

# 차세대 도서관 목록<sup>1)</sup>

(Next Generation Library Catalog)



심 경

shim@irisnet.co.kr

정보학박사

한국도서관협회 평생회원

(주)아이리스넷 대표

예전에 문헌정보학이 도서관학이라고 불리던 시절이 있었다. 이른바 신세대 사서들에게는 어쩌면 “Once upon a time…” 처럼 오래 전 이야기로 들릴지도 모르겠지만, 필자가 대학 3학년 때쯤 “정보학”이라는 과목이 국내에 처음 소개되었고 우리는 한번도 구경한 적이 없는 컴퓨터를 이용한 시스템, COBOL프로그램언어, 네트워크 등에 대해 배우게 되었다. 한없이 낯설기만 하고, 그래서 웬지 우리가 침범(?)해서는 안될 것 같은 영역을 공부하면서, 그래도 우수한 성적을 기대했으나 결과는 “B밖에 C를 뿌렸을 뿐”이었다. 성적은 하늘이 내리시는 것이라고 우리는 말했지만, 아무튼 고민은 도대체 도서관학과 정보학이 뭐가 다르냐는 문제였다. 그런데 대학 졸업 후 10여 년이 지나서 정보학과 비교하여 “library science is institution-oriented”라는 말을 우연히 듣고 “아이”라는 대오각성을 하였다. 그 기쁨도 잠시, 전자도서관의 출현과 더불어 “벽 없는 도서관(library without walls)”이라는 표현이 등장하면서 뭔가 다시 원점으로 돌아온 느낌이다. 기존 도서관학 또는 도서관서비스가 도서관이라는 “기관” 또는 “건물”에 한정되어 있었다면 지금은 문자 그대로 그 벽의 한계를 넘어가고 있다.

1) 이 글은 Breeding, M. (2007). Next generation library catalogs. *Library Technology Reports*, 43(3)을 많이 참조하였음을 알립니다.

차세대 도서관 목록도 그런 움직임 중 하나이며 이들 중 많은 부분은 우리가 이미 알고 있거나 국내 시스템에는 구현이 되어있는 것도 있다. 하지만 이 글에서는 차세대 도서관 목록이 무엇이며 왜 시작되었는가를 알아보고, 그들의 기능적 요소가 무엇인가를 살펴봄으로써 단편적 사례가 아닌 전체적인 “흐름”을 개관하여 향후 방향성 확립에 도움이 되었으면 한다.

## 차세대 도서관 목록이란 무엇인가?

차세대 도서관 목록은 next generation catalog 또는 next generation library catalog라고 영어로 표현된다. 이 명칭은 영어로 보나 한글로 보나 오해의 소지가 있다. 그 명칭 중 “차세대(next generation)”라고 하면 “새로운 것을 미래에 개발하자”는 것으로 인식될 수 있으나 뒤에서 설명하겠지만 사실은 기존 도서관 목록의 모델을 승계하면서 발전시키는 것을 의미하며 그에 적용되는 많은 기술은 이미 존재하고 있다. 또한 “도서관 목록(library catalog)”이라는 부분도 서지레코드의 품질 또는 포맷과 같은 목록레코드와의 관련성을 연상시키지만 여기서는 Online Public Access Catalog(OPAC)의 목록(catalog)을 의미한다. 짧게 말하면 이는 국내 도서관자동화시스템에서 검색모듈 또는 검색인터페이스라고 칭하는 부분이다. 그래서 브리딩(Breeding)은 이를 new generation of library interfaces 또는 next generation of library interfaces라는 표현이 의미전달에 더 명확하다고 하였다. 실제로 차세대 도서관 목록은 서지 목록이 아니라 인터페이스이고 툴이며 기술적 측면으로는 데이터베이스와 색인의 조합이라고 할 수 있다.

그런데 차세대 도서관 목록에 대하여 모두가 공감하는 단일 정의가 있는 것은 아니다. 다만 일반 웹 검색은 빠르고 쉬우며 즉각적인데 반하여 도서관 시스템은 대상자료가 다분히 책 위주로 구성되며, 인터페이스는 혼란스럽고, 이용하기 어려우며, 시간이 소요되고, 불편하므로, 다양한 웹 경험으로 변화하는 이용자 기대치를 고려하여 만족시키자는 것이다.

## 왜 차세대 도서관 목록이 필요한가?

요즘 도서관 이용자들은 웹에 익숙하여 예전과는 달리 현저하게 다른 서비스를 요구하므로 도서관 목록이나 도서관에서 제공되는 그 이외의 인터페이스도 웹 상에 서비스되는 인기검색 사이트와 동일한 스타일과 정교함을 가져야 한다는 것이다. 특히 기존 도서관 목록(legacy library catalog)은 다음과 같은 한계를 가지고 있으므로 이를 발전시켜야 한다는 것이다.

- 만족할 만큼 직관적이지 못하며 복잡한 검색인터페이스
- 널리 쓰이는 이용자 인터페이스 컨벤션<sup>2)</sup>과 일관성 결여
- 검색결과와 적합성 또는 이용자 관심도에 따른 정렬 불가능
- 대상자료 범위의 제한성
- 인쇄자료에 집중되어 전자 콘텐츠 취급 미흡
- 온라인 콘텐츠를 이용자에게 제공 불가능
- 도서관 이용자가 참여할 사회 네트워크(social network) 기능 부족

위에 기술된 기존 도서관 목록의 단점은 인터넷 상 웹 사이트들의 기능을 기준으로 보았을 때의 이야기이다. 그러나 이를 다른 측면에서 보면 기존 도서관 목록의 복잡한 검색인터페이스는 정교한 검색결과를 제공하며 도서관 목록은 태생이 소장자료에 대한 발견 및 위치정보 제공을 목적으로 하였지 최근 급속하게 확대되는 디지털 매체(born digital)를 염두에 둔 것은 아니었다. 따라서 대상자료의 범위가 제한적일 수 밖에 없으며 자료의 원문을 온라인 상에서 제공하지 못했지만 소장자료에 대하여는 아주 정교한 기능성을 제공한다. 예를 들면 이용자가 로그인하여 자신이 대출한 자료 리스트에서 반납예정일을 확인하고 대출갱신을 하며 연체료를 지불하고, 이외에도 다양한 요청을 할 수 있는 기능이 그들이다.

따라서 차세대 도서관 목록은 기존 도서관 목록의 장점과 일반 웹에서 제공되는 새로운 추세를 통합하여 변화하는 이용자 요구수준에 근접하려는 의도로 시작되었다.

### 차세대 도서관 목록의 기능 요소

차세대 도서관 목록이 추구하는 바는 현재 도서관 목록의 부족한 부분을 개선하는데 있으며 향상되어야 할 기능성으로 지적되는 것은 다음과 같은 점이다.

- **대상자료 범위 확장:** 기존 도서관 소장자료의 서지단위 제공이라는 한계를 넘어 소장자료의 서지단위 검색을 세분하여 기사단위의 검색을 제공하고, 자료 원문의 온라인 디스플레이를 가능하게 하며, 사진, 필사본, 지역신문, 족보자료 등 자관 디지털 도서관 장서로부터 콘텐츠 검색과 디스플레이, 그리고 기관 리포지터리의 콘텐츠는 물론 상용 데이터베이스 및 웹 자료 등 제공 데이터 범위를 넓혀야 한다는 것이다.
- **통합검색기능:** 기존 도서관 목록 이외에 전자자원 컬렉션을 제공하는 각종 데이터베이스의 증가로 이용자는 원하는 자료를 어디서 찾아야 하는지 혼란스러워 이들에 대한 통합검색(federated search)이 필수이다.

2) 인터페이스 컨벤션(interface convention)이란 컴퓨터 인터페이스에서 일반적으로 널리 통용되고 받아들여지는 표준이나 규범 등을 일컫는다. 예를 들어, Windows에서 cut-&-paste와 같은 것을 컨벤션이라고 한다.

- **콘텐츠 제공:** 이는 해당 도서관이 전자자료의 원문을 무료, 도서관 구독, 또는 pay-per-view등을 통하여 접근할 수 있을 경우에 해당되기는 하지만, 서지검색 톨의 제공 이외에 자료의 원문을 온라인 상에서 직접 이용자가 볼 수 있도록 제공할 필요가 있다는 것이다. 이를 위하여 링크시스템을 통합검색시스템과 연동하기도 한다.
- **통합 아니면 분산:** 과거 도서관자동화시스템(Integrated Library System: ILS)의 발전은 통합시스템이라는 개념으로 이루어졌다. ILS는 목록, 수서, 연속간행물, 대출 모듈 등이 내부적으로 잘 통합되어 있으나 그 시스템을 벗어나서 다른 시스템과 통합할 수 있는 표준 프로토콜을 가지고 개발되지 않았다. 예를 들면, 전자자원관리시스템(Electronic Resource Management System: ERMS)이나 통합검색시스템과 OpenURL 링크, 그리고 다른 유틸리티 등은 ILS와는 독립적으로 운용된다. 그 이유는 시스템들 간 데이터 교환을 위한 표준 프로토콜이 없으며<sup>3)</sup> 자동화 업체들 간의 경쟁으로 발생하는 것이다. 다양한 시스템과 데이터베이스처럼 분산된 대상을 도서관이 자체적으로 통합을 하는 것은 고비용이 소요되지만 방법이 없지는 않다. 한 가지 방법은 검색엔진에 ILS를 비롯한 다른 시스템 및 데이터베이스의 데이터를 반입하여 통합인터페이스를 제공하는 것이다. 하지만 이는 데이터 반출, 반출 데이터 포맷의 재구성, 로딩 및 색인 추출 과정의 부담이 크며 특히 ILS의 대출상태와 같이 수시로 변화하는 모든 데이터의 동기화<sup>4)</sup>가 기술적으로 단순하지만은 않다. 두 번째 방법은 ILS 등의 모든 데이터를 검색엔진에 색인으로 구축하는 것이 아니라 일부 검색에 필요한 부분을 색인하고 나머지 정보는 이용자의 요구가 있을 때 데이터베이스에 연결하여 불러오는 것이다. 이는 국내 도서관자동화시스템에서 일반적으로 사용되는 방식으로 효과적이다.

위에 기술한 내용은 결국 차세대 도서관 목록을 위한 비전이라고 할 수 있다. 이러한 넓은 관점에서 현재 도서관 목록인터페이스가 가진 불합리성을 수정하는 방법은 일반적으로 검색대상자료의 영역 확장, 보다 현대적인 검색인터페이스 기법과 검색결과를 순위화하는 검색엔진의 채용, 그리고 원문 또는 모든 형태의 미디어까지 이용자 PC에서 보여주기 등이 주류를 이루나 그 핵심은 도서관이 제공하는 모든 정보에 대한 단일 접근인터페이스가 있어야 한다는 것이다. 이러한 차세대 도서관 목록의 기능을 세분하여 살펴보자.

- (1) **풍부한 콘텐츠(Enriched Content):** 기존 도서관 목록은 텍스트 기반 디스플레이 위주이므로 시각적 내용을 제시하고, 이용자에게 제공되는 정보의 양을 증가시키기 위하여 다른 소스에서 콘텐츠를 가져오는 것이 필요하다. ILS로부터 제공되는 기본적인 서지정보와 혼합될 수 있는 콘텐츠의 범주는

3) 최근 국내 도서관자동화시스템도 XML기반 Web Services를 통한 표준 프로토콜을 이용하여 타 시스템과 연계 및 통합을 지원하는 것이 있으나 이는 표현대로 "최근"의 일이다.

4) 검색엔진을 별도로 사용하는 것은 데이터베이스의 데이터를 검색엔진이 색인하여 이용자 검색에 반영하는 것이다. 그러나 데이터베이스는 수시로 신규레코드가 생성되고 수정되며 삭제되는데 이를 검색엔진이 실시간으로 반영하는 것은 말처럼 쉬운 일은 아니다. 일반적으로 검색엔진과 데이터베이스 간의 동기화는 이용자가 줄어드는 새벽시간에 이루어진다.

자료의 표지 이미지 (북자켓, 영화 케이스 또는 저자를 나타내는 다른 시각적 표현을 포함하는 표지 이미지), 목차정보, 요약, 리뷰 등이 있다. 물론 이는 새로운 아이디어가 아니며 몇 년 전부터 이어온 추세이나 점점 더 많은 비율의 책과 도서관의 다른 자료들이 보강된 콘텐츠를 필요로 한다는 것이다. 외국의 경우 이를 전문으로 제공하는 회사가 있으며<sup>5)</sup> 국내에도 존재한다. 그러나 벤더로부터 제공되는 데이터는 그 건수보다는 품질과 저작권 문제에 대한 확인이 필수조건이라 하겠다.

- (2) 패싯<sup>6)</sup> 네비게이션(Faceted Navigation): 단순 검색결과와 제시가 아니라 검색결과 내에서 개별 데이터가 가진 “측면”을 활용하여 네비게이션 또는 “결과 내 재정렬/재검색”이 가능하도록 하는 것이다 (그림 1 참조). MARC 레코드는 패싯을 만들 수 있는 충분한 요소를 가지고 있어 여러 가지 범주를 사용할 수 있다. 예를 들면, 개인명, 지리명, 장르, 주제, 출판년의 범위, 매체형태, 언어 등은 MARC 레코드로부터 쉽게 얻어지는 패싯의 범주이다. 하지만 구조화되지 않은 데이터의 경우 패싯 구성이 어려워 MARC 레코드 외에 패싯의 생성이 가능한 원시데이터는 아주 드물다.



그림 1. 미국 퀸즈 공공도서관 차세대 도서관 목록 인터페이스

- (3) 키워드 검색(Keyword Searching): 구글 등 검색엔진처럼 키워드 검색창을 가지고 간편한 검색 기능을 제공하자는 것이다. 하지만 도서관 검색인터페이스는 키워드 검색창만으로는 충분하지 않고 상세검색(advanced search) 페이지를 추가로 마련하여 보다 정확한 검색기능을 함께 제공하는 것이 바람직하다. 또는 키워드 검색을 수행하더라도 그림 1과 같은 패싯 네비게이션 기법을 더하여 입력보다는 “클릭”으로 검색결과를 “제한”하는 방법도 생각해 볼 필요가 있다. 이는 기존 검색방식인 제한검색에서 “검색->검색결과 내 제한”으로 순서를 바꾼 것이다.
- (4) 검색결과와 적합성 순위화(Relevancy): 웹 검색엔진 이용경험이 많은 사용자들의 기대치 중 하나는 검색결과와 순위화이다. 요즘 많은 웹 검색엔진은 여러 종류의 적합성 순위화 방법으로 결과를 정렬하여 제공한다. 이 순위화는 많은 검색결과 중 가장 적합한 항목이 리스트의 최상단에 나타나도

5) 미국의 경우 Bowker 의 자회사인 Syndetic Solutions(www.syndetics.com)가 목록 레코드의 디스플레이를 풍부하게 하는 콘텐츠를 도서관에 제공하는 주도적 역할을 한다.

6) 패싯 (facet)은 필자가 이는 범위에서는 량기나단의 콜론분류법에서 처음 사용된 것이 아닌가 한다. 이는 원래 다이어그램의 절단된 한 면을 의미하나, 우리 분야에서는 대상 자료 또는 용어의 특정 “측면”을 의미한다.

록 하는 것인데, 불리안 검색기법을 사용하는 도서관 환경에서는 사실상 어렵거나 효과가 의심된다. 어쨌든 검색엔진에서 적합성 순위화를 하는데 그 방법은 일단 검색일치항목을 DB에서 가져온 후, 질의문이 여러 개의 키워드로 구성되어있는 경우, 그 키워드가 검색항목에 몇 개가 출현하는가로 우선 순위화하고, 특정 검색어의 출현빈도가 검색대상 자료에서 높으면 적합성을 다시 상향시킨다. 다른 방법은 질의문에 포함된 키워드가 검색항목에 출현하는 위치를 가지고 정한다. 즉, 질의문의 키워드를 제목 필드에 가진 문헌이 본문에 가진 문헌보다 높은 적합성 가중치를 갖도록 하는 것이다. 하지만 이와 같은 기술적 순위화(technical ranking)가 반드시 검색결과 내에서 중요성이나 이용자 관심도를 반영한다고 단언할 수 없다. 따라서 사회적 척도(social measures)를 채택할 수도 있는데, 예를 들면 검색된 항목에 링크된 문헌의 수나 그들의 품질, 또는 검색항목에 포함된 인용문헌의 수나 그들의 품질 등을 순위화에 사용할 수도 있다. 이와 같이 검색결과 순위화를 위하여 검색된 항목의 중요성이나 이용자 관심도를 표현하는 다양한 실마리를 이용할 수 있는데, 다른 예로 자료의 대출빈도는 그 자료의 인기를 반영한다고 할 수 있고, 도서관에서 해당 자료를 소장한 복본의 수나 종합목록에서 해당자료를 소장한 도서관 수 등도 이와 유사한 척도가 될 수 있다.

- (5) 철자오류 수정("Did You Mean...?"): 검색엔진에서 기대되는 기능 중 다른 하나는 자주 발생하는 질의문의 철자오류 감지기능이다. 이는 구글, 야후, 아마존 등에서 "Did you mean...?"이라고 검색이 가능한 검색어를 제시하는 것을 말한다. 이를 구현하는 알고리즘은 다양하지만, 좋은 "Did you mean...?" 기능은 단순한 철자오류 점검을 넘어 원래 질의문보다 더 많은 결과를 얻을 수 있는 용어를 제시하는 것이다. 이 기능은 잘못된 검색결과나 "검색결과 없음"이라는 메시지보다 이용자에게 무엇인가를 제안하므로 훨씬 유익하다.
- (6) 추천기능(Recommendations): 일반적으로 전자상거래에서 나타나는 기능으로 관련 자료에 대한 정보를 적극적으로 제공하는 것이다. Amazon.com을 예로 들면 "X를 구매한 고객은 Y도 구매를 했습니다"라는 추천기능이 있다. 이러한 상업분야의 추천기능이 도서관에는 적합치 않을 수 있으므로 도서관에 적합한 추천기반을 위하여 이용자 검색행위 데이터를 상세히 분석해 볼 필요가 있다.
- (7) Web 2.0의 이용자 참여(Enabling User Contribution): 웹 2.0 개념에 의하면 자원이라는 것은 정보의 일방적 제시가 아니라 이용자 참여를 유도하는 것이다. 차세대 도서관 목록에 이 접근 방식이 적용될 수 있는 방법은 다양하다. 외부 자원으로부터 보충된 풍부한 콘텐츠 외에도 목록은 이용자의 추가적인 콘텐츠 제공에 의존할 수도 있다. 즉, 이용자에게 목록 레코드에 자료의 등급 부여를 유도하거나 서평을 달아 자신의 의견을 피력하게 하는 것이다. 또 다른 이용자는 이미 있는 서평에 의견을 추가하거나 자신의 서평을 달도록 할 수 있다. 하지만 도서관 이용자들에게 이와 같은 참여를 이끌어 낼 수 있을지는 아직 미지수이다. 이외에 태깅(tagging)은 웹 2.0과 관련된 또 하나의 기능으로 이용자가 자신만의 비공식 용어를 사용하여 관심있는 자료에 그 용어를 부여하는 것이다. 이는 이용자가 자신의 태그를 임의로 생성하여 나중에 해당 자료를 찾는데 사용하는 것이므로 어떤 태그는 지극히 개인적 관심에 국한되겠지만 다른 것들은 보다 공동체지향적(community oriented)일 수도 있다. 그 예로 폭소노미(folksonomy)가 있으며 이는 특정 관심그룹(community)사이에서 정보를 분류하기 위하여 개발된 태그의 세트이다. 얼핏 이보다 훨씬 정교한 방식을 사용해 온 도서관에서 이런 태그의 사용은 반박의 여지가 있으나 이용자가 부여한 태그는 전통적으로 도서관에서 사용해 온 주제명의 보완역할을 할 수도 있을 것이다.

- (8) RSS: 기존 웹 페이지를 통한 서비스 이외에 RSS(really simple syndication 또는 rich site summary)를 이용한 콘텐츠 배포는 이용자가 콘텐츠를 보다 쉽게 이용할 수 있는 길을 열어준다. RSS는 단순한 XML 프로토콜을 통하여 관련된 일련의 자료를 제공하며 오늘날 널리 사용되고 있다.

위에 나열한 기능은 차세대 도서관목록 구성요소 중 일반적인 것들이며 각 도서관 또는 도서관자동화시스템 개발자들은 다소 다른 의견을 가질 수 있다. 그러나 이러한 노력의 공통적인 흐름은 기존 도서관 목록의 기능을 발전시키는 것이고 이를 통하여 도서관 이용자에게 보다 강력하고 사용하기 쉬운 툴을 제공하는 것이다.

## 결언

차세대 도서관 목록은 Library 2.0과 그 맥락을 같이 하며 일견 동일한 것처럼 착각을 일으킨다. 하지만 그 대상이 전자는 시스템 인터페이스에 제한적이며 후자는 도서관 서비스 전반을 지향하는 것으로 그 차이가 뚜렷하나, 도서관이 변화해야 하고 이용자 요구에 적응해 가야 한다는 공통된 메시지를 가지고 있다.

외국의 경우 특정 ILS에 제한되어 또는 독립적으로 구현된 차세대 도서관 인터페이스 패키지들이 존재한다. 가장 널리 보급된 것은 네덜란드 회사인 Medialab Solutions BV의 제품인 AquaBrowser<sup>7)</sup>이다. 이 제품은 공공도서관을 주 고객으로 하였으나 차츰 관중을 확장하며 세계적으로 보급되고 있다. 이외에도 Endeca Technologies의 Endeca, Innovative Interfaces의 Encore, Ex Libris의 Primo, OCLC의 WorldCat Local 등이 있다.

국내에서도 이러한 제품을 수입하면 어떨까 하는 생각을 할 수 있다. 그러나 Unicode기반의 외국제품은 일률적으로 검색에 n-gram방식<sup>8)</sup>을 사용하고 있다는 문제가 있다. 이는 과거 우리가 사용해 온 형태소분석 방식의 한글검색에 비하여 원하지 않는 검색결과를 제공하는 심각한 문제가 있다. 또한 외국시스템은 국내 제품처럼 판매도 하지만 리스(lease)의 개념과 이용자 수를 기반으로 가격을 산정하여 구매비용이 비싸며(약 3억 수준) 유지보수 비용도 국내 제품처럼 구매자가 “우긴다고” 내려가지 않고 연간 10~20%를 착실히 요구하므로 경제적 측면을 심사숙고 해야 한다<sup>9)</sup>. 

7) 그림 1은 AquaBrowser를 적용한 사이트로서, 접근경로는 <http://www.queenslibrary.org/> 이다.

8) N-gram 방식은 대상 스트링을 몇 개의 음절로 나누어 색인을 하는가에 따라 n이 바뀌는데 한글은 두 음절씩 색인하는 bi-gram방식이 제일 적합하다고 한다. 그러나 “정보검색”의 경우 bi-gram 방식은 “정보, 보검, 검색”의 세 단어로 색인을 만들기 때문에 원하지 않는 “보검”의 검색결과가 나온다는 것이다. 필자의 경험에 의하면 국내 이용자들에게는 옹서가 되지 않는 검색결과이다.

9) 이 가격은 Breeding, M. (2004). Integrated Library Software: a guide to multiuser, multifunction systems. *Library Technology Reports*, 40(1)의 p. 22에 나타난 ILS 비용이다.