

인터넷 특허정보 데이터베이스의 탐색기능에 관한 연구

A Study on the Search Functions of Internet Patent Information Databases

최경화 (Kyung-Hwa Choi), 이란주(Lanju Lee)

동덕여자대학교 문헌정보학과

Dept. of Library and Information Science, Dongduk Women's Univ.

본연구의 목적은 4개의 인터넷 특허정보 데이터베이스의 탐색기능을 비교분석 함으로서 각 데이터베이스의 활용도를 제시하고 더 나아가 인터넷 상의 특허정보 데이터베이스의 이용자 우호적인 인터페이스 설계를 위한 기초 자료로 사용하려고 한다. 연구방법으로는 비교적 인지도가 높은 국내의 Patrom과 국외의 USPTO, QPAT-US, IBM을 선택하여 데이터베이스의 수록항목, 주요항목, 검색방법, 멀티미디어 검색기능, 웹으로 지원 가능한 정보자원의 형태를 중심으로 비교분석한다.

1.서론

세계각국의 경쟁이 날로 치열해져 가는 가운데 현대의 기업은 기술개발과 기술혁신의 전 과정에 걸쳐서 특허와 관련된 정보의 요구가 급속히 증가하고 다양화되고 있다. 이에 특허정보에 대한 신속하고 정확한 조사가 필수요건이 되고 있다. 한편 데이터베이스에 수록되는 여러 가지 정보 중에서 특허정보는 특허정보 발생량이 매우 많고 구성이 복잡하여 여러기관과 국가에서 다양한 종류의 데이터베이스를 제공하고 있다. 그런 가운데 우수한 기술경쟁력을 보장해 주는 특허의 중요성이 점차로 증가되고 있다. 탐색기능 평가에 관한 선행연구로 국내에서도 검색엔진을 이용한 연구들이 발표되었으나 (정영미, 김성은, 1997;

이란주, 1997; 이란주, 최경화;1997), 아직까지 인터넷 특허정보 데이터베이스에 대한 효율적인 검색을 도울 수 있는 연구는 없었다.

따라서 본연구의 목적은 인터넷 특허정보 데이터베이스의 탐색기능을 비교분석 함으로서 각 데이터베이스의 활용도를 제시하고 더 나아가 인터넷 상의 특허정보 데이터베이스의 이용자 우호적인 인터페이스 설계를 위한 기초 자료로 사용하려고 한다. 연구방법으로는 비교적 인지도가 높은 국내의 Patrom과 국외의 USPTO, QPAT-US, IBM을 선택하여 데이터베이스의 수록항목, 주요항목, 검색방법, 멀티미디어 검색기능, 웹으로 지원 가능한 정보자원의 형태를 중심으로 비교분석한다.

2.데이터베이스의 탐색기능 비교분석

데이터베이스의 탐색기능을 비교분석하기 위

하여 전통적인 데이터베이스와 CD-ROM 데이터베이스 품질 평가 요소 가운데 인터넷 데이터베이스에 적용될 수 있는 것들로 선정하였다; (1) 데이터베이스의 수록항목 (2) 데이터베이스의 주요항목 (3) 데이터베이스의 검색방법 (4) 멀티미디어 검색 가능여부 (5) 웹으로 지원 가능한 정보자원의 형태.

2.1 데이터베이스의 수록항목 비교분석

특허정보의 검색에 있어서 중요한 것은 각 데이터베이스 내에서 초록 수록 유무와 전문 제공여부를 확인하는 것이다. 또한 데이터베이스가 어떤 항목들을 수록하고 있는지를 파악하여 검색 가능한 것과 불가능한 것을 판단해야 한다. 특히 검색시 중요한 항목들은 서지사항, 청구범위, 초록, 도면, 전문 명세서 등이 있다. 또한 이처럼 데이터베이스에 수록되어 있는 정보의 포괄성은 데이터베이스 품질을 내용적으로 평가할 때 중요한 요소가 된다(이응봉, 1996). IBM의 경우 상용온라인 데이터베이스에서 제공되는 정보로서 인터넷에서는 이제 막 서비스되기 시작하고 이용자에게 중요하다고 할 수 있는 이미지 정보가 제한적으로 제공되고 있다. 그러나 아직 전문은 제공하지 못하고 있는 실정이다. 한편 Patrom의 경우에는 특허정보 검색을 자주 하는 경우 전문이 필요할 때마다 요금을 지불하고 요청하는 것보다는 CD-ROM을 구입하면 색인과 C&I CD-ROM을 통하여 이미지와 연결해 전문명세서를 볼 수 있는 장점이 있다.

2.2 데이터베이스의 주요항목 비교분석

특허정보의 경우에는 특히 수록정보의 소급성이 중요한데 수록범위가 가장 넓은 것은 IBM이며 가장 최근에 생긴 Patrom이 가장 적은 량의 데이터를 가지고 있다. 한편 정보가 발생되어 데이터베이스에 수록된 후 검색되기까지의 시간의 폭 즉, 데이터의 갱신주기는 짧으면 짧을수록 데이터베이스의 품질은 높다고 할 수 있다. 그러나 소급탐색이 필요한 특허정보 데이터베이스의 경우 초록과 같은 부가 정보의 부여를 위해 다소간의 시간지체가 있을 수 있는데 부가 정보의 품질을 손상시키지 않는 범위내에서 갱신주기는 가능하

면 짧은 것이 좋다. 주목할 점은 QPAT-US의 경우 검색식Setting이 가능하다는 점인데 이는 일반적으로 상용데이터베이스에서 적용하는 방법으로 수행한 검색식마다 따로 저장해 두는 번호를 말하며 이를 "Set"라고 말한다. 이러한 Set형태는 일반적으로 인터넷 데이터베이스에서는 활용하지 않는 방법인데 QPAT-US는 상용 온라인 데이터뱅크인 Questel-Orbit에서 운영하고 있어 동일한 검색방법을 웹에서 구현한 것으로 앞으로 이와 같은 형태의 검색법을 지원하는 데이터베이스가 많이 개발될 것으로 기대된다. 이러한 Set의 가장 큰 장점은 Set간 상호 연산자 검색이 가능하다는 점을 들 수 있다.

2.3 데이터베이스의 검색방법 비교분석

(1) Patrom(<http://patrom.dwe.co.kr/index1/html>)

Patrom에서는 and와 or의 부울린 검색과 인접 연산자 사용 및 절단검색 그리고 필드제한 검색 등을 지원하고 있다. 검색하기 위해서는 먼저 출력양식을 선택하고 데이터베이스 선택을 통하여 연도를 선택할 수 있으며, 권리구분으로 특허나 실용신안 또는 전체(둘다)를 선택할 수 있다. 그리고 부울린 연산자와 인접 연산자를 이용하여 검색어를 지정하고 출력건수를 선택한다. 사용자 등록을 하여 중량제를 이용할 경우에는 화면에 보여지는 리스트 한건당 5원의 요금이 부과되므로 출력건수를 적절히 조절해야 한다. 또한 검색된 결과의 리스트에서 서지사항과 요약정보를 볼 수 있고 그 화면에서 공고 및 등록여부를 찾아볼 수 있다. 마지막으로 원문을 보기 위해서는 이미지 CD-ROM을 구입해야 한다.

(2) USPTO (<http://patents.uspto.gov>)

미국 특허청에서 제공하는 초록검색 데이터베이스는 부울린검색과 고급검색, 특허번호 검색, 미국특허분류 데이터베이스의 네가지 경우로 나누어 서비스되고 있다. 먼저 USPTO에서의 부울린검색은 검색어가 2개 이하일 때 쉽게 검색할 수 있는 것으로 검색어와 검색 필드만 알고 있으면 되므로 초보자도 쉽게 할 수 있다는 장점을 갖고 있다. 그러나 검색어

3개 이상을 가지고 고급검색은 할 수 없는 단점을 가지고 있다.

다음으로 고급검색은 데이터베이스 검색에 익숙해진 이용자에게 만족할 만한 결과를 가져올 수 있는 유용한 도구로서 검색하고자 하는 거의 모든 종류의 검색이 가능하다는 장점을 갖는 반면, 초보자가 사용하기 위해서는 약간의 암기사항과 유의사항들이 있어 숙련된 사용을 위해서는 시간과 노력이 필요하다. 부울린검색에서 사용하는 연산자를 비롯하여 절단검색과 구절검색이 가능하고 다양한 필드검색을 제한해서 할 수 있다. 그러나 부울린검색과 다른 점은 검색창이 하나로 구성되어 있다는 점이다. 그리고 부울린검색을 사용할 때는 필드검색을 메뉴상에서 선택할 수 있으나 고급검색에서는 직접 입력해야 한다. 한편 결과 순위 (Rank Results) 부분에서 연대순 정렬 (Chronologically) 과 적합성에 따른 정렬 (By Relevance)이 추가된다는 것들을 들 수 있다. 마지막으로 미국 특허분류 데이터베이스에서는 미국 특허청이 자체적으로 운영하고 있는 미국식 분류를 통한 자료검색이 가능한데 IPC 체계를 따르는 우리로서는 큰 효용을 바랄 수는 없지만, 미국식 분류를 검토함으로써 모호한 기술내용을 명확히 파악할 수 있고, 검색된 문헌에 대한 불명료한 부분을 보완할 수 있는 효과가 있다고 할 수 있다.

(3) QPAT-US(<http://qp2.qpat.com/index.html>)

QPAT-US는 상용 데이터베이스의 검색법을 그대로 가져왔기 때문에 다양한 필드를 이용한 정밀한 검색이 가능하다. 즉 main search page에서의 "Field Tag"에서 필드 검색법을 통하여 자세한 검색법에 대한 설명을 볼 수 있고, "Search Operators"에서는 연산자 사용법을, "Search Examples"에서는 검색에 대한 상세한 자료를 얻을 수 있다. 검색상 QPAT-US의 가장 특징적인 면은 대부분의 인터넷 특허정보 데이터베이스들이 지원하지 못하는 검색식 Setting이 이루어진다는 점으로 비록 한정되기는 하지만 검색전략에 도움을 줄 것이며 아직까지는 대부분의 상용데이터베이스

와 CD-ROM 데이터베이스에서만 지원해 주고 있는 기능 가운데 하나라고 할 수 있다.

기본적인 부울린 연산자와 함께 USPTO 데이터베이스에서 지원하지 못하는 인접 연산자들을 지원하고 있으며, 구절검색과 절단검색 그리고 다양한 필드제한 검색이 가능하다. 특히 필드제한 검색시에는 Basic Supergroup, Quick Supergroup, 그리고 이밖에도 연도 설정에 있어 <, > 등을 사용하여 범위를 설정하거나 "field=num(field=1998)" 설정으로 해당년도만을 선택할 수도 있다.

QPAT-US는 상용온라인 서비스인 Quest-Orbit의 검색방법을 그대로 적용한 것으로서 "main search page"에서 검색을 할 수 있고 Default Field와 Date Range는 메뉴를 선택할 수 있다. 그러나 연산자 기호나 필드 지정 등은 검색자가 지정해 주어야 하므로 초보자보다는 필드의 tag 등 검색법에 익숙한 숙련자에게 더욱 유용하다.

(4) IBM (<http://patent.womplex.ibm.com>)

IBM은 본 논문에서 언급되는 특허정보 데이터베이스 중 가장 많은 검색방법을 가지고 있다. 홈페이지 처음 부분에서 볼 수 있는 Simple text search로부터 시작해서 and/or/and not을 이용하는 Boolean text search, 특허번호로 검색하는 Patent number search, 부울린 연산자를 비롯하여 인접 연산자와 절단검색 그리고 필드검색이 가능한 Text search 등으로 나뉘어져 있다. 연도별 검색에 있어서 현재는 "1995-현재"와 "1971-현재"의 두가지 옵션만이 가능하지만 앞으로 각 연도별 검색과 출간 날짜별 검색이 추가될 예정이다.

특허정보 이미지를 보기 위해서는 "view image"를 클릭하므로써 1974년에서부터 1997년까지의 것들을 볼 수 있다. 그러나 어떤 특허정보의 경우에는 이미지를 담는데 있어서의 어려움 때문에 누락되기도 하지만 계속적으로 추가될 예정이다.

2.4 멀티미디어 검색가능

멀티미디어 데이터베이스는 기존의 텍스트 위주의 데이터 저장에서 탈피하여 화상, 동영상, 음성 등의 멀티미디어 정보를 저장, 검색,

편집할 수 있는 것으로 이러한 멀티미디어 이용자 인터페이스는 데이터 정의와 데이터 조작시 멀티미디어 정보를 위한 추가적인 기능이 제공되어야 한다. 데이터의 검색시 기존의 문자나 숫자위주의 데이터베이스에 대한 검색은 내용 기반으로 하는 것이 효율적이었으나 멀티미디어 데이터의 경우는 브라우징 방법이 효율적이며 질의와 브라우징을 통합적으로 제공하는 인터페이스의 필요성이 대두되고 있다(최중민, 1996). 그러나 아직까지 인터넷 특허 정보 데이터베이스들은 대부분이 문자와 텍스트위주로 구성되어 있으며 특허정보에서 중요하게 평가되고 있는 이미지 또한 IBM을 제외하고 아직까지 지원하지 못하고 있다.

2.5 웹으로 지원 가능한 정보자원의 형태

특허정보를 웹상에서 검색하는데 있어서 웹으로 지원가능한 정보자원의 형태는 원문정보가 될 것이다. 그러나 웹으로 원문을 지원하는 데이터베이스는 아직까지는 CD-ROM과 연동한 국내의 Patrom 뿐이다. 나머지 데이터베이스는 원문제공 신청을 별도로 해야만 제공 받을 수 있고 한 번 요구 할 때마다 일정액을 지불해야 한다.

앞으로는 다중플랫폼을 지원할 수 있고, 여러 종류의 데이터베이스를 탑재할 수 있으며, 데이터베이스를 선택적으로 브라우저 상에서 선택할 수 있는 웹의 특성을 최대한 이용하는 탐색기능들의 추가적인 기능들이 필요하다. 이것은 또한 인터넷 특허정보 데이터베이스를 이용하는 또다른 의의가 될 것이다.

3. 결론

일반적으로 이용자들이 특허정보 데이터베이스를 이용할 때 가장 많이 사용하고 또 검색결과에 많은 영향을 미치는 탐색기능에 있어서, 수록항목은 초록과 서지사항을 비롯하여 4개의 데이터베이스가 비슷한 수준을 보여주고 있다. 주요항목간 비교에서는 서로 다른 특징들을 보여 주었는데 특허정보는 특히 소급탐색이 중요하므로 수록범위가 중요한데 가장 망라적인 데이터베이스는 1971년부터 데이터를 가지고 있는 IBM이며 갱신주기 또한 수시로 이루어지고 있다. 한편 검색방법에 있어

서는 연산자 사용과 필드제한 검색을 비롯한 절단검색 등은 4개의 데이터베이스 모두가 지원하고 있다.

그러나 인터넷 데이터베이스의 특징적인 면을 고려한 멀티미디어 자료나 웹을 이용한 정보자료의 활용은 아직까지 제대로 이용되지 못하고 있다. 국내의 Patrom만이 원문 CD-ROM을 웹에서 이용할 수 있는 것을 제외하고 나머지 데이터베이스들은 이 기능을 제대로 활용하지 못하고 있다. 또한 아직까지는 상용 온라인 데이터베이스에서 지원하는 고급 검색 등은 완전히 지원하지 못하고 있는데, Quest-Orbit사에서 서비스하고 있는 QPAT-US는 상용 온라인 서비스에서 제공하고 있는 검색식 Setting을 지원하여 Set간 상호 연산자 검색이 가능하도록 하고 있다. 검색법에 있어서는 IBM이 네가지 방식의 가장 다양한 형식을 지원하며, 검색에 능숙한 이용자에게는 QPAT-US가 가장 적절하고 초보자에게는 Patrom과 USPTO가 가장 친절한 검색법을 보여주고 있다.

본 연구결과가 인터넷 특허정보 데이터베이스의 보다 나은 활용에 기여하기를 바란다. 후속연구로 탐색기능 비교분석 연구 결과를 토대로 인터넷 특허정보 데이터베이스의 효과적인 사용을 위한 이용자 우호적인 인터페이스를 고려한 데이터베이스 설계모형을 제공하기 위한 연구를 발표할 예정이다.

(참고문헌)

- 이란주, 1997. "월드 와이드 웹 (World Wide Web) 검색엔진의 효율적인 검색방법 고찰 - Lycos, InfoSeek, Alta Vista, Excite, Yahoo를 중심으로". 도서관 52(2): 29-43.
- 이란주, 최경화, 1997. "국내 웹검색도구의 특성 및 탐색기능 평가에 관한 연구". 한국문헌정보학회지 13(3): 75-108.
- 이용봉, 1996. "데이터베이스의 품질에 관한 비평적 평가", 국회도서관보, 제33권 4호.
- 최중민, 1996. "데이터베이스와 사용자 인터페이스" 데이터베이스월드. 10월호