



## “구글 프린트” 프로젝트에 관한 고찰

이향국\*

계속 발전하고 있는 정보 컴퓨터 통신 기술의 발달은 일상생활뿐만 아니라 도서관에도 큰영향을 주고 있다. 특히 도서관이라는 개념을 실제공간(장소)에서 가상공간으로 확대하는데 큰 기여를 했다.

구글의 “구글 프린트” 프로젝트는 지금까지 우리가 생각해오던 “도서관”의 정의를 시대에 맞추어 다시 생각하게 할 수 있는 분기점이라고 한다면 너무 큰 과장이 될까? 현재 저작권 문제로 당분간 구글의 “library project”가 중단된 상태이지만 이 프로젝트가 계속 진행된다면 도서관계에 영향이 있을뿐만 아니라 도서관 이용자에게는 기술의 이점을 이용, 도서관장서의 사용에 많은 편리함을 제공할 수 있기 때문에 도서관에서 종사하는 사람으로 이 글을 나누고자 한다. ■ 필자 주

### “구글 프린트” 프로젝트는 무엇인가?

2004년 12월 하버드대학 도서관은 인터넷 검색엔진 회사 “구글”과 손잡고 하버드대학이 소장하고있는 장서들을 디지털화해서 웹상에서 검색하고 전문을 읽을 수 있도록 하는 “구글 프린트(Google Print)”라는 프로젝트를 시작한다고 발표했다.<sup>1)</sup>

4만권의 책들이 우선적으로 시작될 것이며 추후 하버드가 소장 하고있는 1500만권들의 책들이 디지털화 되어질 것이라고 했다. 이 소식은 도서관 이용자, 도서관인, 그리고 출판계등에 적지않은 반향을 일으켰다.

1950년대 컴퓨터의 출현으로 정보가 전자형태로 만들어진것을 시작으로 1960년대초 최초로 컴퓨터를 이용해서 검색을 할 수 있는 의학관련의 MEDLAS<sup>2)</sup>가 개발되었고,

\* 하버드 엔칭도서관, hkleee@fas.harvard.edu

1) Harvard libraries and Google announce pilot digitalization project with potential benefits to scholarworldwide, <http://www.news.harvard.edu/gazette/daily/2004/12/13-google.html>

2) Medical Literature Analysis and Retrieval System의 첫글자를 딴것으로, 1960년 미국의학도서관(NLM)에서 사용하기 위해 개발된 의학관련 전자 서지 정보 시스템, <http://www.dialog.com/about/history/pioneers1.pdf>

1971년 MEDLAS의 온라인판인(Online dial-up databases search) MEDLINE<sup>3)</sup>, 1972년 DIALOG가 처음으로 상업용 데이터베이스의 서비스를 시작했다. 그후 1980년대 중반 CD-ROM의 개발은 정보가 전자형태로 저장되고 검색할 수 있는 많은 데이터베이스들의 형성을 도왔다. 1960년대 ARPANet(Advanced Research Projects Agency Network)과 1970년대 미국 국방부의 국방관련 연구와 커뮤니케이션을 돕기 위해 개발되어진 커뮤니케이션 네트워크는 1990중반 웹의 개발로 정보의 전달과 생산뿐만아니라 저장과 검색에도 큰영향을 미쳤다. 또 GUI(Graphical User Interface)의 개발은 글뿐아니라 그림, 소리, 비디오 등 다양한 형태의 정보가 인터넷에 저장되고 검색이 가능하도록 하는 기능을 제공했다. 웹이 도서관에 끼친 가장 큰 영향중 하나가 데이터베이스를 컴퓨터 본체에 저장하거나 CD-ROM을 사지 않아도 정보의 저장과 검색이용이 가능하게 한 것이다. 그후 많은 웹데이터베이스들이 개발되어 의학 법률 등 특정정보의 데이터베이스 뿐만아니라 정기간행물의 기사를 웹에서 검색하고 그 전문을 읽을 수 있는 서비스를 제공해 오고 있다.

현재 많은 책들이 저작권에 묶여있어 “구글 프린트”는 전문을 보여주는 기능보다는 필요한 책을 찾고 발견하도록 하는 기능이 크지만 도서관에 소장된 모든 책들의 전문을 웹페이지에서 검색하고 읽는다는 것은 보다 더 진일보한 도서관 서비스의 혁명이라고 할 수 있다.

1990년초 부터 영국, 미국, 한국을 포함한 여러나라에서 고서, 사진같은 특별한 자료를 중심으로 장서의 디지털화가 시작되긴 했지만, 원본에서 디지털화되기까지 사람의 손을 많이 필요로 할 뿐만아니라 고성능의 스캐너, 바코드 리더기등의 장비와 소프트웨어 등이 구비되어야 하는 간단하지 않은 과정이며 그에 따른 고비용, 디지털화 되는 동안 이용자의 사용제한 등 현실적으로도 도서관이 소장하고 있는 막대한 장서를 디지털화 한다는 것은 그리 간단한 과제는 아니다. 많은 예산과 시간뿐 아니라 기술적인 문제가 수반되기 때문이다. 그 예로 뉴욕공공도서관 소장의 장서를 디지털화 한다면 어느 정도의 예산과 시간이 소요될까? 한 권의 책을 디지털화하는데 \$10의 경비가 소요되고<sup>4)</sup> 30분의 시간이 든다고 가정한다면 2000만 권의 책을 소장하고있는 뉴욕공공도서관의 디지털 라이브러리 프로젝트는 약 1,131년의 시간과 2000억(\$200 millions)이라는 천문학적인 경비가 들 것이다.

구글과의 협력이 없이는 도서관이 소장하고 있는 모든 장서를 디지털 하는 것은 시간이나 경제적으로 불가능했을 것이라고 언급한 뉴욕공공도서관장 외 다른 협력 도서관의 담당자들의 말처럼 “구글 프린트” 프로젝트는 구글의 기술이 겸비된 거대한 경제력과 우수한 도서관 장서와의 합작품이라 하겠다. 소장도서의 디지털화는 도서관의 책이라는 인쇄매체가 물리적인 의미에서 가상 또는 사이버로 더욱 많이 확대되는 계기가 될 것이다.

만약 저작권의 문제가 없는 책인 경우, 미국 뉴욕주 뉴욕시 소재 뉴욕 공공도서관 소장의 희귀본을 서울에서 인터넷이 연결된 컴퓨터로 검색하고 전문까지 읽을 수 있는 것이다. 그러면 구글이 추진하고 있는 이 야심찬 프로젝트 “구글 프린트”는 무엇이며 과연 도서관, 도서관인, 이용자에게 어떤 영향을 줄 것인가?

3) MEDlas onLine에서 따온 미국 의학정보 온라인 데이터베이스로 1971년 미국의학도서관을 이용할 수 없는 외부 이용자를 위해 개발되었다.  
<http://www.dialog.com/about/history/pioneers1.pdf>

4) Google is adding major libraries to its database, John Markoff and Edward Wyatt,  
<http://www.nytimes.com/2004/12/14/technology/14google.html>

“구글 프린트”의 사명은 세계에 흩어져 있는 정보를 이용자가 이용하기 쉽도록 정리하여 주는것으로 오프라인에서 이용할 수 있는 정보를 온라인에서도 이용자가 쉽게 접근해서 이용할 수 있도록 하는 것이다.<sup>5)</sup> 다시말해서 출판사와 저자들의 저작권을 존중하면서 책들의 “가상의 카드목록”을 제공하는 것이다. 이것은 인터넷 중심의 문화에 부응하는 시대적인 발상으로 차세대의 목록이라고 표현해도 될 것이다.

“구글 프린트”에서 제공되어지는 그 많은 책의 원천은 무엇일까? “구글 출판사 프로그램(Google Publisher Program)”과 “구글 프린트 도서관 프로젝트(Google Print Library Project)”를 통한 두 개의 프로그램에 계약한 도서관들과 출판사들이 이 프로젝트의 소스이다. “구글 출판사 프로그램”은 출판사(저자)와의 프로젝트이고 “구글 프린트 도서관 프로젝트”는 구글과 파트너 협정을 맺은 도서관들과의 프로젝트이다. 구글 출판사 프로그램의 경우 출판사에서 구글회사로 책을 보내면 책의 전문을 디지털화 하고 목차를 “구글 프린트” 검색데이터베이스에 넣어서 이용자가 검색을 할 경우 검색이 가능하게 하는 것이다. “구글 프린트”에 들어가는 모든 책들은 책의 저작권에 관계없이 책의 전문이 스캔되어 디지털화 되어진다. 이는 모든 페이지를 구글 자체의 데이터베이스 색인에 넣어서 이용자의 효과적인 검색을 돕도록 하기 위한 것이다.

현재 “구글 프린트”에서 검색되어지는 대부분의 책들은 구글 출판사 프로그램에서 제공되어지는 것이다.

현재 구글 도서관 프로젝트는 하버드대를 비롯하여 미시간대, 스탠포드대, 옥스포드대 등 4곳의 대학 도서관과 뉴욕 공공도서관 등 5개의 도서관에서 계약을 시작했다. 프로젝트의 내용은 각각의 도서관마다 상이하다. 현재 구글은 더 이상의 미국 내 대학도서관을 추가할 계획은 없지만 특별한 장서를 소장하고 있는 도서관이나 영어권 이외의 언어를 사용하는 나라의 도서관들과의 협력을 계획하고 있고, 실제로 프랑스, 이탈리아, 독일, 스페인, 네덜란드등 비영어권인 유럽의 소재한 출판사가 이 프로젝트에 파트너로 책들을 구글로 보내기 시작했다.<sup>6)</sup>

미시건 대학의 경우 본 대학이 소장하고 있는 모든 장서(7.4 million)가 스캔되어질 계획이고, 옥스포드 대학의 경우 저작권에 문제가 없는 주로1900년 이전에 간행된 책들에 한하여, 하버드 대학의 경우는 선발적으로 영어로 인쇄된 책들 중에서 무작위로 40,000권이 선택될 것이다, 뉴욕공공도서관의 경우 저작권에 걸리지 않으며 일반시민들의 관심이 많은 분야에 한해 스캔하는 도중 손상 될 가능성이 있는 책들은 제외시켰다. 스탠포드 대학의 경우는 책들을 스캔할 장소로 구글 본사를 정했을뿐 그외 다른 구체적인 사항은 진행중이라 한다.

그러면 서가의 책들은 어떻게 디지털이 되는가?

서가에 있는 책이 스캔이 되어 컴퓨터 파일로 저장되고 다시 본래의 서가에 정리되기까지는 어떤 과정이 필요할까?

인쇄된 책을 바이트(bite)로 저장하기위해서는 스캐닝을 하는것이 첫 번째 과정이다. 도서관의 장서를 스캐닝하기위해서는 고성능 스캐너, 바코드 리더기, 컴퓨터, 또는 책장을 넘기는 기계등과 같은

5) <http://print.google.com/googleprint/about.html>

6) Google print goes to Europ, Susan Kuchinska, [www.internetnews.com/xSP/article.php/3531776](http://www.internetnews.com/xSP/article.php/3531776)

장비가 구비되어 있는 특정장소로 옮겨서 스캔을 한후 다시 제자리에 갖다 놓아야 한다. 작업장소는 기관에 따라서 구글의 본사가 될수도 있고 디지털화하는 도서관이 될 수도 있다. 장서를 옮기는 과정, 스캔하는 과정에서 원본에 손상이 안가도록 하는것은 디지털화 하는 과정에서 가장 중요한 일의 하나가 될 것이다.

과정을 간단히 순서대로 열거해보면,

1. 한 선반 전체의 책들을 순서대로 카트에 싣는다.
2. 바코드 리더기로 카트 위의 책들을 점검한다.
3. 소프트웨어가 도서관 전자목록기록 (electronic catalogue 의 record)을 찾아서 구글 데이터 센터로 보내 책과 일치시킨후, 디지털 할 수 있는 장비가 갖추어진 작업장소로 카트를 보낸다. 한번에 여러 권의 책들을 스캔 할 수 있도록 몇 대의 기계들이 동시에 작업을 하면 시간이 단축될 수 있다.
4. 카트에 있는 순서대로 스캔되어지며, 작업후에도 카트에 원래 순서대로 정리된다.
5. 스캔이 되었는지 확인하고 디지털 파일은 로우포맷(raw format) 형태로 구글 데이터 센터로 보내지며 그곳에서 이용자가 이용할 수 있는 형태로 가공되어진다.

## 검색방법

“구글 프린트” 홈페이지(<http://print.google.com/>)에서 검색하고 싶은 용어를 입력하면 검색어가 들어간 목차를 가진 책들의 리스트가 보이며 화면 오른쪽에 검색어, 검색되어진 페이지의 총숫자와 “구글 프린트”의 데이터베이스를 검색한 시간이 보여진다. 각각의 책에는 스캔되어진 책의 표지가 보이며 서명, 저자, 출판사, 출판년도, 검색어가 포함된 페이지와 문장을 볼수 있어 검색된 많은 리스트 중에서 내가 찾고자 하는것에 좀더 근접한 책을 선별할 수 있는 기회를 준다. 책제목을 클릭하면 저작권에 문제가 없는 책들의 경우는 전문을 볼 수 있지만 저작권에 걸리는 경우 검색어와 관련된 몇 페이지의 내용만 보여지거나 서지사항과 발췌문을 볼 수 있을 뿐이다. 때로는 간단한 문장만 보여지는 경우도 있다.

왼쪽의 상자는 크게 “Book pages”와 “Buy this book” 두 부분으로 나누어져있다. Book Pages 중 “about this book”은 서명, 저자, 출판사, 출판년, 주제, 총페이지수, 크기, ISBN 등의 서지사항과 같은 책과 관련된 보다 많은 정보를 제공한다. 그외 목차, 색인, 저작권 상황, 서평, 책과 관련된 내용이 들어있는 다른 웹페이지들과도 연결된다.

검색자가 책을 구입하고 싶은 경우는 “Buy this book”에 링크된 온라인 책방에서 살 수 있다. 도서관에서 스캔되어진 책의 경우 “Find this in a library”를 클릭해서 우편번호를 입력하면 책을 소장하고 있는 가까운 도서관에 대한 정보도 제공될 것 이라 한다.

저작권에 대한 규정은 1923년 이전에 미국에서 출판된 도서는 책의 전문을 볼 수 있다. 또 미국 정부가 저자인 출판물인 경우도 저작권에서 자유롭다. 미국 외의 나라에서 출판된 책들의 경우 확일화

된 국제 저작권법이 없기때문에 1900년이전에 출판된 경우에 한해서 자유롭긴 하지만 각 나라의 법에 따를 것이라 한다.

## 도서관에 미치는 영향

도서관이라는 물리적인 장소에 꼭 가지 않아도 원하는 책을 볼수 있다면 과연 앞으로 도서관은 어떻게 변할 것이며, 새로운 형태의 도서(e-book, book on the Web, digitalized book) 출현에 따른 사서의 역할은 어떻게 될것인가? 이용자에게는 어떤 영향이 있는가?

인터넷이 연결된 컴퓨터만 있으면 시간과 장소에 구애됨이 없이 편리한 장소에서 원하는 책을 읽을 수 있는 것은 이용자의 입장에서 "구글 프린트"의 가장 큰 매력이라 하겠다. 또 이용자의 요구가 많은 도서의 경우도 다른 이용자가 책을 반납 할 때까지 기다리거나, 다른 이용자가 요구시 반납해야 하는 번거로움이 없어진 것이다. 즉 다른 이용자와 책을 공유해서 볼 필요가 없어진 것이다.

또 도서관의 입장에서도 이용요구가 많은 책일지라도 복본을 구입할 필요가 감소 되어진다.

종이형태의 도서관 장서가 바이트(bits) 형태의 웹페이지로 전환되면 과연 도서관은 어떻게 변할까? 도서관이 단지 인쇄형태의 책들을 보관하는 책창고의 의미로 변하지는 않을까?

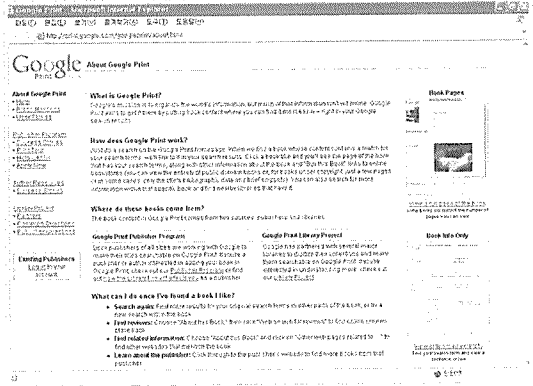
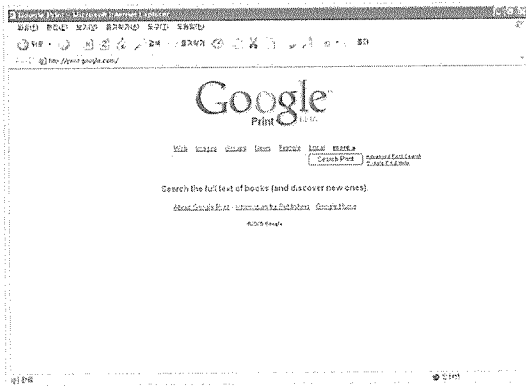
도서관에서 책을 대출하는 것이 길거리에서 동전을 넣고 공중전화를 사용하는 것처럼 시대에 뒤떨어진 일이 되지는 않을까? 그러나 텔레비전, 인터넷 방송 등의 출현으로 라디오를 듣는 사람들이 사라지는 않을까 했던 기우처럼 여전히 인쇄된 매체를 선호하는 사람들이 있을 것이며 빠른 인터넷이 구비된 도서관의 시설 이용을 위해, 또 참고질문을 하기 위해서, 전문을 디지털하는데 있어서 저작권에 걸려 전문을 인터넷에서 볼 수 없을때 책들을 보기 위해서 등 여러이유에서 도서관을 찾는 이용자 수는 감소하지는 않을 것이다.

도서관의 책들이 디지털화 되어지고 디지털 형태의 책들이 증가한다해도 도서관 서비스가 확대되는 것이지, 인쇄 형태의 장서 수집이 감소하는 것을 의미하지는 않을 것이다. 인쇄형태로 출판되는 책들이 디지털화 형태의 책을 위한 대치용으로 구입되는 일은 없을것이다. 도서관인에게는 장서 중 디지털 되어진 책들을 기존의 인쇄본의 전자목록 데이터베이스에 디지털 되어진 새 형태의 책에 대한 새로운 정보를 추가해야 하는 업무가 추가될것이다. 또 계속적으로 새로운 소프트웨어의 개발과 그것을 이용한 도서관 업무의 증가는 도서관인에게 끊임없이 지식습득을 위한 노력을 요구할 것이다.

이용자의 입장에서는 "구글 프린트"프로젝트의 특혜를 가장 많이 누릴수 있다. 인터넷이 연결된 컴퓨터는 시·공간을 초월해서 무료로 원하는 책을 검색하고 읽을 수 있게 하는 것이다. 전문이 제공되지 않더라도 직접 책방에 가지 않고 그 책에 대한 정보를 습득할수 있고 책의 구입결정에 많은 도움을 받을 수 있을 것이다.

현재 저작권에 대한 문제점들이 지적되고 있고, 원본이 손상될 가능성 등의 문제점이 있지만 원본 손상의 문제나 작업 시간의 단축 등은 끊임없이 개발되는 장비와 기술개발의 발전으로 점차 극소화 될 것이다.

그의 또 하나 지적되는 것이 손으로 쓴 원고류를 해독할 수 있는 광학식 문자판독 인식기술(OCR : Optical Character Recognition)이 아직 없는 것이다. 이런 경우 전문의 글자를 일일이 다시 입력해서 원래의 디지털된 파일에 연결해서 검색이 되도록 해야하기 때문에 많은 시간과 인력이 필요한것이다. 위에 열거한 점들이 개선된다면 “구글 프린트”는 협력도서관들이 경제적인 부담이 없이 장서를 디지털화해서 인터넷을 통해 무료로 전세계 사람들이 사용할 수 있게하는 진일보한 도서관의 서비스를 제공하는 것이다. 또 “구글 프린트”의 대중화는 아마도 현존하는 도서관의 검색 웹페이지를 구글의 사용에 익숙해 있는 사용자들의 편의를 위해 구글의 검색창처럼 바뀌어야 하지는 않을까?



〈구글 프린트 홈페이지와 안내 페이지〉