴 등) -검색결과와의 적합성 -검색결과 제공의 다양성 -원문의 링크/복사서비스 -링크정보의 다양성 -검색보조도구(색인 등)
	비용 Cost	* 품질기준으로 표시되었으나 평가항목에 제외	-구입비용 -이용비용	-DB 이용비용의 적절성	-도입/이용 비용의 적절성 -무료 정보원 이용성
	네트워크/하드웨어 Network & H/W		-접근성 -접근속도 -접속방식		<접근성> -접근속도, 접근방식 -네트워크/하드웨어 환경
기타 / 효과성	일반적인 요구사항 General facts desired	* 품질기준으로는 표시 없음	-권위(전문성) -통합성 -유일성 -객관성 -이용자 만족도 -데이터작성지침 준수	* 품질기준으로는 표시 없음	<장보적합성> : 주제분야/사용 언어/정보수준 적합성 <장보탐색충족률> : 이용자 요구/타 정보원 대비 성공 <장보유용성> : 정보 가치 <장보영향력> : 서비스를 받은 지역사회에 미치는 영향 <비용대비효과성> : 비용 대비 혜택(효과) 여부 <이용자 만족도> : 이용자가 느끼는 만족도

* 주) 본 표는 김상준(2004)이 선행연구에서 정리한 표에 볼드체가 차별화된 요소를 나타내는 홍현진(2005)의 최근 연구 결과를 통합하여 수정

적 평가를 통해 시도하고자 한다. 그러기 위해서는 국내 연구자의 선행연구를 기반으로 각 평가항목 중 도서관이나 정보센터 실무자가 쉽게 객관적으로 평가할 수 있는 항목을 중심으로 하여 Web DB의 품질평가기준을 추출하였다.

품질평가기준에 따른 측정방법은 다음의 평가기준을 상호 비교하기 위해 객관적 입장에서 쉽게 확인할 수 있는 내용을 중심으로 실시하였다. 이와 같이 쉽게 확인할 수 있는 객관적 평가요소를 중심으로 평가기준을 정한 이유는 Web DB 도입을 위한 선정단계의 실무적 품질 평가에는 이론적 품질평가기준과는 달리 전문가나 이용자를 동원한 실질적 평가를 하기가 시간 측면이나 예산 측면에서 어렵기 때문에 실현가능한 평가기준을 채택하였기 때문이다.

이러한 평가기준의 종류와 정의는 본 연구의 평가기준을 위해 근거대상으로 삼은 <표 2>의 5개 선행연구와 동일하므로 서술은 생략하고, 본 연구에서 학술 Web DB 도입을 위해 도입 전에 실무적으로 적용 가능한 내용을 중심으로 측정한 항목은 다음과 같다(김상준 2004).

- 1) 최신성 : 금년(2005년) 레코드 수, 발행빈도 빈번한 저널의 최근호 수록현황
- 2) 정확성 : 입력 데이터의 오류 사례, 오류데이터의 양(영국식과 미국식 영어, 철자 오류)
- 3) 완전성 : 공표된 레코드 수와 실제 레코드 현황, 레코드의 고유성이나 중복률
- 4) 포괄성 또는 수록범위 : 전체/연도별 레코드 수, 초록 보유비율

- 5) 일관성 : 저널 명의 표기 방법의 통일성, 레코드 구조의 일관성(대소문자 구분, 색인어의 일관성, 띄어쓰기)
- 6) 편의성 또는 사용 용이성 : 중복 데이터 제거, 시소러스 규모, 검색결과 활용의 편리성
- 7) 지원성 또는 사용자 지원성 : 전자 저널 원문링크, 원문복사와의 연계
- 8) 검색성 또는 접근성 : 검색속도, 시소러스 유무, 최신 레코드의 초록 보유비율
- 9) 비용 : 구입비용과 상대적 타당성
- 10) 네트워크/하드웨어 : 인터넷 외 추가 소요 사항
- 11) 기타 : 검색 가능한 요소를 중심으로 각 DB의 특성을 알 수 있는 검색에 의한 사례조사

5. Web DB 품질의 비교평가 결과

4개 DB의 비교평가를 위한 검색은 2005년 4월 14일부터 4월 21일 사이에 집중적으로 실시되었고, 1차 보완을 위한 검색이 5월 18일부터 5월 25일 사이에 실시되었으며, 2차 보완을 위한 검색이 7월 14일부터 7월 21일 사이에 실시되었다. 다만, NDSL 관련 정보 중 이용자 검색 기능만으로 불가능한 비교정보는 NDSL 운영 센터의 협조로 수집한 데이터를 함께 사용했고, WoS도 전체 데이터를 검색하기 위해 Thomson 한국지사의 협조로 많은 부분의 데이터를 수집했음을 미리 밝혀 둔다.

본 연구의 평가기준으로 삼은 데이터의 품질과 서비스 품질 영역의 10개 기준과 1개 사례

조사를 포함하여 각 세부 측정 항목별로 비교 대상 4개 Web DB의 품질을 비교평가 한 결과는 다음과 같다.

5.1 최신성(Currentness)

수록 레코드의 최신성을 알아보기 위해 각 DB의 2005년도 레코드 수를 검색해 본 결과, 조사시점이 4월이기는 하지만 <표 3>에서와 같이 NDSL에서는 35만5,070건인데 비하여, PubMed에서는 19만3,191건이고, Scopus에서는 22만1,373건을 수록하고, WoS에서는 44만9,676건을 수록하고 있었다.

WoS의 건수에는 수록논문의 발행연도가 아니라 DB 제작연도를 기준으로 연도가 검색되므로 2005년도 레코드에는 2004년도와 2005년도 논문이 혼재되고, 조사시점이 4월임을 감

안하면 2004년도 레코드의 비중이 상당히 높게 나타날 가능성이 있다. 이 점을 감안 하더라도 WoS에서 2005년도 레코드가 가장 많고, 특히 과학기술 분야의 SCI Expanded는 2005년도 전체 WoS의 91.9%인 41만3,273건을 차지하여 WoS의 세부 분야 중 그 비중이 가장 높았다. 더구나 SCI Expanded의 모든 레코드가 WoS의 모든 레코드에서 차지하는 비중이 88.5%에 달해 WoS를 SCI Expanded와 동일시해도 될 정도로 WoS 내 과학기술 분야 SCI Expanded의 비중이 높았다. NDSL도 WoS 다음으로 많은 최신 논문을 수록하고 있고, 전체 데이터 대비 2005년도 데이터의 비중도 1.5%로 가장 높았다. 그 이유는 NDSL도 KESLI 전자 저널 권소시업 출판사로부터 메타 데이터를 수시로 협조 받아 DB를 구축하므로 데이터 갱신주기가 매일이면서 출판사와 협력

<표 3> 수록년도별 레코드 수

연도	NDSL		PubMed		Scopus		SCI Expanded		WoS	
2005년만	355,070	1.5%	193,191	1.2%	221,373	0.8%	413,273	1.4%	449,676	1.4%
2004년만	1,799,613	7.9%	613,870	4.0%	1,327,221	4.9%	1,057,061	3.6%	1,146,116	3.5%
2000-2005	9,376,508	40.9%	2,948,436	19.0%	6,619,026	24.3%	5,512,979	19.0%	6,040,830	18.4%
1997-2005	14,078,763	61.4%	4,326,399	27.9%	10,114,603	37.1%	8,375,498	28.8%	9,221,657	28.1%
1985-2005	20,457,194	89.3%	9,023,352	58.2%	19,101,694	70.1%	17,311,118	59.6%	19,367,323	59.0%
1974-2005	21,954,636	95.8%	11,982,119	77.3%	24,877,227	91.3%	23,355,686	80.4%	26,338,487	80.2%
1966-2005	22,349,296	97.5%	13,640,524	87.9%	26,948,982	98.9%	26,095,687	89.8%	29,478,392	89.8%
1966 이전	561,636		1,869,686		(311,243)		2,955,801		3,358,965	
전체 레코드	22,910,992		15,510,210		27,260,225		29,051,488		32,837,357	

* 주) 위 결과는 DB별 필드 구조의 가변성으로 실제와 상이할 수 있고, NDSL은 프로시딩 500만여건을 제외.

관계가 치밀한 PubMed 이상으로 신속하게 DB 구축이 이루어지기 때문이다. 그러나 PubMed에 비해 주제분야가 다양한 Scopus의 2005년도 레코드 수가 상대적으로 부족한데, 그 이유는 Scopus내 Medline 데이터의 반입에 의한 갱신이 늦어지거나 Medline과 Scopus 자체 데이터와의 중복 레코드 제거작업에 시간이 많이 소요되거나, 수록 저널의 논문 데이터에 대한 출판사의 협조체제가 부족하여 입력이 지연되는 것으로 판단된다.

최신 연도인 2004년도와 2005년도의 레코드 수가 DB내 전체 레코드 수에 대비한 비중에서는 NDSL은 9.4%, Scopus는 5.7%, PubMed는 5.2%이며, WoS는 4.9%로 나타나 전반적으로는 역사가 오래된 DB 보다는 최근에 구축된 NDSL의 최신 데이터 비중이 두드러지게 높게 나타났다. 그리고 전체 데이터 대비 논문 이용이 집중되는 2000년 이후 레코드의 비중도 최근에 DB로 구축되기 시작한 NDSL이 40.9%나 되어 가장 높고, Scopus가 24.3%로 그 다음이며, DB 구축의 오랜 역사와 전통을 자랑하는 PubMed와 WoS가 각각 19.0%와 18.4%로 상대적으로 낮게 나타났다.

따라서 위 2005년도 수록 레코드의 건수가 연간 수록건수에 대비한 비중에 따른 최신성을 알아보기 위해 전년도인 2004년도 대비 가장 최근년도인 2005년도 레코드의 비중은 WoS가 39.2%로 가장 높고, PubMed가 두 번째인 31.5%이며, NDSL은 19.7%이고, Scopus가 16.7%로 나타나 Scopus의 데이터가 최신성 비중이 가장 낮고 전체 수록 저널의 60% 정도는 인쇄 저널을 입수하여 입력하는 관계로 DB 구축의 신속성이 떨어지는 것으로 나타났다.

또한 DB에 실제 수록되어 있는 저널의 최신성과 지연시간(타임 래그)을 시범적으로 비교하기 위해 생명과학 분야를 중심으로 매주 발행되는 *BBRC*(*Biochemical and Biophysical Research Communications*), *Nature*, *Science* 및 격주로 발행되는 *Cell* 저널의 최신 이슈 수록상황을 조사해 본 결과 <표 4>처럼 나타났다.

조사대상 4개 저널의 경우 NDSL과 PubMed는 색인은 하지 않고 출판사와 거의 동시에 서비스를 하고 있어 타임 래그가 크게 발생하지 않으며, Scopus는 평균 3주 미만의 타임 래그가 발생하고, WoS는 평균 3주 이상

<표 4> 주요 저널의 최신 이슈별 DB 수록 지연시간

저널명 / 출판사	NDSL	PubMed	Scopus	WoS
<i>BBRC</i>	330(4)	330(3)/1주 지연	330(2)/2주 지연	329(3)/5주 지연
<i>Nature</i>	434(7034)	434(7034)/지연 없음	434(7030)/4주 지연	434(7033)/1주 지연
<i>Science</i>	308(5720)	308(5720)/지연 없음	307(5717)/3주 지연	307(5717)/3주 지연
<i>Cell</i>	121(1)	121(1)/지연 없음	120(6)/2주 지연	120(5)/4주 지연

의 타임 래그가 발생하고 있다. 그리고 Scopus에서는 계열사 저널인 BBRC가 WoS보다 최신성이 우수했지만 Nature의 경우 WoS보다 최신성이 떨어져 Elsevier 계열사의 저널과 외부 출판사 저널 간에 최신성이 상반될 가능성이 엿보인다. 결국 NDSL과 PubMed는 거의 실시간으로 서비스된다고 보면 되지만, Scopus와 WoS는 계열사 저널이 아니면 출판 후 약 3-5주 정도가 지연되어 DB에 서비스되고 있다고 하겠다.

특이한 현상이지만, PubMed의 경우 심지어는 출판사 사이트에서 본격적인 서비스 이전이라도 PubMed에서 검색가능한 경우가 있어 출판사측보다도 더 신속히 서비스되는 경우도 관측되었다. 또한 PubMed에서는 '2004[dp] AND 2005[dp]'로 검색 시 "Gene, 2005 Feb 14;346C:71-81, Epub 2004 Nov 11."와 같이 전자 저널 발행년도와 인쇄 저널 발행년도가 혼재하는 현상이 발생하고 있었다. 그 이유는 2004년도에 논문게재 허가가 나서 전자 저널로 미리 서비스하는 논문인 경우 몇 달 뒤인 2005년에 정식 인쇄 저널이 발행되면 이 정보를 기준으로 데이터를 다시 정비하게 되어 있기 때문이다. 그래서인지 다른 DB에서는 연도 중복이 목격되지 않았지만, PubMed에서는 각 연도별 레코드 수의 합계와 연도에 대한 범위를 주어서 검색한 레코드 수의 결과가 불일치하는 현상이 발생하고 있었다. 따라서 전자 저널과 인쇄 저널이 병존하는 추세에 따라 최신성을 추구하는 PubMed에서는 논문인용에 혼

란을 일으킬 수 있는 문제점이 드러났다.

5.2 정확성(Accuracy)

영국식과 미국식 영어의 구분을 위해 Color와 Labor 및 Encyclopedia의 영국식과 미국식 용어를 검색해 본 결과, Color와 Labor에 대해서는 PubMed에서만 어휘통제가 되어 동일한 결과가 나왔고, 다른 DB에서는 전혀 어휘통제가 되지 않았다. Encyclopedia에 대해서는 다른 DB와 마찬가지로 PubMed에서도 어휘통제가 되지 않았다. 따라서 영국식과 미국식 영어에 대해 다른 DB에서는 전혀 어휘통제가 되지 않은 채 원시자료에 표기된 대로 입력하고 있었지만, PubMed는 Color와 Labor 등 부분적으로 어휘통제를 하고 있어 정확성이 높다는 사실이 밝혀졌다.

또한 철자 에러 정도를 확인하기 위해 February, Guatamala, Misssion, Government, Fransisco, Grammer, Recieve, Wensday, Seperate, Conditons, Bussiness, Infomation, Retrival 등 13개 검색어로 검색해 본 결과(Grammer는 일부 DB에서 인명도 같이 검색됨), NDSL에서는 모두 1,397건의 에러가, PubMed에는 모두 650건의 에러가, Scopus에는 모두 1,867건의 에러가, WoS의 경우 504건의 에러가 발견되었다. 따라서 WoS의 철자 에러율이 양적인 측면에서도 4개 DB 중 가장 낮고(표 5 참조), 전체 레코드 수 대비 철자 에러율도 미미하지만

〈표 5〉 정확성 측정결과

항목	검색어	NDSL	PubMed	Scopus	SCI Expanded	WoS
영국식과 미국식 영어 구분	Color	84,175	69,216	135,010	91,730	97,379
	Colour	37,553		49,533	29,283	30,047
	Labor	32,094	86,075	83,010	27,177	67,788
	Labour	23,670		42,199	8,703	19,736
	Encyclopedia	2,342	309	783	1,025	4,824
	Encyclopaedia	432	33	175	122	218
철자 예러	Febuary	7	5	34	15	17
	Guatamala	9	4	14	6	8
	Misssion	3	0	2	0	0
	Goverment	87	23	215	32	58
	Fransisco	24	27	46	6	7
	Grammer*	98	377*	110*	31	48
	Recieve	259	19	114	33	35
	Wensday	0	0	0	0	0
	Seperate	513	45	324	131	143
	Conditons	320	131	862	131	132
	Bussiness	16	0	14	3	4
	Infomation	40	12	90	24	34
	Retrival	21	7	42	17	18
	소 계	1,397	650	1,867	429	504

* 주) Grammer는 PubMed와 Scopus에서 일부 저자명으로도 검색되어 상대적으로 높게 나타남.

WoS가 가장 낮게 나타나 WoS의 정확성이 가장 높다고 볼 수 있다. 반면 최근에 구축된 Scopus와 초록이 많지 않은 NDSL에서 철자 예러 양이 오랜 전통을 자랑하는 PubMed와 WoS 두 DB보다 상대적으로 많아 대체적으로 DB 구축 초기에는 질적인 측면보다 양적인 측면에 치중함을 엿볼 수 있다. 따라서 NDSL과 Scopus에는 지속적인 DB 품질관리의 필요성

이 높다고 하겠다.

레코드 내 오류나 데이터의 정확성 측면에서 발견된 특이한 사실은, PubMed의 경우 출판사 측에서 데이터 수집 협조를 받는 관계로 “*Med Lett Drugs Ther.* 2006 Apr 11;47(1205):31-32.” 저널에서 출판사 측의 연도 관련 서지작성 오류가 PubMed의 검증과정 없이 그대로 반영되어 데이터 1건의 발행년도

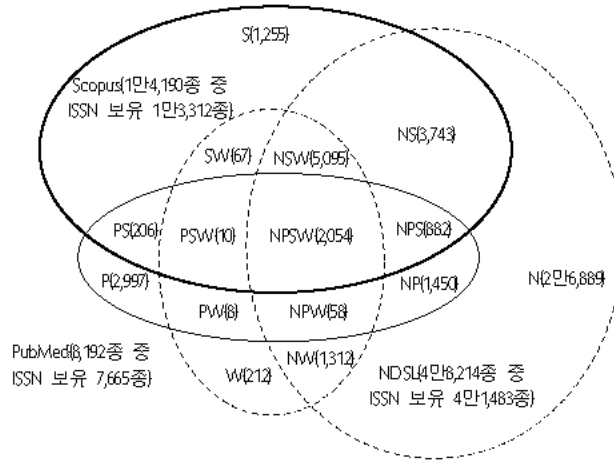
가 2006년도로 검색되는 현상을 발견하였다. NDSL과 Scopus에서는 확인할 수 없었지만, WoS의 경우에도 동일한 저널 47권에 대해 '25 26'과 '27+'라는 형식으로 2건의 논문이 검색되면서 발행년도는 2005년도로 나타났다. 이와 같은 현상은 DB를 직접 구축하지 않고 외부에 의존하는 NDSL과 Scopus는 물론 Medline에 수록되기 전 임시로 데이터를 수록하는 PubMed에서 발생될 개연성이 높다고 하겠다. 따라서 외부 데이터를 도입하여 DB를 구축한다고 하더라도 데이터의 정확성을 위해 더 많은 신경을 써야만 할 것이다.

5.3 완전성(Completeness)

공표된 레코드와 실제 레코드 수의 차이를 홈페이지 등의 자료와 실제 검색결과를 비교해서 조사해보았다. 그 결과 NDSL은 홈페이지 첫 화면에 'NDSL 데이터베이스 구축현황'이라는 곳에 최근의 DB 구축현황을 그림과 표로 나타내어 이용자의 정보파악을 쉽게 해주고 있다. 또한 홈페이지 자료와 실제 검색결과 사이에 큰 격차도 나타나지 않아 홈페이지 자료의 최신성도 유지되고 있었지만 'NDSL 데이터베이스 구축현황'의 작성 기준 날자가 제시되지 않아 아쉬움이 남는다. PubMed는 홈페이지에 DB 구축현황에 대한 자료를 찾을 수 없다. 다만, 도움말 등의 자료에는 1966년 이후 Medline 자료 약 1천5백만건과 1950-1965년의 OldMedline 자료 약 200만건 등 1,700만

건 정도가 수록되는 것으로 보이지만, 실제 데이터는 1902년 데이터부터 존재하며, 총 레코드 건수가 약 1,551만건에 그치고 있고, 공표된 레코드가 실제 레코드 수보다 많아 홈페이지 도움말 등의 자료에 대한 신뢰성이 떨어진다. Scopus도 공표된 자료에 의하면 1966년 Medline 데이터 이후부터 수록되었다고 하지만 실제 공표된 것보다 더 과거분 데이터가 존재하여 불일치가 발생하고 있다. WoS의 홈페이지에는 DB에 구축된 레코드 현황자료를 찾을 수 없다.

레코드의 고유성을 간접적으로 평가하기 위해서 4개 DB 간에 수록 저널의 중복률을 조사해보면 다음 <그림 2>와 같이 나타낼 수 있다. 각 DB 운영처에서 제공하는 수록 저널 리스트에서 ISSN 중복을 확인하여 중복을 제거한 후 ISSN이 없는 저널은 제외하고 ISSN을 보유한 저널만을 대상으로 비교하기 위해 최종 정리한 후 전산 프로그램을 이용하여 비교 조사하였다. 저널 수는 NDSL은 총 4만8,214종에서 ISSN 중복제거 후 4만7,558종이지만 ISSN 보유 저널이 4만1,483종이며, 뉴스지 등 Medline 보다 수록 저널 종수가 급증한 PubMed는 총 8,192종에서 ISSN 중복제거 후 8,116종이지만 ISSN 보유 저널이 7,665종이며, Scopus는 총 14,190종에서 ISSN 중복제거 후 14,126종이지만 ISSN 보유 저널이 1만3,312종이며, WoS는 총 9,320종에서 ISSN 중복제거 후 8,859종이지만 ISSN 보유 저널이 8,816종이다. 조사 결과 4개 DB에 공통인 저



〈그림 2〉 4개 DB 수록 저널의 중복정도와 고유성

널은 2,054종이며, 3개 DB에 공통인 저널은 5,095종과 882종과 58종과 10종이며, 2개 DB에 공통인 저널은 3,743종을 포함하여 6가지로 나타나며, 각 DB에 고유한 저널은 NDSL 2만6,889종과 PubMed 2,997종과 Scopus 1,255종과 WoS 212종으로 나타났다. 따라서 NDSL과 PubMed는 고유성이 높다고 할 수 있지만 NDSL은 국내의 영문학회지가 다수 누락되고, PubMed는 소식지 등이 포함되어 추가된 고유한 저널의 질적인 부분이 의심되며, Scopus는 WoS의 SCI 저널을 모두 포함하려는 정책이라지만 아직은 그러하지 않고 Scopus에 없는 데이터를 Medline 데이터로 보충하면서 데이터 중복체크가 완벽하지 않아 일부 중복된 데이터를 포함하는 경우가 있고, 저널 선정을 엄격하게 유지하는 WoS는 212종 외에는 다른 DB에서도 검색이 가능하다.

레코드의 필드 구조를 평가하기 위해 DB의

필드 명과 내용을 보면, 4개 DB에 공통인 필드는 필드 명이 다르더라도 실질적으로 동일한 필드는 논문제목, 저자명, 초록 및 연도이고, NDSL에서는 저자주소 필드가 구축되기는 하지만 검색이 불가능하며, 구축되는 필드도 단순하고, PubMed에만 있는 필드는 Named Person 등이지만 Medline 형태로 보면 필드가 다양하게 존재하고, Scopus도 검색결과 화면에서 보여주는 필드가 NDSL처럼 단순하다. 반면 WoS는 검색결과 활용화면에서 다양한 선택이 가능해 필드의 사용이 조금 더 자유롭다.

5.4 포괄성(Coverage) 또는 수록범위(Scope)

DB에 수록된 레코드의 총 건수는 NDSL은 프로시딩 500만여건을 제외한 논문만 2,291만 992건이고, 생명과학 및 의학 분야의 PubMed

는 1,551만210건이며, Scopus는 Medline 데이터를 포함해서 2,726만225건이고, WoS는 3,283만7,357건으로 수록년도의 차이를 감안 하더라도 WoS가 가장 많이 수록하여 포괄성이 가장 높다고 할 수 있다. 각 DB의 2004년도 레코드 수를 검색해 본 결과, <표 3>과 같이 NDSL은 179만9,613건이며, PubMed는 61만 3,870건이고, Scopus는 132만7,221건이며, WoS는 114만6,116건 이었다. 최근 1년분(2004년도) 전체에 대한 레코드 수로만 보면 생명과학 분야를 중심으로 수록하는 PubMed나 엄격한 저널 선정과정을 거치는 전학문 분야의 WoS보다 인문·예술, 사회 및 자연과학 전학문 분야를 수록하는 NDSL이나 사회 및 자연과학 분야를 수록하는 Scopus가 최신 1년 단위(2004년분)의 레코드를 더 많이 포괄적으로 수록하고 있었다. 그렇지만 생명과학 분야만을 수록하는 PubMed가 광범위한 주제분야를 가진 다른 DB의 절반 정도 레코드 수를 가지고 있으므로 생명과학 관련학문 분야의 연구 활동이 매우 활발하고 전학문 분야에서 높은 비중임을 볼 수 있다.

과학기술 분야 중심의 4개 DB 특성상 최근 년도의 논문에 대한 이용이 집중된다는 연구결과를 반영하여 각 DB의 최근 9년분인 1997년도 이후 레코드 수를 검색해 본 결과, <표 3>과 같이 NDSL은 1,407만8,763건이며, PubMed는 432만6,399건이고, Scopus는 1,011만 4,603건이며, SCI Expanded는 837만5,498건인 반면 WoS는 922만1,657건 이었다. 따라

서 최근 9년분에 대한 포괄성은 NDSL이 가장 높고, Scopus는 WoS보다 약간 높으며, 주제 범위가 좁은 PubMed가 가장 낮다.

아울러 전체 데이터 대비 1985년 이후 레코드의 비중은 NDSL이 89.3%나 되어 가장 높고, Scopus가 70.1%로 그 다음이며, 59%의 WoS와 58.2%의 PubMed가 비슷한 수준으로 나타났다. NDSL의 경우 최근 20년 정도의 데이터가 전체 데이터의 절대다수인 90% 정도를 차지하였다. 따라서 1985년 이후 20여년간의 레코드 수나 최근 1년분 전체인 2004년도 레코드 수에서 NDSL이 4개 DB 중 가장 많아 초록이 없는 데이터가 많기는 하지만 포괄성과 데이터의 양적인 측면에서는 우수함을 볼 수 있다. 그렇지만 NDSL의 초록 보유율은 29.7%로 나타나 논문의 제목 중심으로 검색하게 되면 정확성이 떨어지고 데이터가 존재함에도 불구하고 검색되지 않을 확률이 높아 주제어 검색에 많은 한계점을 가지게 된다.

한편, Scopus는 과학기술 선진 출판국가가 아닌 국가에서 발행되는 학술 저널을 상대적으로 많이 포함하고 있는데, 1만4,190종의 저널 리스트 중 미국 저널이 4,607종(32.5%), 영국이 2,923종(20.6%), 네덜란드가 902종(6.4%), 독일이 806종(5.7%)의 저널을 수록하고 있다. 또한 브라질 66종, 러시아 199종, 인도 171종, 중국 365종, 일본 366종에 비해 한국은 전체의 0.3%인 37종(*Molecules and Cells*는 홍콩 Springer에서 출판되어 제외)을 수록하고 있어 Scopus 내 한국 저널의 비중이 경쟁국가에

비해 상대적으로 낮게 나타나 국내 학술출판 분야의 육성이 필요한 실정이다. 위의 주요국가 저널 외에 Scopus에는 Medline DB의 유일 저널이 저널 리스트 상에 전체의 7.9%인 1,114종(또는 다른 자료에 의하면 1,250종)이 포함되어 있었다.

5.5 일관성(Consistency)

저널 명 표기의 일관성을 조사해보면, PubMed와 Medline 데이터에서는 “*Biochimica et biophysica acta* 1619(1), 1 8”로 나타나고, NDSL의 저널 명 검색결과 “*Biochimica et biophysica acta, General*

subjects”로 나타나지만, Scopus 데이터에는 “*Biochimica et Biophysica Acta General Subjects* 1619(1), 1 8”과 Medline 형태의 데이터 2개가 동시에 나타나고, WoS에는 “*BIOCHIMICA ET BIOPHYSICA ACTA GENERAL SUBJECTS* 1619(1): 1 8 JAN 2 2003”로 나타나 Scopus에는 Medline 데이터와의 통합과정에서 중복 데이터의 제거 문제가 완벽히 해결되지 않았음이 드러났다.

일관성 항목 중 대소문자 구분은 4개 DB 모두 구분하지 않고 있어서 최근의 추세를 반영하고 있다고 볼 수 있다.

또한 색인어의 일관성을 알아보기 위해 databases, database, data base, db로 검색

〈표 6〉 일관성 측정결과

항목	검색어	NDSL	PubMed	Scopus	SCI Expanded	WoS
대소문자 구분	DATABASE	93,641	62,716	160,579	67,108	71,292
	database					
	UNION	45,730	27,979	64,018	23,210	46,268
	union					
색인어의 일관성	databases	93,641	42,202	160,579	26,368	28,823
	database		62,716		67,108	71,292
	data base	35,538	37,839	96,684	9,157	9,986
	db	44,276	16,698	71,439	33,150	35,797
뛰어 쓰기	cd rom	5,334	1,220	3,961	1,897	3,233
	ed-rom	5,277				
	cdrom	145	835	115	80	187
	on line	221,703	422,831	825,105	10만 이상	10만 이상
	on-line	49,388	11,654			
	online	49,037	14,897	43,612	23,293	29,330

해 본 결과, 4개 검색어에 대한 완벽한 어휘통제가 이루어진 DB는 없었고, NDSL과 Scopus는 단수와 복수를 모두 동일하게 처리했지만, PubMed와 WoS 두 DB는 단수와 복수를 달리 취급하는 차이점이 밝혀져 어휘통제가 이루어지지 않아 일관성이 떨어졌다(〈표 6〉 참조).

그리고 띄어쓰기에서 발견된 특이한 현상은 PubMed에서는 'cd rom'과 'on line' 사이에 처리 방식이 서로 다르며, 대체로 NDSL 이외에는 'on line'과 'on line'처럼 단어 사이의 공백과 대수()는 구분하지 않고 동일하게 처리하고 있어 검색결과가 동일하였다. 그렇지만, NDSL과 PubMed는 다른 DB와 달리 원시 자료에 따라 'on line'과 'on line' 및 'online' 등 3가지로 색인하고 있었다.

5.6 편의성 또는 사용 용이성(Ease of Use)

PubMed의 기본이 되는 Medline 데이터는 Scopus와 WoS는 물론 BMN(BioMed Net), BP(Biosis Previews), Embase, SciFinder 등 관련분야의 여러 DB에 통합되어 서비스된다. 따라서 Scopus와 WoS의 경우 자체 DB와 Medline 데이터를 동시에 검색할 수 있어서 이용자의 편의성과 검색 망라성을 높여주고 있다. 물론 통합 운영하는 DB에서 자체 DB와 Medline 데이터의 중복 데이터 제거 정도와 제거의 정확성에 따라 통합의 효과와 효율성이 좌우될 것이다. 따라서 과학기술 분야 중 가장 광범위한 의학 및 생명과학 분야의 Medline

데이터를 포함하는 Scopus나 WoS와 달리 NDSL에는 Medline 데이터가 완벽하게 포함되지 않아 과거분 생명과학 분야 자료의 품질과 데이터의 충실도가 떨어져 NDSL에도 반드시 Medline 데이터가 포함되어야 포괄성과 함께 편의성이라는 DB의 품질이 높아진다고 할 수 있다.

다만, WoS의 경우 'ISI Web of Knowledge' 매뉴얼에서 WoS와 Medline을 동시에 검색하면, 검색결과에서 동일한 문헌에 대해 중복 데이터는 제거된 상태에서 보여준다. 그렇지만, WoS와 PubMed를 동시에 검색하면 검색결과에서 동일한 문헌에 대해 중복 데이터는 제거되지 않고, 두 DB의 검색결과를 각기 보여주는 차이점이 있다. 따라서 WoS의 경우 Medline과 PubMed의 차이점을 정확히 인식하지 못하면 검색방식 선택에서 오해가 발생할 가능성이 나타났다. 이러한 현상은 Scopus에서도 Scopus 자체에서 구축된 데이터는 Medline 데이터와 중복될 경우 자체 데이터를 사용하지만, 중복되지 않는 Medline 유일 1,250여종 저널의 데이터는 그대로 받아들여 통합 서비스하는 관계로 Scopus 내 데이터 형태에서 이중구조를 가지게 되었다.

또한 약품 및 화학 정보는 PubMed의 MeSH에는 7,000개 정도가 수록되어 있지만, Scopus의 Emtree에는 2만개 정도가 수록되어 약품 및 화학 정보 검색에 Scopus가 더 편리할 수 있다. 그러나 PubMed는 Nucleotide, Protein Sequences, Protein Structures,

Complete Genomes, Taxonomy, OMIM 등과 동일하게 Entrez 검색 시스템에서 운영되므로 서지정보뿐만 아니라 추가적인 정보검색이 가능하다.

NDSL의 검색결과는 초록 보기, 원문 열어 보기, 소장기관 보기, 원문 신청하기, 개인 파일 화, 바꾸니 담기, E mail이나 PC로 다운 등을 할 수 있지만, 개인용 논문관리 프로그램으로 다운이 불가능하다. 그리고 PubMed의 검색결과는 Text, File, Clipboard, E mail, Order 등으로 활용할 수 있고 개인용 논문관리 프로그램으로 다운도 가능하다. 또한 Scopus는 검색결과를 출력할 수 있고, E mail로 송부도 가능하고, My List에 임시 저장은 물론, EndNote, ProCite, Reference Manager, RefWorks 등 4개의 개인용 논문관리 프로그램을 위한 RIS format과 ASCII format 형식으로도 다운이 가능하여 편리하고, 웹 자원 Scirus(<http://www.scirus.com/>) DB와 특별히 연결하여 검색결과를 동시에 볼 수 있으며, 검색결과의 분석기능이 편리하게 설계되어서 누구나 쉽게 이용할 수 있어서 유용하다. 반면에 WoS는 1회 500건 미만으로만 검색결과를 출력 및 저장할 수 있지만, E mail로 송부도 가능하고, EndNote 등 4개 개인용 논문관리 프로그램을 위한 다운로드 가능하고, 최순순/제1저자순/수록정보원순/피인용 횟수순으로 정렬도 가능하고, 지정된 논문이 피인용 시 E Mail로 Alert 서비스도 가능하고, 구독중인 JCR(Journal Citation Report) 데이터 및 소

장정보로 바로가기도 가능하고, 단백질 및 DNA 서열정보는 물론, 화학 반응정보와 화학 구조도도 그럴 수 있고, 공동 참고문헌 빈도순으로 정렬된 관련문헌을 추가적으로 보여주는 'Related Record'도 가능하고, 'Order Record'를 할 수 있고, 2,000건의 검색결과에 대해 8개 항목으로 된 검색결과 분석기능인 'Analysis Results'도 포함되어서 매우 유용하다. 그리고 WoS에는 Thomson ISI 사의 다른 DB와 상호연계가 매우 우수하지만, 다른 상용 DB를 구독하는 경우에만 연계효과가 있어 예산이 부족한 국내 도서관계에는 실질적으로 많이 사용이 되지 않고 있다.

5.7 지원성 또는 사용자 지원성(Customer support)

원문 링크 측면에서, NDSL은 4만7,558종 수록 저널 중 1만8,870종(39.7%)이 원문에 링크되어 있어서 원문 링크 비율은 낮지만, 링크된 전자 저널 총수로는 4개 DB 중 가장 많아 지원성이 높다고 하겠다. PubMed에서는 Medline 수록 4,714종 저널이 링크되고 있어서 전체 Medline 수록 저널의 약 98% 정도가 전자 저널 출판사 사이트와 직접 링크되어 있다. 그리고 PubMed의 특이한 기능으로는 검색결과 중 전자 저널이 링크되어 있고, 자관에서 해당 전자 저널을 구독하는 경우 자관의 소장정보와 함께 전자 저널 구독 정보를 사전에 입력하여 두면 PubMed 검색결과를 보여주는

과정에 자관 구독 전자 저널 여부를 쉽게 알 수 있도록 자관의 등록된 아이콘을 표시하여 원문을 링크해 주는 LinkOut 기능이 있다. PubMed에서 제공되는 이 기능은 Scopus나 WoS보다 원문 입수와 활용 측면에서 더 편리한 기능이다. Scopus는 제작사인 Elsevier 등 일부 출판사 전자 저널만 원문 링크 되어서 검색 후 곧바로 원문을 볼 때 자관에서 구독하는 저널일지라도 원문 링크가 일부에 그쳐 불편한 측면이 있다. WoS는 159개 출판사 약 7,000종의 전자 저널과 원문 링크 되어서 검색 후 곧바로 원문을 볼 때 자관에서 구독하는 저널이면 7,000종은 곧바로 볼 수 있다.

또한 NDSL에서는 논문 검색결과 화면에 소장기관 보기와 원문신청 하기 메뉴가 있어서 곧바로 비용이 저렴하고, 편리한 국내 도서관에 우선적으로 DDS를 신청할 수 있어 편리하다. 이러한 DDS 서비스에는 KESLI 및 NDSL의 DDS만 참여하는 396개 도서관 중 248개 기관이 DDS에 참여 중으로 서울대학교 등 일부 도서관이 참여하지 않았지만, 많은(참여율 63%) 도서관이 참여하여 신속하고 편리하게 원문복사 서비스를 상호 주고 받고 있다. PubMed에서는 검색 후 자관 비소장 자료에 대해서는 Loansome Doc이라는 DDS 서비스와 연계되어 있어서 북미지역의 경우 손쉽게 원문복사 신청을 할 수 있어서 서비스에 가입한 도서관이라면 이용자에게 더 편리한 서비스라고 하겠다. 물론, 무료로 이용가능한 OA(Open Access) 전자 저널의 링크 서비스도

PMC를 포함해서 풍부하다. Scopus는 검색결과 화면에서 비소장 저널에 대한 원문신청 기능이 없어 불편할 수 있다. 그러나 WoS에서는 InfoTrieve를 통한 ISI Document Solution으로 원문입수가 가능하다.

검색결과 분석기능은 두 상용 DB인 Scopus와 WoS에서 제공하는 기능으로 제한적이긴 하지만, Scopus가 인용한 정보와의 링크 측면에서 더 편리하다. 이 분석기능은 가장 대표적인 이용자 위주의 고급 부가기능으로 NDSL과 PubMed에서는 벤치마킹하여 도입을 검토할 필요가 있다. 그리고 PubMed에서는 검색결과를 'All'과 'Review'로 나누어 보여주므로 연구초기 단계나 다른 분야에 대한 사전지식이 없을 경우 Review 형태의 자료만 구별하여 이용할 수 있어 편리하다.

5.8 검색성(Searchability) 또는 접근성(Searching)

NDSL에는 색인집이나 시소러스가 없지만, PubMed에는 MeSH가 있고, Scopus에는 GEOBASE Subject Index, MeSH, FLX terms, WTA terms, Regional Index, Species Index, Ei thesaurus는 물론 Emtree라는 역사를 자랑하는 색인집과 시소러스가 있어서 이용자의 검색성을 도와줄 수 있다. WoS도 검색 단계에서 'terms list(dictionary)' 사용이 가능한 것으로 보아 색인어 통제장치가 있다고 판단된다.

또한 검색어 제한(Limits), 검색진행 과정(History), SDI 서비스(Alerts) 등은 4개 DB에 모두 포함되어 있어서 본질적인 차이를 찾을 수 없었다. NDSL은 검색단계에서 출판년도를 제한하여 검색할 수 있고, PubMed에서는 출판년도로 검색 시 연도가 범위를 벗어나도 검색결과를 보여 주지만, Scopus는 검색단계에서 출판년도를 제한하여 검색할 수 있음은 물론 범위를 벗어난 연도는 결과를 보여주는 화면으로 넘어가지 않고 검색식 입력화면에 계속 놓이게 된다. 반면, WoS는 검색 직전에 검색대상 데이터의 DB 입력년도를 선택하고 검색을 시작하게 된다.

그리고, 검색화면이 가장 단순하면서 가장 많이 사용하는 무료 PubMed에 비해 NDSL, Scopus 및 WoS는 검색속도 면에서도 느린 편이다. 특히, NDSL에서는 대량의 건수가 검색될 때 너무 느린 편이어서 문제점으로 지적된다.

초록이 있는 레코드 건수가 전체로는 Scopus가 NDSL이나 PubMed보다 많다. NDSL의 초록 보유율은 2005년 데이터는 42.1%이고, 2000년 이후분은 22.6%이며, 1985년 이후분은 25.6%로 나타나 최근 데이터는 출판사에서 초록과 함께 메타데이터를 협조 받는 관계로 초록 보유율이 상대적으로 높지만, 과거 데이터를 포함한 초록 보유율은 전체적으로 29.7%(NDSL 검색 시스템 화면상에서는 27% 수준으로 표시)에 그치고 있다. WoS는 1991년 데이터부터 저자 초록이 존재한다.

실제 검색성능을 평가하기 위해 2004년도

에 발행된 'stem cell' 자료를 찾기 위해 필드를 제한하지 않고 NDSL에서 'stem cell'로 검색하면 5,542건이 검색되고, PubMed에서 "stem cell" AND 2004[dp]로 검색하면 4,785건이 검색되지만, Scopus에서 'TITLE ABS KEY("stem cell") AND PUBYEAR IS 2004'로 검색하면 4,358건이 검색되고, WoS에서는 DB 수록년도가 2004년분 중 'TS=(stem cell)'로 검색하면 5,012건이 검색되어 초록이 많지 않은 NDSL에서 더 많은 검색건수가 나왔다. 다시 'stem cell'이라는 검색어를 자료의 제목으로 한정시켜 2004년도 논문을 검색하면, NDSL에서는 4,719건이고, PubMed가 1,657건이고, Scopus에는 1,627건이고, WoS에서는 2,747건이 검색되어 NDSL에서 가장 많은 검색결과가 나타나 위에서와 비슷한 결과를 가져왔다.

검색성능을 더 알아보기 위해 과학기술 분야의 생명과학과 인문·사회과학 분야의 주제어와 저자명을 검색하여 보았다. 그 결과, 인문·사회과학 분야의 주제어는 Scopus와 달리 NDSL과 생명과학 분야의 PubMed에는 건수가 적었고, 생명과학 분야 주제어의 경우 각 DB마다 조금씩 차이를 보였다. 특히 PubMed는 자동용어매핑(Automatic Term Mapping)에 의해 MeSH 색인어로 질의어를 변경하여 처리하는 관계로 'HIV protease inhibitors'로 입력하면 "HIV protease inhibitors"와 같이 구로 인식하여 동일하게 처리하는 경우가 있었다.

〈표 7〉 검색성 평가를 위한 주제어와 저자명 검색결과

항 목	검색어	NDSL	PubMed	Scopus	SCI Exp. (1992-현재)	WoS (조사 불가)	
주제어 (생명 과학)	stem cell	35,371	142,858	72,080	34,851		
	stem AND cell	35,371	96,272	72,080	55,016		
	"stem cell"	30,821	40,594	43,882	34,851		
	stem cell research		1,387	84,346	3,461	527	
			1,387	57,441	3,461	1,992	
			623	821	826	527	
	HIV protease inhibitors		3,160	6,384	5,965	1,122	
			3,160	7,945	5,965	5,611	
			913	6,384	1,051	1,122	
주제어 (인문 사회 과학)	volumetric turbine flow meters	2	0	52	0		
		2	0	52	1		
		0	0	0	0		
	rotational angular momenta	206	6	1,311	15		
		206	6	1,311	193		
		13	6	9	15		
	spontaneously broken gauge field theories	120	0	112	0		
		120	0	112	94		
		0	0	0	0		
	economic globalization	1,095	255	3,082	25		
		1,095	255	3,082	226		
		217	18	296	25		
저자명	Axel, R(노벨의화학상)	110	149	143	51		
	Buck, L B(노벨의화학상)	13	38	87	34		
	Ciechanover, A(노벨화학상)	87	148	152	103		
	Hershko, A(노벨화학상)	67	107	108	50		
	Rose, I(노벨화학상)	110	269	541	23		
	Gross, D J(노벨물리학상)	56	120	242	119		
	Politzer, H D(노벨물리학상)	19	10	6	9		
	Wilczek, F(노벨물리학상)	151	48	135	155		

* 주) WoS는 Thomson 한국지사의 협조 거절로 조사가 불가능해 SCI Expanded로 비교했으며, 주제어는 LaGuardia의 논문에서 사용된 검색어를 재사용하되 주제어 그대로, AND로 연산, " "로 검색 등 3가지 검색결과를 순서대로 표시했으며, 저자명은 2004년도 노벨상 과학 분야 수상자 (조사일: 2005. 7. 14-21)

또한 PubMed는 저자를 입력하면서 그동안 4회에 걸쳐 수록정책을 변경한 관계로 논문에 수록된 모든 저자가 검색되지는 않는다. 저자명이 같은 경우, 동일한 표기의 저자명 키워드를 가진 연구자가 여러 주제분야에 걸쳐 동시에 존재하는 저자명일 경우 검색건수에서 큰 차이가 날 가능성이 있고(표 7) 참조), 전반적으로 NDSL은 저자명 검색에서 지극히 저조한 결과를 보여주고 있다. <표 7>에서 나타난 것처럼 저자명 검색에서 전반적으로 Scopus가 풍부하게 검색되기는 하지만 일관성이 없고, Scopus와 WoS와 직접 비교하기는 힘들지만 자연과학 분야의 저자명 검색에서 Scopus가 더 망라적으로 검색되어 선행연구(LaGuardia 2005)의 결과와 상반되었다.

주목할 점은 'stem cell research'와 같이 주제어를 입력 후 검색할 경우, NDSL과 Scopus는 주제어 사이의 공백을 'AND' 연산자로 인식하여 처리하지만, WoS는 주제어 사이의 공백을 구로 인식하여 "stem cell research"와 같이 검색하며, PubMed는 사례간에 서로 다른 결과가 나타나 처리 원칙을 찾아볼 수 없다. 또한 PubMed에는 검색필드를 제한하지 않고 검색하면 MeSH에서도 검색되는 관계로 생명과학 분야의 주제어 중 핵심 분야는 다른 DB보다 풍부하게 검색되는 편이다.

5.9 비용(Cost)

상용 Scopus나 WoS는 비용 측면에서는 불

론, 무료 DB인 NDSL과 PubMed보다 불리하다. 그러나 구독비용만이 상용 DB의 이용을 저해하는 요인이라고 보이지는 않는다. 더구나 무료 NDSL과 PubMed를 널리 사용하는 분야에서는 국내에서 널리 인지되고 많이 사용되는 PubMed 외에 PubMed와 수록 저널의 중복도가 높은 Scopus와 WoS 등 다른 상용 DB를 추가로 사용해야 한다면, 굳이 PubMed 외에 다른 DB를 사용해야만 획득 가능한 고유(유일)한 정보의 효과와 구독비용 간의 근거가 다른 상용 DB 도입의 최대 관건이라고 판단된다.

현재 20여개 국내기관이 금년부터 시범적으로 이용하다가 2006년 8월까지 구독여부(구독하면 시범기간도 소급해서 구독기간으로 적용)를 최종 결정할 수 있는 Scopus는 혁신적 시장 진입 프로그램을 준비하여 고가의 WoS 중심 국내 인용정보 DB 시장에 적극적인 자세로 도전하고 있다. 그러나 시장진입 후 조만간 구독가격을 정상화시켜 올리지 않을까 하는 의구심을 도서관 실무자들은 갖고 있다. WoS의 경우 그동안 유지해 온 고가정책은 Scopus의 시장진입으로 완화되어 제공년도를 무료로 더 늘려 제공함으로써 실질적인 가격인하 정책을 최근 펼치고 있다. 그렇지만 1900년까지 소급되어 구축되는 Century of Science를 포함하여 모든 데이터를 완전히 갖추려면 커다란 규모의 예산을 필요로 하여 국내의 많은 도서관(현재 약 23개 기관 구독 중)에서는 일부 연도의 데이터만 구입하여 불완전한 검색이 이루어지는 현실이어서 최신년도의 정보가 핵심정보임에는

틀림없다고 하더라도 문제점이 아닐 수 없다.

결국 무료 NDSL과 PubMed의 이용이 활성화된 연구현장에서 상용 Scopus나 WoS DB의 도입가치는 중복된 데이터에 대해서는 얼마나 부가적이고 독자적인 추가 가공을 해 부가가치를 창출했느냐와 중복되지 않는 레코드의 독자적 가치와 그 우수성 및 예산 규모에 따른 자판의 필요성에 달려있다고 보인다.

5.10 네트워크/하드웨어(Network & H/W)

4개 DB는 인터넷의 Web으로 제공이 가능한 학술 DB여서 네트워크 및 하드웨어 설치에 대한 부담은 없다.

다만, 실제로 측정을 하지는 않았지만 무료 PubMed에 비해 NDSL과 Scopus 및 WoS를 검색을 위한 접근성과 접근속도 및 접속방식에서 일반적으로 더 느리고 불편한 편이다.

5.11 기타(Others) : 사례조사

4개 DB에 수록된 2005년도 상반기 한국인 저자의 *Cell*, *Nature* 및 *Nature* 자매지, *Science* 등 주요 저널에 실린 논문을 분석하기 위해 NDSL은 저자의 주소검색이 불가능해 제외하고, PubMed는 책임저자의 주소만 검색되고 공동저자로 참여한 경우는 검색이 불가능해 논문의 주도적 연구기관이 한국 기관인 경우만 검색이 가능하므로 검색결과가 Scopus나 SCI Expanded와 상이함(거의 대부분 더 적음)을 감안한 사례조

사 결과 다음과 같은 차이점이 발견되었다.

SCI Expanded에서는 총 34건이 검색되었지만, 2004년도분 4건과 논문이 아닌 자료 1건 등 5건을 제외하면 29건이었다. 반면 Scopus에는 27건이 검색되었고, PubMed에는 공동저자로 한국인이 참여한 논문이 제외되어 한국 연구기관이 주도한 논문 20건만 검색되었다. 특이한 사실은 PubMed와 SCI Expanded에는 Scopus와 같은 계열의 회사인 Elsevier사에서 최근 발행한 *Cell* 121(1): 25-36 논문이 Scopus와 달리 아직 수록되지 않았다. 그리고 SCI Expanded에는 Scopus보다 3건의 가장 최신 논문이 더 많이 포함되어 최신성 측면에서 WoS가 Scopus보다 앞서고 있음을 보여주었다. 또한 SCI Expanded에는 Scopus와 중복되는 26종 중 11종에서 피인용 정보가 Scopus보다 더 많아 두 DB간에 연구 최전선에 있는 핵심 저널을 중심으로 피인용 정보의 횟수와 내용에 대한 최신성에 상당한 격차를 보였다.

따라서 피인용 정보가 검색가능한 Scopus와 WoS를 사용할 때는 DB의 최신성과 수록저널의 범위가 달라 피인용 횟수가 두 DB에서 상이할 수 있다는 가능성과 함께 Scopus에는 아무리 인용정보 링크의 정확률이 높고 소급하여 계속 인용정보를 구축하고 있다지만 1996년 이후 최근 10년분의 피인용 정보만이 현재 포함되었다는 사실을 반드시 알아야 될 필요가 있다. 또한 국내에서 만들어진 NDSL에 저자의 소속기관에 대한 검색이 불가능하여 비교할 수 없다는 사실이 안타깝다.

〈표 8〉 2005년도 상반기 주요 저널의 한국인 논문 분석결과

저널 및 논문	PubMed (피인용 불가)	Scopus (피인용 횟수)	SCI Expanded (피인용 횟수)
<i>Cell</i> , 120(2): 209-221.	Y	Y(1)	Y(1)
<i>Cell</i> , 120(3): 395-406.	Y	Y(1)	Y(2)
<i>Cell</i> , 121(1): 25-36.		Y(3)	
<i>Nat Biotechnol</i> , 23(1): 63-68.	Y	Y(2)	Y(3)
<i>Nat Biotechnol</i> , 23(5): 591-599.	Y		Y
<i>Nat Cell Biol</i> , 7(2): 137-147.	Y	Y(2)	Y(2)
<i>Nat Genet</i> , 37(3): 305-310.	Y	Y	Y(1)
<i>Nat Genet</i> , 37(5): 478-485.		Y	Y(1)
<i>Nat Immunol</i> , 6(2): 211-218.	Y	Y	Y(1)
<i>Nat Mater</i> , 4(2): 147-151.	Y	Y(2)	Y(2)
<i>Nat Mater</i> , 4(5): 383-387.		Y	Y
<i>Nat Mater</i> , 4(5): 399-402.	Y	Y	Y(1)
<i>Nat Med</i> , 11(4 Sup): S12-15.	Y	Y	Y
<i>Nat Rev Mol Cell Biol</i> , 6(5): 376-385.	Y	Y	Y
<i>Nat Rev Neurosci</i> , 6(6): 449-462.			Y
<i>Nat Struct Mol Biol</i> , 12(1): 38-45.	Y	Y(1)	Y(1)
<i>Nat Struct Mol Biol</i> , 12(5): 423-428.	Y	Y	Y(1)
<i>Nature</i> , 433(7025): 541-545.	Y	Y	Y(1)
<i>Nature</i> , 433(7027): 717-719.		Y(2)	Y(3)
<i>Nature</i> , 434(7033): 622-625.		Y	Y
<i>Nature</i> , 434(7033): 625-628.		Y(1)	Y(2)
<i>Nature</i> , 434(7034): 743-746.	Y	Y(1)	Y(1)
<i>Nature</i> , 434(7035): 921-926.	Y	Y	Y(2)
<i>Nature</i> , 434(7036): 980-986.		Y	Y(1)
<i>Nature</i> , 435(7038): 87-90.	Y	Y	Y
<i>Nature</i> , 435(7040): 347-353.	Y	Y	Y
<i>Nature</i> , 435(7040): 370-373.		Y	Y
<i>Science</i> , 308(5724): 1040-1042.		Y	Y
<i>Science</i> , 308(5727): 1442-1446.		Y	Y
<i>Science</i> , 308(5729): 1777-1783.	Y		Y(1)
<i>Science</i> , 309(5731): 121-125.	Y		

* 주) SCI Expanded의 2004년분이 4건과 비논문 1건은 비교에서 제외(조사일: 2005. 7. 14-21).

6. 결 론

지금까지 DB 품질평가에 관한 국내외 선행 연구 결과를 중심으로 학술 Web DB의 도입을 위한 품질평가기준을 정립하고, 그에 따라 쉽게 측정이 가능한 객관적 평가를 위한 입장에서 무료이용이 가능한 NDSL 및 PubMed와 상용 Scopus 및 WoS DB의 품질을 비교평가해 보았다.

비교대상 4개 Web DB의 품질평가기준별 주요 측정결과를 항목별로 정리하면 다음과 같다.

1) 최신성 : NDSL은 전체 데이터 대비 2005년도 데이터의 비중 1.5%는 물론, 최신년도인 2004년도와 2005년도 레코드 수가 DB내 전체 레코드 수에 대비한 비중 9.4%와 2000년 이후 레코드의 비중도 40.9%로 가장 높다. WoS에서는 일부 2004년도 레코드가 포함되기는 했지만 2005년도 최신 레코드가 가장 많고 2004년도 대비 2005년도 레코드의 비중도 39.2%로 가장 높다. 그러나 Scopus는 16.7%로 가장 낮고 2005년도 레코드 수도 상대적으로 적어 DB 구축의 신속성이 떨어지고 있다. 저널의 최신성(타임 래그)을 비교하기 위해 *BBRC*, *Nature*, *Science* 및 *Cell* 저널의 최신 이슈 수록상황을 조사해 본 결과 NDSL과 PubMed는 일단 색인은 하지 않고 출판사와 거의 동시에 서비스를 하고 있어 타임 래그가 거의 발생하지 않으며, Scopus와 WoS는 평균 3주 정

도의 타임 래그가 발생하고 있다.

2) 정확성 : 영국식과 미국식 영어의 구분은 Color와 Labor에 대해서는 PubMed에서만 어휘통제가 되어 동일한 결과가 나왔고, Encyclopedia에 대해서는 다른 DB와 마찬가지로 PubMed에서도 어휘통제가 되지 않았다. February 등 13개 검색어로 철자 에러를 검색해 본 결과 WoS의 철자 에러율이 4개 DB 중 양적인 측면과 비율에서 가장 낮게 나타나 WoS의 정확성이 가장 높은 반면, 최근에 구축된 NDSL과 Scopus는 외부 데이터를 도입하여 DB를 구축할 경우 데이터의 정확성을 위해 지속적인 DB 품질관리의 필요성이 높다.

3) 완전성 : 작성 기준 날짜가 제시되지 않아 아쉽지만 NDSL은 홈페이지에 'NDSL 데이터베이스 구축현황'을 제공하고 있고, PubMed와 Scopus는 공표된 레코드가 실제 레코드 수와 달라 홈페이지의 자료에 대한 신뢰성이 떨어진다. 저널 중복도와 고유성을 조사한 결과 4개 DB에 공통인 저널은 2,054종이고, 필드 명이 다르더라도 내용이 같은 4개 DB에 공통인 필드는 논문제목, 저자명, 초록 및 연도이다.

4) 포괄성 : 레코드의 총 건수는 수록년도의 차이를 감안하더라도 WoS가 가장 많이 수록하여 포괄성이 가장 높다. NDSL은 초록 보유율은 29.7%에 그치지만, 1년분이 완전히 수록된 2004년도 레코드 수와 최근 9년분인 1997년도 이후 2005년까지의 레코드 수

- 및 1985년 이후 레코드의 비중도 89.3%나 되어 가장 높다.
- 5) 일관성 : 저널 명 변경과 권호 형태가 불규칙적일 경우 저널 명 표기가 DB 별로 다양하며, Scopus에는 Medline 데이터와의 통합과정에서 중복 데이터의 제거 문제가 완벽히 해결되지 않았다. 대소문자 구분은 4개 DB 모두 구분하지 않고 있으며, 색인어의 일관성에 완벽한 어휘통제가 이루어진 DB는 없었다. NDSL과 Scopus는 단수와 복수를 모두 동일하게 처리했지만, PubMed와 WoS는 단수와 복수를 달리 취급하며, 대체로 NDSL 이외에는 'on line'과 'on line' 처럼 단어 사이의 공백과 대시()는 구분하지 않고 동일하게 처리하고 있어 검색결과가 동일하다.
- 6) 편의성 : Scopus와 WoS의 경우 검색결과 분석기능은 상용 DB의 고급화된 부가가치 기능이며, 자체 데이터와 PubMed의 기본이 되는 Medline 데이터를 동시에 검색할 수 있어서 이용자의 편의성과 검색 망라성을 높여준다. NDSL에는 Medline 데이터가 완벽하게 포함되지 않아서 의학 및 생명과학 분야 과거 자료의 품질과 데이터의 충실도가 떨어지며, WoS의 경우 Medline과 PubMed를 동시에 검색하려면 Medline과 PubMed의 차이점을 정확히 인식해야 정확한 선택이 가능하다.
- 7) 지원성 : NDSL 수록 저널의 39.7%가 원문에 링크되어 있어서 비율은 낮지만 링크된 전자 저널 종수로는 4개 DB 중 가장 많아 지원성이 높고, 논문 검색결과 화면에 소장 기관 보기와 원문신청하기 메뉴가 있어서 곧바로 국내 도서관에 우선적으로 DDS를 신청할 수 있어 편리하다. PubMed에서는 Medline 수록 4,714종 저널이 링크되어 있어서 전체 Medline 수록 저널의 약 98% 정도가 전자 저널 출판사 사이트와 직접 링크되어 있고, 검색결과를 'All'과 'Review'로 나누어 보여주므로 Review 형태의 자료만 구별하여 이용할 수 있어 편리하다.
- 8) 검색성 : NDSL에는 시소러스가 없지만, PubMed에는 MeSH가 있고 Scopus에는 MeSH와 Ei thesaurus는 물론 Emtree가 있다. 다른 DB와 달리 WoS는 검색 직전에 검색대상 데이터의 DB 입력년도를 선택하고 검색을 시작해야 한다. 생명과학 분야 주제어의 경우 각 DB마다 조금씩 차이를 보였지만, 전반적으로 NDSL은 저자명 검색에서 지극히 저조하고, 주제어를 입력하여 검색할 때 NDSL과 Scopus는 주제어 사이의 공백을 'AND' 연산자로 인식하여 처리하지만 WoS는 주제어 사이의 공백에 대해 구로 인식하고 있다.
- 9) 비용 : 현재 20여개 기관이 금년부터 시범 이용 중인 Scopus는 혁신적 시장진입과 안정적 평가기간을 제공하고 있지만 가격 인상의 우려가 있고, WoS의 경우 과거의 모든 데이터를 완전히 갖추려면 커다란 규모의 예산을 필요로 하여 국내의 많은 도서관에

서는 일부 연도의 데이터만 구입하여 불완전한 검색이 이루어지고 있다. 국내에서 널리 인지되고 많이 사용되는 NDSL과 PubMed 외에 저널의 중복도가 높은 Scopus나 WoS 등 다른 DB를 추가로 사용해야 한다면 획득 가능한 고유(유일)한 정보의 효과와 비용 간의 근거가 상용 DB 도입의 최대 관건이다.

- 10) 네트워크 : 4개 DB는 인터넷의 Web으로 제공이 가능한 학술 DB여서 네트워크 및 하드웨어 설치에 대한 부담은 없지만, 무료 PubMed에 비해 NDSL과 Scopus 및 WoS는 검색을 위한 접근성과 접근속도 및 접속방식에서 일반적으로 더 느리고 불편한 편이다.
- 11) 기타 사례조사 : 2005년도 상반기 한국인 저자의 Cell 등 주요 저널에 실린 논문을 분석해 보니 PubMed는 논문의 주도적 연구기관이 한국 기관인 경우만 검색이 가능하며 3개 DB 중 가장 적으며, SCI Expanded는 Scopus와 중복되는 26종 중 11종에서 피인용 횟수가 Scopus보다 더 많아 두 DB 간 연구 최전선의 핵심 저널을 중심으로 피인용 정보의 횟수와 내용에 대한 최신성에 상당한 격차를 보이고 있으므로 피인용 횟수가 두 DB에서 상이할 수 있다는 가능성과 함께 Scopus에는 1996년 이후 최근 10년분 피인용 정보만이 포함되었다는 사실을 반드시 알려야 될 필요가 있다. 이상과 같이 2개 무료 DB와 2개 상용 DB의

품질비교를 수행한 결과, 학술 Web DB 도입을 위한 품질평가를 실무적으로 수행할 목적으로 선행연구를 적용하기 위해서는 평가기준의 문제가 아니라 평가 방법의 적용이 문제였다. 또한 Scopus와 WoS DB 도입의 필요성은 인용정보를 수록하여 피인용 횟수와 주제의 선행 연구현황에 대한 검색이 가능하고 검색결과와 분석기능이 우수하다는 점과 무료 DB에 포함되지 않는 서지정보가 상용 DB에서 검색되는 정도와 필요성 등을 중심으로 자관의 예산 상황 대비 예상되는 효과성 정도에 따라 결정되어야만 한다. 따라서 학술 Web DB에 대한 정밀한 품질평가를 전제로 상용 DB 도입의 타당성을 검토하고, DB 품질평가에 근거한 도서관내 웹 자원으로 선정하여 도입하고, DB 품질평가 결과를 반영하여 이용자의 요구에 맞게 이용자교육 및 도서관정보 서비스가 이루어져야 효율적인 정보 서비스가 가능할 것이다.

본 연구는 국내 선행연구의 DB 품질평가기준을 중심으로 상용 DB 도입 검토를 위한 DB 평가에 표준적인 품질평가기준을 적용했다는 점에 의의를 둔다. 따라서 본 연구에 적용된 품질평가기준과 방법으로 다른 상용 DB의 품질평가 측정이 이루어진다면 조금 더 쉽고 객관적인 DB의 품질평가가 가능할 것이다. 하지만 본 연구의 부족한 점을 보완하기 위해서는 무료 DB를 사용할 수 있는 상황에서 상용 DB의 도입에 따른 비용/편의 분석과 비용소요 대비 가치에 대한 객관적 평가, 이용자의 DB 서비스 품질에 대한 만족도 조사에 의한 종합적인 품

질평가, 평가기준 사이의 상대적 중요도나 점수화에 대한 방안의 보완, 데이터의 정확률이나 재현율과 같은 고도의 복잡한 품질 분석이 더 필요할 것이다.

참고문헌

- 김기오, 2003. 『국내외 의학관련 웹사이트의 KoreaMed와 PubMed 인지도 비교 연구』. 석사학위논문, 숙명여자대학교 대학원.
- 김기홍, 2000. 『CD ROM MEDLINE과 PubMed의 검색 효율성에 대한 비교 연구』. 석사학위논문, 성균관대학교 대학원.
- 김상준, 2002. 문헌정보학 실무자와 실무중심의 연구활동. 『한국의학도서관』, 29(12): 12-25.
- 김상준, 2004. 의학 분야 Web DB의 품질평가: PubMed와 Embase를 대상으로. 『한국문헌정보학회지』, 38(2): 161-187.
- 김석영, 2002. 과학기술분야 해외 온라인 데이터베이스 평가 연구. 『정보관리연구』, 33(1): 61-80.
- 김성희, 2003. 데이터베이스 평가 모델 개발을 위한 해외사례조사연구. 『국회도서관보』, 40(2): 15-33.
- 문성빈, 최인숙, 최상희, 이성숙, 장혜영, 2002. 과학기술전문정보 서비스품질평가 및 이용자 정보요구 분석. 『정보관리연구』, 33(1): 31-47.
- 안계성, 함정훈, 2000. 『데이터베이스 품질평가 항목』. 서울: 한국데이터베이스진흥센터.
- 연구개발정보센터, 1996. 『DB 선정 및 품질평가기준에 관한 연구』. 과천: 과학기술처.
- 유동숙, 1998. 『의약정보관련 서지사항 데이터베이스의 비교평가』. 석사학위논문, 숙명여자대학교 임상약학대학원.
- 유사라, 1999. 데이터베이스 정보 품질평가의 매타분석. 『정보관리학회지』, 16(1): 157-174.
- 유사라, 2000. Web DB 정보 서비스평가를 위한 기준 측정지표 분석 I. 『한국문헌정보학회지』, 34(3): 25-40.
- 유사라, 2002. 국가과학기술전자도서관 이용자 정보요구와 이용 형태 분석. 『한국문헌정보학회지』, 36(1): 133-156.
- 유재복, 2004. 새로운 인용색인 웹 데이터베이스: Scopus. 『STIMA Bulletin』, 1(창간호): 26-32.
- 유재복, 2005. 웹 인용색인 데이터베이스 Scopus 선정 학술지의 신뢰성 분석: Web of Science와의 비교를 중심으로. 『STIMA Bulletin』, 2(통권2호): 46-59.
- 윤봉자, 1994. 문헌검색도구를 이용한 한국인의 의학연구업적 평가: Medline 수록

- 논문 중심으로. 『한국의학도서관』, 21(2): 75-83.
- 이국희, 최창선, 임남홍, 김호중, 박옥순, 최민영. 1995. 『데이터베이스 품질평가에 관한 연구』. 서울: 한국데이터베이스진흥센터.
- 이만수, 이상열, 신인수. 2003. 서지 데이터베이스의 품질평가에 관한 연구. 『정보관리연구』, 34(3): 21-39.
- 이명희. 1998. ASKERIC 데이터베이스의 품질에 관한 연구. 『한국비블리아』, 9: 221-240.
- 이용봉, 조현양, 류범중, 최재황. 2001. 과학기술분야 데이터베이스의 품질향상을 위한 품질평가 연구. 『한국문헌정보학회지』, 35(2): 109-132.
- 이용봉. 1996. 데이터베이스 품질에 관한 비평적 평가. 『국회도서관보』, 33(4): 13-28.
- 이용봉. 1999. CD-ROM 데이터베이스의 품질평가 및 개선방안에 관한 연구. 『한국문헌정보학회지』, 33(4): 29-46.
- 이제환. 1996. 이용자의 관점에서 본 KRISTAL System과 DB의 품질. 『KOSTI Proceedings』, 1: 240-267.
- 이제환. 1997. 과학기술분야 서지 DB의 품질관리 및 평가방안: KORDIC의 KRISTAL 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.
- 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 평가 모델 개발에 관한 연구』. 서울: 한국교육학술정보원.
- 조순영, 한혜영, 배영천, 박연희, 박동섭. 2003. 『국내 대학도서관의 해외 전자정보 구독 현황 분석 보고서』. 서울: 한국교육학술정보원.
- 한국데이터베이스진흥센터. 1996. 『데이터베이스 표준화연구보고서』. 서울: 한국데이터베이스진흥센터.
- 한국데이터베이스진흥센터. 2003. 『데이터베이스 이용실태 및 정보수요조사 보고서』. 서울: 한국데이터베이스진흥센터.
- 함정훈. 2001. 데이터베이스 품질평가기준 연구. 『Digital Contents』, 2001년 01월

- 호: 60-73.
- 홍현진, 정혜경, 노영희, 이미영. 2005. 해외 비영리기관 소장 학술 데이터베이스 현황조사 및 분석 연구. 『정보관리학회지』, 22(1): 87-104.
- 홍현진. 2005. 웹 기반 데이터베이스의 품질평가 기준 개발에 관한 연구. 『한국문헌정보학회지』, 39(2): 211-235.
- Adachi, Y. 2004. "A new navigational service in STM scholarly information: Scopus." *J. Info. Process. Manag.*, 47(8): 558-562.
- Cooper, M. D. 2001. "Usage patterns of a web based library catalog." *JASIST*, 52(2): 137-148.
- Fingerman, S. 2005. "Scopus: profusion and confusion." *Online*, 29(2): 36-38.
- 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.
- Jacso, P. 2004. "Citation searching." *Online Info. Rev.*, 28(6): 454-460.
- Jacso, P. 2004. "ISI Web of Science, Scopus, and SPORTDiscus." *Online*, 28(6): 51-54.
- Jain, N. C. 2005. "Scopus has wider scope than Science Citation Index." *Curr. Sci*, 88(3): 331-331.
- 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.
- LaGuardia, C. 2005. "Scopus vs. Web of Science." *Library J.*, 130(1): 40-43.
- 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.
- 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.
- Sevinc, A. 2004. "Web of Science: a unique method of cited reference searching." *J. Nat. Med. Assoc.*, 96(7): 980-983.

- 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.
- Voutier, C. 2004. "Web of Science: index not as useful as it appears." *J. Nat. Med. Assoc.*, 96(9): 1240-1241.
- 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. Trewheellar. 1998. "Medline and Embase complement each other in literature searches." *Brit. Med. J.*, 316: 1166.