

디지털 이미징 기술*

Imaging for Libraries and Information Centers

Richard W. Boss

최 선 회**

Tran. by Seonheui Choi

차례

- | | |
|-------|----------|
| 1. 서론 | 3. 기술 기반 |
| 2. 개관 | 4. 사례연구 |

1. 서론

지난 몇 년 동안, 수많은 도서관들과 문서보관소들이 도서, 팜플렛, 학술지 논문, 고문서, 사진, 슬라이드 및 다른 자료들을 디지털 형태로 이미지를 찍기 시작했고, 자체 시설 내에 혹은 자체 시설이 없이도 검색을 위해서 그것들을 저장하고 있다. 어떤 기관들, 특히 대체할 수 없는 자료들의 노화되는 컬렉션을 갖고 있는 도서관들과 문서보관소들에게는, 그 동기가 자료의 보존

이지만, 대부분은 그 동기가 접근을 개선하는 것이다.

1.1 보존

많은 도서관들, 특히 연구도서관들과 문서보관소들은 종이들이 노화, 화학적 불안정성, 환경적 상황, 처리, 자연재해 및 도난 등의 결과로 점점 약해지고 있기 때문에, 소장 컬렉션들의 악화와 손실이라는 지대한 문제로 위협받고 있다. 보존은 도서관 자료나 문서자료들을 물리적으로

* 이 자료는 미국도서관협회에서 발행하는 *Library Technology Reports*, v.37, no.1 (Jan./Feb. 2001)에 수록된 Richard W. Boss의 "Imaging for Libraries and Information Centers."를 요약, 번역한 것이다. 미국도서관협회는 이 보고서의 번역을 허가하며 다음의 정보를 추가하도록 요청하였다: "Library Technology Reports is published by ALA TechSource, a unit of the publishing division of the American Library Association. ALA TechSource did not review this summarization and is not responsible for the accuracy of the information contained in this abbreviated version of 'Digital Imaging Technology' (2001.5.31)". 이 보고서는 도서관에서 사용할 수 있는 여러 가지 디지털 이미징 기술을 요소별로 개관하여 그 특징을 서술하고, 실제로 이미징 프로젝트를 수행하였던 주요한 도서관들의 사례를 소개하고 있다.

** 지식정보관리실 연구원(Researcher, Information Resources Management, KISTI)

안정시키고 그들의 외형을 변형시키지 않고 다른 형태로 재생하는 보존 기법에 의하여 사용하기 위하여 관리하는 것과 관련된 활동들로 구성된다. 재생은 접근을 제한함으로써 원본을 보호할 뿐만 아니라, 원본을 아주 상세히 표현하여, 만약 원본이 상실되었을 때에조차도, 전부는 아니더라도 이용자들의 요구를 맞추어 주도록 한다.

만약에 보존이 도서관이나 문서보관소에서 자료들의 이미지를 찍는 주요한 동기라면, 마이크로필름화가 디지털 이미징보다 훨씬 비용효과가 크고, 더 영구적이다. 선명한 활자체로 마이크로필름화하는 것이 한 장당 \$0.15일 때, 디지털 이미징은 적어도 \$0.31에서 \$0.34가 든다. 책과 팜플렛의 경우는 비용이 현저히 상승한다: 마이크로필름화는 페이지 당 \$2, 디지털 이미징은 \$4.30에 이른다. 사진은 그 비용에 65%를 추가한다.

아카이벌 품질의 마이크로필름은 100년까지도 간다, 그러나 디지털 이미지는 매 10년 정도마다 갱신하거나 재복사해 주어야 한다, 부분적으로 그 이유는 저장기술이 계속적으로 변화하기 때문이다.

1.2 접근

디지털 이미징의 추가비용은, 원본의 상태가 어떻든지에 상관없이, 자료의 잠재적 이용도가 높으면 정당화된다. 디지털 이미지의 검색은 이미지와 동일한 장소에 있는 이용자에게만 한정되는 것이 아니고, 한 시점에 한 이용자에게만 제한되지 않는다. 지리적 제약을 극복하는 것은 종종 이미지

배포 (image distribution)라고 기술된다, 그러나 이 보고서에서는 이용자를 강조하고 있기 때문에 접근(access)이라는 용어를 사용한다.

- 보존과 접근이 서로 배타적이지는 않다. 노스이스트 문서보존센터 (Northeast Document Conservation Center)는 비록 도서관과 문서보관소에 보존용 마이크로필름화 서비스를 제공하기 위하여 세워졌지만, 거기에서 마이크로필름화되는 많은 문서들이 차후에 추가비용으로 디지털이미지로 만들어진다는 것을 보고한다. 마이크로필름은 그리고나서 보존을 위하여 저장되고, 디지털 이미지 형태는 접근을 향상하기 위하여 사용된다.
- 이 보고서에서는 접근이 강조되고 있지만, 그러나 그 동기가 보존인 도서관들에게도 많은 정보가 유익할 것이다.

이미지 프로젝트의 요소들 -- 이미지 찍기, 이미지 저장, 이미지 조직 (편목이나 색인), 이미지 전송, 및 이미지 접근 --이 하나의 시스템을 구성한다. 시스템은 모든 필요한 요소들을 포함하는 툴이나 꾸러미 상품으로서 벤더가 판매용으로 합성할 수도 있고, 혹은 도서관이나 문서보관소가 많은 소스로부터 합성할 수도 있다. 요소들간의 호환성은 표준을 따름으로써 달성된다.

2. 개관

이미지의 일반적 사전의 정의는 재생 (reproduction) 혹은 유사함(likeness)이다,

그러나 대부분의 사전들이 이미징, 디지털 이미징, 이미징 시스템, 이미징 기술, 문헌 이미징, 혹은 전자적 이미징 등을 정의하지 않고 있다.

2.1 정의

이 보고서의 목적을 위하여, 디지털 이미징, 이미징 시스템, 이미징 기술 등등은 이미지들을 찍고, 편집하고, 전자적으로 저장하고, 조직하고, 전송하고, 접근할 수 있도록 결합하는 일단의 기술들로 정의한다. 이미지는 인쇄된 자료의 페이지일 수도 있고, 사진, 슬라이드, 동영상 비디오, 혹은 그밖의 다른 소스 문서나 원문일 수도 있다.

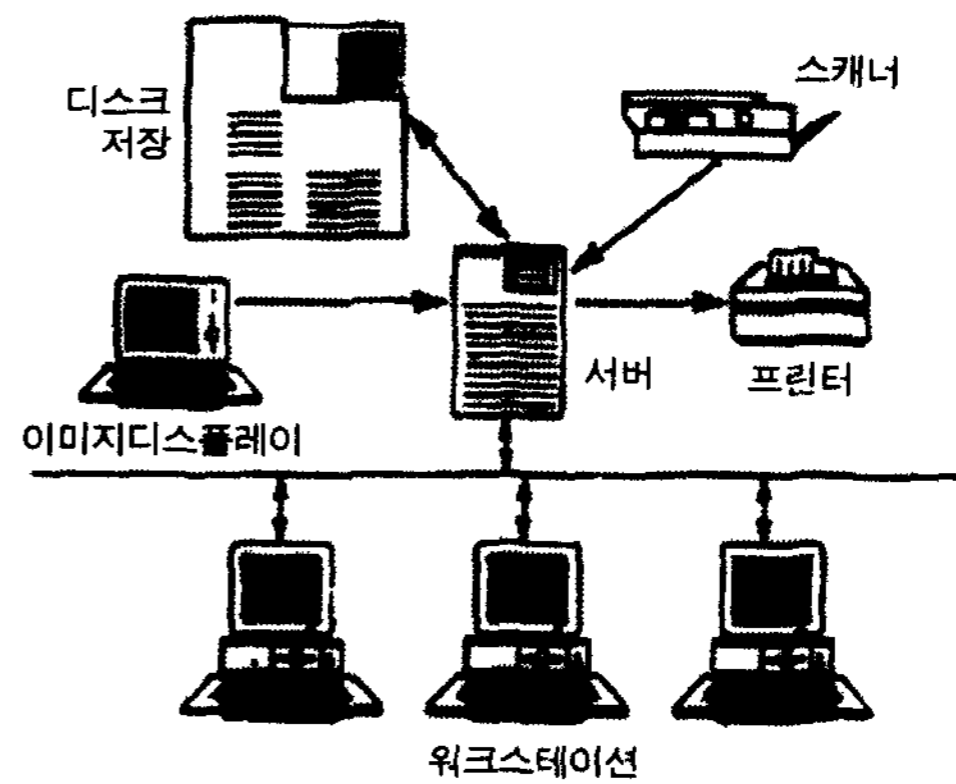
2.2 디지털 이미징을 그래픽 및 전문 (full-texts)과 구별하기

디지털 이미징은 눈으로 볼 수 있는 것을 찍고, 나중에 접근이나 검색을 위하여 그것을 디지털로 저장하는 것에 관련되어 있다는 점에서 그래픽과 다르다: 그래픽은 컴퓨터 안에서 데이터베이스로부터 만들어진 시각적 표현과 관련이 있다. 이미지는 그래픽보다 엄청나게 큰 처리 및 저장조건을 필요로 하고, 따라서, 상이한 하드웨어와 소프트웨어를 요한다.

이미지는 또한 전문과도 구별되어야 한다. 디지털 이미징에서 이미지는 전체가 디지털화되어, 그것은 원본의 복제이다: 전문은 글자-대-글자로 기계가독형태로 변환하는 것이다. 이미지는 이미지를 전체로 검색하는 것만을 허용하고, 전문은 텍스트를 탐색하고 특정한 부분들을 검색하는 것을 허용한다.

2.3 전형적인 이미징 시스템

비록 기술이 응용시스템마다 다르긴 하지만, 전형적인 이미징 시스템의 운영은 극히 간단하다. <그림1>은 요소들과 과정들을 나타낸다.



<그림 1>

소스 문서가 찍히고, 편집되고, 저장 요구를 줄이기 위하여 압축되고, 저장되고, 조직(편목 혹은 색인)되고, 그것에 접근하고자 하는 이용자에게 네트워크를 거쳐 전송된다.

2.4 역사

비즈니스 기술로서, 이미징 기술은 비교적 새로운 것이다. NASA에서 위성 이미지를 향상시키고 조작하기 위하여 사용하였던 복잡한 시스템의 부산물로서 1980년대 초반까지는 주로 과학자들과 기술자들이 사용하였다. 이 무렵 비즈니스용의 이미징 시스템이 등장하기 시작하였다. 이들은 좁게 규정된 과제들을 처리하고 제한된 시장이나 알맞은 일반적으로 주문생산형의 독립적 시스템들이었다.

넓은 시장을 위하여 이미징을 가능하게 하고, 점차 살만하게 해준 기술적 진보에는 다음과 같은 것들이 있다 :

- 이미지의 효율적, 고해상 촬영을 제공하는 기술
- 이미지의 편집을 가속화하는 컴퓨터 처리능력과 기억장치의 개선
- 낮은 비용으로 이미징에 필요한 많은 양의 데이터를 전자적으로 저장하게 하는 저장능력
- 저장 조건을 더욱 줄여주는 데이터 압축
- 이미지를 전송하는데 필요한 큰 대역폭과 높은 데이터 전송율을 허용하는 고속네트워크
- 높은 품질의 이미지를 제공하는 고해상 모니터

2.5 응용기술

이미지 기술은 많은 산업 및 기관들, 특히 대량의 서류업무의 통제가 필수적인 은행, 신용카드 및 보험업에서 찾아볼 수 있다. 이들 분야에서, 이미징은 하드카피 파일과 마이크로폼을 대체하고 있다. 비록 마이크로폼이 비용효율적이고 - 알맞게 통제되는 환경을 가정할 때 - 수명이 길지만, 또한 단점들도 있다. 전문적인 (이용자들에게 별로 인기가 없는) 마이크로폼 판독기가 필요하고, 자료에 대한 접근이 필름이나 피쉬 및 판독기가 소재한 곳으로 제한된다는 것 등이다.

의학계 또한 이미징 기술을 이용하는 데에서 선도적이다. Picture Archive and Communications Systems (PACS)는 병원

들에서 방사선 이미지들의 촬영, 저장, 및 전송용으로 신속하게 필름을 대체하고 있다. 포괄적인 병원정보 시스템을 제공하기 위하여 환자차트 정보를 찍고, 그것을 엑스레이 및 다른 의료기록들과 결합하는 임상기구 시스템을 포함하여, 다른 응용들 또한 일반적이 되어가고 있다.

15년 동안, 도서관들과 문서보관소들은 취약한 문서들을 보전 및 보존하기 위하여, 또한 마이크로폼보다 효율적인 매체로 소스문서에 대한 접근을 제공하기 위하여 전자적 이미징을 이용해 왔다. 그러나 지난 이년여 동안 이미징을 수행한 도서관 및 문서보관소들의 수는 200여개 이상으로 증대되었다.

2.6 이미징 대 마이크로필름화

이미지 찍기의 비용은 마이크로필름화보다 비싸다. 그러나 마이크로필름화가 융통성이 없는 한편으로, 이미징 기술은 얼룩의 제거나 이미지의 다양한 조정 등과 같이 이미지를 편집하는데 많은 선택권을 제공한다. 마이크로필름화와 이미징 사이의 비용격차의 대부분을 대표하는 것은 이미지의 향상능력이다.

저장된 이미지들은 원거리의 이용자들을 포함하여, 동시에 여러 이용자들이 접근할 수 있다. 이미지들은 만약 전송매체가 용량 조건을 처리할 수 있으면 원격으로 검색될 수도 있다. 비록 이미지들이 개별적 단어나 구로는 탐색될 수 없다 해도, 이미지 레코드 세트의 재편집은 언제든지 가능하다.

마이크로필름으로부터 프레임은 특수한

촬영 기구를 이용하여 디지털 이미지를 만들어내도록 처리될 수 있다.

이미징 시스템의 장점

- 요약하자면, 이미징 시스템은 여러 장점들을 갖고 있다. 그중 가장 중요한 것은 소스문서에 대한 개선된 가독능력, 조밀한 저장능력, 접근의 속도 및 어디서든 복수의 이용자들이 동시에 이용가능한 것들이다.

2.7 이미지의 유형

이미지에는 여러 가지의 유형들이 있다: 썸네일(thumbnail), 서비스 및 아카이벌(archival). 썸네일 이미지는 전형적으로, 이용자들이 고품질의 이미지로 접근하기 위하여 시간을 들이고 싶은가를 판단하도록 만들어진 작은 이미지이다. 서비스 이미지는 대부분의 이용자들의 수요에 맞추는 정보를 전달하도록 디자인되어서, 썸네일 이미지보다는 고해상도를 갖는다. 그것은 보통 저장 공간을 절약하고 전송을 위한 대역폭을 줄이도록 압축되어 있다. 썸네일보다 검색에는 시간이 오래 걸린다. 아카이벌 이미지는 보통 압축으로 인한 가공이 없는 압축되지 않은 이미지이다. 이들은 최고 품질의 이미지이지만, 썸네일이나 서비스 이미지들보다 더 많은 저장공간을 요하고 검색 속도가 느리다.

2.8 촬영

지금은 200 dpi(dots per inch)부터 20,000 dpi까지의 해상도 범위로 매 몇 초당 한 장의 속도로 이미지를 찍는 기구들

을 구할 수 있다. 전자는 팩스기계로 나온 것과 동등하고, 후자는 고급의 의학용 그림과 동일하다. 촬영 기구는 낮은 해상도로 30초당 소스문서 하나를 찍는 저렴한 데스크탑 모델(\$400)에서부터, 20,000dpi까지의 해상도로 삼초마다 스캔하는 고가의 제작중심의 기계들(\$40,000 까지)도 있다.

2.9 편집

빈번히 소스 문서, 사진, 지도 및 예술작품은 상태가 좋지 않다. 약할 뿐 아니라, 대비, 색상 및 선명도를 잃었을 수도 있다. 그러나 편집함으로써 소스문서의 물리적 교정 뿐 아니라, 전자적 교정까지도 가능하다.

고해상도 모니터(1,600x1,280dpi)와 알맞은 소프트웨어를 갖는 고성능 PC는 가독성을 향상시키도록 이미지를 편집할 수 있다. 이미지는 망가진 것을 복구하고, 얼룩을 제거하거나, 읽기 어려운 부분들을 나타나게 해준다. 비록 대부분의 비즈니스들은, 표준의 비즈니스 문서들과 서신들을 스캐닝하기 때문에, 이미지를 향상하려고 하지는 않지만, 많은 도서관과 문서보관소들은 질이 좋지 않은 원본들을 교정하고 있다. 도서관과 문서보관소에서 이미지를 편집하는 데 보통 사용되는 소프트웨어는 전문적인 사진사들을 위하여 개발된 Adobe Photoshop이나 Corel Photo-Paint이다.

편집은 디스큐잉(de-skewing: 이미지를 바로잡기)과 크라핑(cropping: 불필요한 테두리를 제거하기), 처럼 소스문서의 외형은 변화시키기 위해서는 아무것도 하지 않는 것으로 제한될 수도 있다: 혹은 밝기, 대비

를 개선하기, 디스펙클링(despeckling: 변색과 흠집을 제거하기) 및 이미지의 질을 개선하기 위한 다른 방법들을 포함할 수도 있다.

2.10 저장과 압축

모든 이미지는 상당한 양의 저장용량을 필요로 한다. 우표크기의 썸네일 이미지는 보통 각각 10에서 35 Kb이고, 신속하게 검색되고 디스플레이된다. 전형적으로 JPEG라는 포맷을 사용하여 압축되는, 서비스 이미지는 각각 300 Kb까지를 필요로 하며, 몇 초안에 검색될 수 있다. 최고급의 이미지인 아카이벌 이미지는 많은 양의 저장 공간을 요한다(각각 몇 Mb), 그래서 종종 자기(磁氣) 디스크 드라이브보다는 CD-ROM이나 DVD에 저장된다.

단일 18 GB 자기 디스크 드라이브는 300만에서 750만 표준크기 페이지 혹은 180장의 예술작품 컬러사진에 맞먹는 양을 담을 수 있다. 그 능력은 디스크의 저장용량 뿐만이 아니라 디지털화된 이미지의 복잡한 압축기법에 기인한다. 이것은 이미지들이 보통 텍스트 파일보다 훨씬 많은 기억용량을 요하기 때문에 중요하다. 압축하지 않고는, 예를 들어, 큰 기술도면 하나가 정보의 8 Mb를 차지하고, 예술작품의 고해상도 이미지는 100 Mb까지를 필요로 할 수 있다.

압축은 무관한 정보를 제거하고 화소(pixel) -- 이미지의 기본요소들 --의 중복되는 패턴을 일종의 코드화된 전자적 shorthand로 응축하기 위하여 컴퓨터를 사용함으로써 달성될 수 있다. 이 방식으로

큰 기술도면이 8 Mb에서부터 300 Kb이하로 축소되면서 이미지는 이 20~30배 압축될 수 있다.

2.11 조직

대부분의 이미징 시스템은 색인 기능을 포함하고 있어, 이미지들이 식별될 수 있다. 비즈니스에서는 색인이 보통 구매주문번호, 보험계약번호, 이용자번호, 혹은 날짜 및 이름이나 디스크립터이다. 도서관들과 문서보관소들은 전형적으로 분류번호나 파일번호 및 주제를 사용한다. 그러나, 많은 도서관들은 그들의 자동화된 도서관 시스템에 완전한 MARC 서지 레코드를 작성하고 이들을 이미지에 연결하기를 택하고 있다. MARC는 소스문서로부터 찍힌 이미지에 공통되는 정보들 중의 어떤 것들은 다루게 되어있지 않기 때문에 한계점이 있다. 디지털 형태로 된 모든 형태의 정보를 위하여 새로운 편목 관행을 개발하려는 작업이 진행중이다.

MARC의 두 가지 대안은 더블린코어와 EAD(Encoded Archival Description)이다. 둘다 훨씬 융통성이 있지만, 그 어느 것도 표준이거나 널리 채용되지는 않는다, 따라서 도서관들이 MARC 서지 레코드를 이용해 오던 기존의 방식으로 유니온 데이터베이스를 구축하기는 어렵다. 더블린코어는 2001년에 NISO(National Information Standards Organization) 표준이 될 것으로 기대된다.

2.12 전송

이미지의 전송은 자주 최근의 네트워크

의 능력에 도전한다. 이미지가 지금 TCP/IP와 Z39.50 컴퓨터간 프로토콜과 같은 표준 파일 프로토콜을 거쳐 네트워크를 통해 전송되고 있다고는 해도, 많은 네트워크들은 적당한 시간 내에 이미지를 전송하는데 필요한 만큼 큰 대역폭을 갖지 못한다. 56 Kbit (초당 56,000 비트) 라인을 통해서 한 장의 아카이벌 이미지를 전송하는 것은 18분까지도 소요된다. 그런 속도로 전송되는 다중의 이미지 파일을 기다릴 준비가 된 이용자들이 별로 없다. 다른 한편으로, 씬 네일은 일초 이하, 그리고 서비스 카피는 오초 정도까지 필요로 할 것이다. 그래서 아카이벌 이미지에 대한 접근이 인터넷보다, 광대역 LAN을 통해서 접근하는 것으로 제한되는 경향이 있다.

이미징 프로그램을 추구하는 기관들은 최소한도로 10 Mbps의 대역폭 및 이상적으로는 100 Mbps를 지원하기 위하여, LAN을 디자인해야 한다.

2.13 접근

도서관이나 문서 보관소는 원거리 이용자에 의한 모니터의 선택을 통제할 수는 없다고 해도, 자기 시설(들)내에서의 이용자들이 이용할 수 있는 고해상도 모니터를 갖추도록 계획해야 한다. 최소한도로, 그들은 적어도 1,024x768줄의 해상도를 지원하는 19인치 컬러 모니터이어야만 한다. 많은 오래된 PC 모니터들은 훨씬 낮은 해상도로 이미지를 디스플레이한다. 따라서 대부분의 스캐너로 달성된 해상도는 상실된다.

프린팅의 선택도 또한 다양하여, 300에서 600 dpi를 갖는 모노크롬용 데스크탑 레

이저 프린터300으로부터 훨씬 비싼 회색조 및 컬러 레이저 프린터들까지 있다.

2.14 이미징의 문제들

전체적인 이미징에서 가장 절박한 문제는 표준의 영역이다. 위에서 지적한 것처럼, 데이터 압축 기법은 이미지 저장을 줄이는데 중요하며, 표준 압축 알고리즘이 필수적이다. 이를 인식하여, Joint Photographic Experts Group(JPEG)과 Motion Picture Experts Group(MPEG)은 정지 및 동영상을 위한 표준들을 채택하였다. 그러나, 다른 비견할 만한 표준들이 있다.

이미지의 전송을 위한 표준들 또한 필요하다. 현재는 Tagged Image File Format (TIFF)가 특히 PC 시장내에서 흔히 사용되고 있고, PDF도 그것이 원래 생겨난 웹상에서만 아니라 기관내의 시스템에서도 점차 많이 사용되고 있다.

디스플레이 기구를 위한 색깔 표준은 컬러 이미징이 실용가능하기 위하여 필수적이다. 그것들은 아직 개발중이다.

컴퓨터 시스템과 네트워크 설계는 저장된 이미지에 대한 개방적인 접근을 지원하기 위하여 발전될 필요가 있다. 보다 분산된 형태의 컴퓨팅, 특히 사용자-서버 설계를 지향하는 현재의 경향은 고속 데이터 전송에서의 진보와 결합되어, 이미지들이 여러 어플리케이션들을 거쳐 통합될 수 있도록 허용하게 될 가능성이 많다.

비용 또한 문제로 남아 있다. 비록 스캐닝 도구들이 지난 오년간 대단한 기술, 가격 및 성과에 도약을 이루었지만, 이미지 데이터베이스를 만들어내는 것과 관련된

시간과 비용은 아직도 높다. 대량의 텍스트가 아닌 데이터를 위한 데이터베이스들에게는 특히 그렇다.

2000년 현재, 비용효과 분석이 이미징의 가치를 입증한 유일한 분야들은 수백만 페이지의 레코드들을 포함하는 은행, 재무, 보험 및 기타 응용들이다. 그래도, 더 작은 시스템들이 보다 대중적이 되고, 가격들이 낮아지고 있다. 또한 더많은 -- 자동화된 도서관 시스템 판매자들을 포함하는 -- 일반적 목적용 시스템의 벤더들이 이미징에 대한 기능성을 향상하기 시작하고 있다.

어림잡아, 도서관이나 문서보관소는 혼합적인 소스문서 (텍스트 및 다양한 크기와 상태의 사진들)를 포함하는 전형적인 이미징 프로젝트는 이미지 당 평균 \$6의 비용이 들 것을 기대해야 한다. 소스문서의 마이크로필름이나 피쉬가 이용될 수 있으면, 디지털로 이미지를 찍기는 이미지당 \$0.20까지 낮은 비용이 든다; 그러나, 만약에 찍은 후에 이미지가 향상되어야 한다면, 상당한 추가 비용이 발생한다. 이미지가 판권으로 보호되고 있는가를 결정하고, 만약에 그렇다면, 판권소유자에게 적당한 수준의 보상금을 확정하는 것 또한 처리되어야 한다.

2. 15 터키 시스템의 이용가능성

은행, 보험회사 및 병원들을 위한 상품들을 포함하여, 많은 전문화된 이미징 시스템들이 시장에 나와 있다. 특정하게 도서관과 문서보관소들을 위하여 디자인된 시스템들이 바로 지금 등장하기 시작하고 있다.

이 보고서는 도서관들과 문서보관소들

에 의하여 사용할 가능성이 있는 -- 하드웨어, 소프트웨어, 장치, 훈련 및 지속적인 지원을 함께 가져오는 벤더에 의하여 제공되는 -- 터키시스템에 초점을 둔다.

2.16 문서 이미지 관리 시스템의 유형

다른 기관들마다 다른 정보 저장과 검색 요건을 갖고 있기 때문에, 다른 유형의 문서 이미지 관리시스템들이 있다. 그것들은 다음과 같이 분류될 수가 있다:

- 데스크탑 시스템은 단일 사용자 응용을 위하여 디자인되었다. 이들은 적은 수의 이미지들 -- 보통 수천 이하 -- 을 처리한다. 그들은 보통 고성능 PC에 구성된다. 비록 가격이 \$10,000 이하에서 시작하지만, 한번에 단 1명의 사용자만이 사용할 수 있기 때문에 별로 좋은 선택이 아니다.
- 엔트리 레벨이나 중간범위 서버에 구성되는, 부문 시스템(departmental system)은 5명에서 100명까지의 동시적인 사용자들 및 수십만 이미지를 처리한다. 이 시스템들은 보통 사용자들이 LAN, WAN이나 인터넷을 거쳐서 접근할 수 있도록 지원한다. 가격은 \$250,000에서 시작하여 그 이상까지 갈 수도 있다.
- 범기업적 시스템은 큰 기관의 문서들과 사용자들을 모두 수용하도록 디자인되었다. 이들은 수백명의 동시 사용자들 및 수백만의 이미지들을 수용할 수 있다. 이 시스템들은 전형적으로 초마이크로-, 미니 혹은 메인프레임에 기반한 것들이다. 이미지 워크스테이

선이 직접 연결되거나 LAN으로 통할 수도 있다. 여러 장소들이 WAN으로 연결되어지기도 한다. 인터넷 접근도 가능하다. 가격은 \$250,000 이상에서 시작한다.

2.17 문서 이미지 관리 시스템을 고르기

이미징 시스템과 벤더들을 평가하기 전에, 자신의 응용 프로그램을 철저히 이해하는 것이 필수적이다. 문서 이미지 관리 응용 프로그램들은 두 가지 유형으로 범주화될 수 있다:

- 워크플로우 혹은 거래 처리 시스템: 이 시스템들은 문서들의 처리에 중점을 두고, 그 과정을 자동화하는 것을 목표로 한다. 보기로는 구매주문, 인보이스, 신용카드요금 및 보험계약 등을 포함한다.
- 저장 및 검색 시스템. 이 시스템들은 특정하게 다량의 문서들을 저장하고 검색하도록 디자인되었다. 이 문서들은 다양한 유형과 포맷들일 수 있다.

이 연구에서는 도서관 및 문서보관소들에서 중요한 응용이 되는 후자에 초점을 둔다. 불행히도, 대부분의 문헌들은 이 기술의 주요한 이용자들인 재무, 보험 분야 및 정부의 이용자들에 우선순위를 두고 있기 때문에, 워크플로우 혹은 거래 처리 시스템을 강조하고 있다.

다음에서, 이미징 프로젝트의 목적이 보존인지 개선된 접근인지를 결정해야만 한다. 전자는 이미지를 찍고 약간의 변형만으로 혹은 변형 없이 저장하는 것을 강조한

다; 후자는 이미지를 찍고, 외형을 변형하는 대가를 치르더라도 가독성을 촉진하기 위하여 향상시키는 것을 강조한다.

이미징 시스템의 규모는 접근을 목표로 할 때에 현저히 커지게 된다. 왜냐하면 보통 수요가 있을 것 같은 소스문서의 이미지를 찍는 것이 그 우선순위이기 때문이다.

2.18 기획의 중요성

이미징 기술의 투자는 단지 숫자 상의 문제는 아니다. 진보된 기술이 어떻게 소개되든, 필요한 상당한 기획과 연장된 수행 스케줄이 있다. 문서 이미징은 단순히 새로운 도구이기보다는 새로운 과정을 대표한다. 비록 기술적 전문성이 이 시스템들을 디자인하고 설치하는데 필요하지만, 성공을 위해서는 조직의 목표와 인사 문제를 이해하는 것도 마찬가지로 필수적이다.

3. 기술 기반

여기에서는 간략하게 이미징에서 중요한 일곱 가지 기술 기반들을 소개한다: 그래픽, 이미지 유형, 해상도, 명료성 (legibility), 파일 형식, 이미지 촬영 인터페이스 및 이미지 드라이버.

3.1 그래픽

앞에서 디지털 이미징과 그래픽을 구별하였다. 디지털 이미징과 그래픽의 보다 기술적인 구별은 래스터 그래픽 (Raster Graphics)과 벡터 그래픽 (Vector Graphics)로 할 수 있다.

(1) 래스터 그래픽

래스터 그래픽(보다 적합한 용어는 래스터 이미징)에서, 이미지는 일련의 점들로 나타내진다. 래스터 유닛은 스크린 상에서 구별할 수 있는 점이다. 가장 단순한 형태의 래스터 그래픽은 그림이 각 점이 0(흰색을 대표)이나 1(검은색을 대표)을 의미하는, 이차원적 점의 배열로 되어 있는 모노크롬 (bitmapped) 이미지로 되어 있다. 대부분의 이미지 촬영 기구 -- 및 모든 팩스 기계들 --는 비트맵된 래스터 이미징을 사용한다. 이 상황에서, 해상도는 그림의 매 인치를 대표하도록 사용되는(화소라고 불리우는) 점들이다; 인치 당 점이나 화소가 많으면 많을수록, 디지털 표현은 원래의 이미지를 더욱 정확하게 묘사한다.

(2) 벡터 그래픽

벡터 그래픽은 기본적인 구성 블록으로서 줄과 원 같은 도표적 요소를 채용하고, 컴퓨터-지원 디자인(CAD) 같은 응용 프로그램들에 많이 사용된다; 이 응용은 그래픽이라고 대중적으로 알려져 있으므로, 이 연구에서는 언급하지 않기로 한다.

3.2 이미지 유형

주요한 이미지의 유형(bit depth라는 용어로도 쓰인다)은 5가지가 있다: 1-비트 흑백 혹은 bi-tonal, 4-비트 회색조 (grayscale), 8-비트 회색조, 8-비트 컬러, 24-비트 컬러. (32-비트 컬러는 막 등장하기 시작했다). 비트는 단일 비트가 0이나 1로 대표되는 기본 단위이다.

1-비트 흑백 이미지는 비트가 흑이거나

백일 수 있다는 것을 의미한다. 이것이 bi-tonal이다. 흑과 백 사이에 음영은 없다. 실제로, 이 유형은 그림 없이 인쇄된 텍스트를 저렴하게 이미징하는 것에만 사용된다.

4-비트 회색조 이미지는 그림이 없는 대부분의 인쇄된 책자에 적합한, 16-음영의 회색조를 포함한다. 이는 명목상으로 흑백으로 된, 그러나 실제로는 음영과 잉크의 밀도와 종이의 색조의 다양성을 갖는 고서들에도 적합하다.

8-비트 회색조 이미지는 256 음영의 회색을 포함한다. 이것은 고문서들과 상당한 음영을 갖는 인쇄된 텍스트들을 위하여 가장 흔하게 사용된다.

8-비트 컬러는 8-비트 회색조와 비슷하지만, 각각의 비트가 256색들 중의 하나일 수 있다. 이것은 질보다 경제성이 더 중요할 때에만 24-비트 컬러를 대신하여 사용된다. 저장이 저렴해진 지금, 스캐닝에서는 빈번하게 사용되지 않지만, 서비스 카피를 저장하는데에는 종종 사용된다.

24-비트 컬러는 16.8백만 색들을 포함한다. 이것은 각 비트가 16.8백만 색들 중의 하나를 포함하는 잠재력을 갖고 있어, 적당한 비용으로 고급의 질을 제공하기 때문에, 도서관들과 문서보관소들에서 이미지 찍기를 위한 이상적인 이미지 유형이다. 24-비트 컬러로 일단 찍히면, 이미지는 24-비트 컬러는 archival 카피로 유지해놓고, 사용하기 위하여 8-비트나 회색조로 복사될 수 있다.

• 지금 32- 및 42-비트 컬러도 있다. 그것들은 주로 전문적인 사진작가들 및 그래픽

픽 아티스트들이 사용한다.

Bit depth는 스캐너가 찍을 수 있는 색깔의 숫자만을 의미하는 것이 아니라, 열거될 수 있는 단계들의 최대한의 숫자도 의미한다. 더 많은 비트들은 일반적으로 프린터가 인쇄할 수 없거나 스크린이 모든 색깔들을 디스플레이할 수 없을 때일지라도, 더 나은 이미지를 재생할 수 있다.

- 촬영 기구의 이론적인 능력이 완전히 달성되는 경우는 드물다. 예를 들어, 24-비트 촬영 기구는 색깔에 16.8백만의 값들 중 하나를 부여하는 능력을 갖는 것으로 표현된다. 그러나, 실제로는, 가장 좋은 것도 262,144 색깔들의 값 -- 전형적인 컬러 사진의 동적범위에 모자라는 -- 으로 제한되어 있다. 이런 이유로, 전문적 사진작가들은 32-비트 촬영 기구들, 심지어는 42-비트 촬영 기구들을 이용하기 시작했다.

3.3 해상도(Resolution)

해상도는 인치당 점의 수(dots per inch: dpi)로 결정된다. 인치당 점이 많을수록, 더 많은 정보가 이미지에 관하여 저장되어진다. dpi가 높을수록, 저장되어야 있는 파일의 크기가 커진다. 100 dpi의 1-비트 흑백 이미지는 겨우 10Kb의 저장공간만을 필요로 한다; 400 dpi의 24-비트 컬러는 475 Kb의 저장공간을 필요로 한다.

현재 600-dpi 촬영 기구들 및 저장은 매우 저렴하여, 100, 200, 300, 혹은 400 dpi로 이미지를 찍고 저장하는 것에 대한 요구는

많지 않다. 큰 문제는 그것을 600 dpi로 할 것인가 1,200 dpi로 할 것인가이다. 코넬, 옥스퍼드, 및 버지니아 같이 유명한 디지털화된 센터들의 대부분은 600 dpi로 촬영하고, 매스터 카피는 오프라인으로 CD-ROM에 저장하고 서비스용으로는 400-dpi 카피를 만든다. 그러나 점점 많은 도서관과 문서보관소들이 색깔을 포함하는 소스 문서들을 갖고 작업할 때, 1,200dpi로 촬영하고 있다.

- 명료성은 bit depth와 해상도의 결합의 결과이다. 8-비트 화소는 1-비트 화소보다 더 많은 정보를 찍는다, 따라서 흑백조로 하기보다 회색조로 낮은 해상도를 사용하는 것이 동등한 정도의 가독성을 달성하는 것이 가능할 수도 있다.

3.4 파일 포맷

이미지 포맷은 거의 어떤 플랫폼이나 소프트웨어 시스템으로 이동하기 때문에, 네 가지 공통적인 것이 있다: TIFF, JPEG, GIF, 및 PDF.

TIFF(Tagged Image File Format) 파일은 파일의 매스터카피용으로 널리 사용되고 있다. 일단 저장되면, TIFF 파일은 다른 하드웨어 및 소프트웨어 시스템을 갖는 컴퓨터로도 검색하고 읽을 수 있다. TIFF 이미지는 조작하기가 쉽다. 압축될 때에도, 정보를 잃어버리지 않는다 (무상실 압축: lossless compression).

JPEG(Joint Photographic Experts Group)파일은 공간 제약이 있는 시스템들을 통한 웹 보기와 이동에 널리 사용된

다. 파일이 압축되면, 약간의 정보를 잃어버리게 되므로, 압축은 손실이 있다. 그러나, 이미지가 질적으로 현저히 축소된다는 것을 의미하지는 않는다. 만약 그것이 24-비트 컬러로 스캔이 되면, 그 손실은 사람의 눈에는 띄지 않는다. JPEG 이미지는 질을 잃어버리지 않고 확대될 수는 없다. 비록 JPEG 이미지가 아카이빙의 목적으로는 대중적이지는 않지만, 서비스 카피로는 흔히 사용된다. TIFF 파일로부터 JPEG 카피를 만드는 것처럼 스캐닝할 때 TIFF와 JPEG파일을 둘다 만들어내는 것이 흔하다.

GIF(Graphic Interchange Format)는 256 색깔에 제한된 더 오래된 형식이다. TIFF 처럼, GIF도 손실 없는 압축형식을 사용할지만, 더 적은 저장공간을 필요로 한다.

모든 전술한 이미지 형식들이 국제 표준에 바탕을 두고 있지만, PDF (Portable Document Format)는 Adobe가 개발한 특허된 형식이다. PDF는 Adobe Acrobat Reader의 매력적인 특징들 때문에 웹에서 널리 사용되고 있다, 따라서 이미징 시스템의 널리 보급된 형식이 되어가고 있다.

3.5 이미지 촬영 인터페이스

이미지 촬영 인터페이스 -- 전형적으로 스캐너 --는 어떤 다른 주변기기들과 마찬가지로 PC와 인터페이스되어야 한다. 여러 가지 선택들이 있다: IDE, SCSI, SCSI-2, SCSI-3, USB 및 Firewire.

(1) IDE

IDE는 PC와 그것의 주변기기 구성요소

들 사이에 가장 흔한 인터페이스이다. 이것의 주요한 장점은 저렴한 가격이지만, 모든 선택권들 사이에 가장 열등한 탐색 시간을 제공한다. 보통 PC에서 조작되는 데이터보다 훨씬 많은 바이트를 포함하는 이미지로 작업할 때, 보다 저렴한 비용은 데이터가 이미지 촬영기구로부터 PC까지 천천히 움직일 때 야기되는 낮은 생산성을 상쇄하는 것 이상이다.

(2) SCSI

SCSI는 여러 유형으로 나온다: SCSI, SCSI-2, SCSI-3. 모두 IDE보다 신속한 탐색 시간을 제공하지만, 비용이 더 들고 설치하기가 훨씬 어렵다. 그럼에도 불구하고, SCSI가 이미징 응용프로그램으로서는 훨씬 좋은 선택이다.

SCSI 인터페이스를 갖는 대부분의 제품들은, 또한 Ultra2로도 알려져 있는 SCSI-2와 같이 나온다. 이것은 초당 40Mb로 데이터를 이동한다. SCSI-3이나 Ultra3은 초당 160 Mb로 이동한다. 거의 모든 이미지 촬영 기구들은 SCSI-2나 SCSI-3를 지원한다.

(3) USB

USB(the universal serials bus)는 아직은 초창기이다. 이것은 데이터를 초당 최고 15Mb의 속도로 전송하며, SCSI-3이나 Firewire보다 상당히 느리다. 이것은 키보드, 마우스, 타블렛, 모뎀, 전화, CD-ROM 드라이브, 프린터 및 기타 저속-내지 중속 외부 주변기기를 총별 스타 토폴로지로 연결하도록 디자인되었다. 이것의 주요한 장

점은 설치가 쉽고, PC 뒤에 많이 늘어져 있는 케이블들을 제거할 수 있다는 것이다. 소비자용 모델 이미지 촬영 기구로는 적합하지만, 전문적 모델로 사용하기에는 적당하지 않다.

(4) IEEE 1394

IEEE에서 채택, 유지하는 이 표준은 원래 Apple이 개발하였다. 종종 이 표준에 잘못 적용되는 이름인 FireWire는 Apple이 표준에 적용시킨 것이다. IEEE 1394, 최근에는 IEEE 1394b판이 나왔는데, SCSI보다 몇배나 빠르게 초당 3.2Gb까지의 속도로 데이터를 전송할 수 있게 한다. 최고의 데이터 전송 속도는 트윈피이버 케이블연결을 사용해서만 가능하다. 표준 4쌍 케이블은 초당 200Mb까지 지원한다. 모든 PC들이 다 IEEE 1394를 지원하지는 않는다.

3.6 이미지 드라이버

이미지 드라이버는 소프트웨어 응용과 소통할 수 있기 위하여 이미지 촬영 기구에 필요하다. 이미지 드라이버용으로 ISIS와 TWAIN이라는 표준이 있다. 이미지 드라이버가 없이는, 촬영된 이미지를 향상시키거나 저장하도록 워크스테이션으로 이동하기 위하여, 작업 중인 응용프로그램을 나와, 하드웨어 드라이버를 찾아서 열고, 기구의 옵션들을 정하고, 이미지를 얻고, 디스크에 저장하고, 하드웨어 드라이버를 닫고, 응용으로 돌아가고, 그런 후에야 디스크로부터 이미지 파일을 찾아서 읽을 수 있다.

(1) ISIS

ISIS는 Cornerstone Corp.의 자회사인 Pixel Translation가 개발한 특허제품이다. 최근까지, 거의 모든 전문적 수준의 이미지 촬영 기구는 ISIS만을 지원했고, Pixel Translation은 판매되는 거의 모든 이미징 제품에서 인세를 받았다. 지난 이년간 많은 스캐너, 소프트웨어, 및 보드 벤더들이 인세를 지불하지 않고 자기들의 제품들을 통합하기 위하여 산업의 모든 사람들이 이용할 수 있는 표준을 창안하도록 함께 모였다. 이들은 많은 소비자-수준의 스캐너에서 사용되는 TWAIN 이미지 드라이버 프로토콜을 감독하는 그룹인, TWAIN Working Group에게 접근하였다.

(2) TWAIN

TWAIN 1.7로 지정된, 전문적 수준 이미지 촬영 기구용 TWAIN 프로토콜은 1999년에 채택되었다. 불과 몇 달안에 대부분이 자기들의 최신 제품 생산 라인을 TWAIN과 일치하게 만들었다. 비록 소수가 ISIS를 옵션으로 제공하는 하지만, TWAIN이 최근의 표준이 되는 추세이다.

4. 사례연구

4.1 선구적 노력들

이 기관들은 5년 이상 전에 이미징 프로젝트를 수행했던 선구적인 사례들이다.

- (1) 콜럼비아 대학교 에이버리 도서관 (Columbia University Avery Library)

도서관에 의하여 맨 처음에 수행된 이미

징 프로젝트들 중의 하나는, 1985년에 콜럼비아 대학교의 에이버리 도서관에서 추진된 프로젝트 AVIADOR(Avery Videodisc Index of Architectural Drawings on RLIN) 이다. 이 프로젝트의 주요한 목표는 41,000의 건축도면들에 대한 기계가독형 편목정보를 제공하고 각각의 도면의 이미지를 전자적 저장 기구에 통합한다는 것이었다. 그 의도는 이용자들이 모두 원본을 손대지 않고도, 전체의 소장자료들을 뒤적이거나 직접 특정한 도면으로 갈 수 있게 하는 것이었다. 콜럼비아의 건축도면들의 역사적 컬렉션에 대한 접근이 현저히 개선되었을 뿐만 아니라, 이 프로젝트는 그것들을 보존하는데에도 도움이 되었다.

초기의 프로젝트에서, 프랭크 로이드 라이트 (c.546 항목), 라이트의 선배 루이스 헨리 설리번 (c.175 항목), 대중적인 건축가 휴 페리스 (c.310 항목) 같은 건축가의 작품들 및 맥킴, 미드 & 화이트 (c.560 항목)의 샘플을 포함하는, 도면들은 에이버리에 소장된 40개의 뛰어난 고문서 컬렉션을 대표한다. 프로젝트 동안 편목된 가장 작은 컬렉션은 제임스 켄워 주니어 (58항목)이고, 일라이 잭크 칸 컬렉션로부터의 8,248 도면들은 오랜 기간 동안의 그의 회사의 발전을 대표한다.

AVIADOR는 이미징 기술이 아직 새로운 것일 때 추진되었다. 비교적 새로운 디지털 스캐너와 광학 저장기구에 의존하기보다, 나중에 고품질 사진이나 네거티브로부터 디지털이미지를 만들어낼 의도로, 자료들을 촬영하고 아날로그 비디오디스크에 저장하도록 결정되었다.

AVIADOR은 RLIN의 메인프레임에 연결된 PC에 사용하도록 디자인되었다. RLIN은 비디오디스크에 있는 도면의 41,000 정지화면 이미지의 각각을 RLIN의 적합한 목록레코드로 연결해주는 인터페이스를 위한 프로그램을 고안하였다.

이용자들은 다음과 같이 이용할 수 있었다:

1. 다양한 텍스트 색인 (건축가, 제목, 건물명, 장소 등)을 이용하여, RLIN에서 알맞은 서지 레코드(들)을 탐색한다.
2. 원하는 레코드가 식별되고 IBM PC 모니터 상에 디스플레이 될 때, RLIN 레코드내에서 특정한 도면의 등록번호에 커서를 놓는다.
3. 그리고 나서 IBM PC의 적당한 기능 키를 누름으로써 비디오디스크 모니터에 일치하는 시각적 이미지를 불러들인다.

이용자들은 대신 이렇게 할 수도 있다:

1. 비디오디스크에서 이미지를 스캐닝함으로써 그래픽 색인으로 시작한다.
2. 원하는 이미지가 식별되고, 비디오 모니터 상에 디스플레이될 때, IBM PC 키보드에 적당한 기능 키를 누름으로써 일치하는 서지 레코드를 검색한다.

50 카피의 비디오디스크가 필요한 소프트웨어와 이용자 매뉴얼을 포함하여, 각각 \$500에 생산되었다. 이미지 파일은 나중에 디지털포맷으로 전환되어 자기디스크 드라

이브에 저장되었다.

(2) 포드자동차 기술 정보 센터 도서관
(Ford Motor Technical Information
Center Library)

자동화된 도서관시스템에 연결된 초기의 이미징 파일들 중의 하나는, 포드 자동차회사의 주요 과학 기술도서관인 포드자동차 기술 정보 센터 (TIC: Technical Information Center) 도서관의 것이다. 1989년에, 이 도서관은 이용자들에게 회사 기술 보고서의 전문에 대한 데스크탑 전자적 접근을 제공하고, 도서관의 온라인 목록과도 통합될 수 있는 문서-이미징 시스템을 RFP에 발주했다. 도서관은 이미 모든 도서관 기능들을 위하여 COMSTOW Information Service의 BiblioTech 온라인 통합 도서관 소프트웨어를 사용하고 있었다.

TIC 직원들은 포드자동차회사 내에서 많은 이용자들을 위한 회사보고서의 중앙보관소의 필요성을 오랫동안 인식하고 있었다. 도서관은 연구, 제품 및 생산기술 직원 보고서들의 컬렉션을 오랫동안 유지하고 있었으나, 다른 연구 혹은 기술그룹이 생산한 보고서들을 위한 중앙 장소는 없었다. 결과적으로, 가치있는 정보가 그래야 하는 만큼 널리 공유되지 못하였다. TIC 직원들은 광학 저장 및 이미징 기술을 도서관의 물리적 공간의 필요를 증대시키거나 정보를 탐색할 다른 장소를 추가하지 않고도 전체 기관에 회사의 기술정보에 대한 접근을 제공하는 방법으로서 보았다.

BiblioTech 시스템은 Digital Equipment Corp. (DEC) Micro Vax 3900에 32Mb의 RAM 및 2Gb 이상의 자기저장용량을 갖

고 설치되었다. 성공적인 입찰자는 디지털 스캐너를 사용하여 이미지를 촬영하고 CD-ROM에 그것들을 저장하는 것을 제안하였다. 도서관은 제안을 받아들이고, 나중에 BiblioTech 시스템과 디지털 이미지를 담고있는 PC-기반 CD-ROM 기구들 사이에 인터페이스를 설치하였다.

CD-ROM에 대한 접근경험을 갖고, TIC 직원들은 어떤 이미징 시스템도 기존의 정보관리 시스템 위에 세워지고, 완전히 통합되어야 한다고 결론을 내렸다. 포드의 경우에, 이것은 이미징 시스템이 BiblioTech 시스템에 대한 백엔드 시스템 역할을 해야 하고, BiblioTech는 색인-탐색 도구 역할을 해야 한다는 것을 의미했다. 문서가 BiblioTech에 배치될 때, 전문이미지의 이용가능성은 레코드에 표시되고, 간단한 키누르기의 연속으로 화면을 지우고 이용자에게 이미지를 가져오게 한다. 이용자는 그리고 나서 필요하면 왔다갔다하고 확대하면서 문서 사이를 이동할 수 있다. 문서를 본 후에, 이용자들은 자기들이 전에 이미지를 보고 있던 바로 그 지점에서 BiblioTech로 되돌아 갈 수 있다. 이용자들은 자기들의 워크스테이션에서 이미징된 문서를 프린트하거나 데스크탑으로 팩스할 수도 있다. 인터페이스가 매끄럽게 되기까지는 몇 년이 걸렸다.

(3) 국립농업도서관 텍스트 디지털화 프로그램
(National Agricultural Library Text Digitizing Program)

마이크로컴퓨터-기반 스캐닝 시스템은 1988년 1월에 NAL에 설치되었다. 4,000 페이지 이상의 판권이 없는, 양식 (養殖)에

관한 출판물들을 비트맵된 페이지 이미지와 ASCII 텍스트를 만들어내기 위하여 스캔하고 디지털화하였다. 텍스트는 Unibase의 TextWire Plus를 사용하여 색인되었고, 결과로 나온 데이터베이스 'Aquaculture I'는 1989년 3월에 참여하는 랜드그랜트 도서관들에게 CD-ROM으로 배포되었다. 도서관들의 역할은 전달매체, 검색소프트웨어 및 디스크 자체의 내용을 평가하는 것이었다. 추가로, ASCII 텍스트만으로부터의 검색과 페이지 이미지가 있는 ASCII 텍스트 검색의 비교도 수행되었다.

두 번째 디스크, 'Food, Agriculture and Science'는 Consultative Group on International Agricultural Research(CGIAR)에 의하여 생산되었고, 1989년 9월에 랜드그랜트 도서관들과 전세계의 CGIAR 사무소들에 평가를 위하여 보내졌다. 그것은 CGIAR 사무소들 자체에서 지정된 CGIAR 자료들을 수록하였고, Knowledge Access로부터의 KAware2 검색소프트웨어를 사용하였으며, 원문과 그래픽을 둘 다 포함하였다. 프로젝트는 매우 성공적이어서 CGIAR은 CD-ROM의 전체 시리즈 생산을 위촉하였다.

1990년, 세 번째 데이터베이스가 배포되었다: 버몬트 주립대학에서 제작된 캐나다인 산성비에 관한 간행물을 수록하고 있는 2장의 디스크 세트. 미국교육부로부터의 두 Title II-C 그랜트는 간행물들을 식별, 스캔 및 편목하기 위한 노력을 지원하였다. 이 자료들은 판권허가와 더불어, 캐나다 정부의 여러 기관들이 이용할 수 있었다. "산성비: 캐나다 정부문서"라고 제목이 붙은 이들 CD-ROM들은 KAware2 검색소프트웨어

어를 사용하고, 전문 및 전면 이미지를 포함한다.

네 번째 및 최종적인 평가 디스크는 NAL의 고엽제 에이전트 오렌지에 대한 대규모 특수 컬렉션으로부터 선정된 4100 페이지들을 포함하였다. NATDP에 의하여 제작되고 'Agent Orange'라고만 제목이 붙은 이 디스크는 Personal Library Software로부터 Windows Personal Librarian 검색소프트웨어를 사용하고, 선정된 문서들의 전문, 4100의 고해상도 페이지 이미지 및 200의 저해상도 페이지 이미지 샘플들을 포함하였다. 이것은 1990년 10월에 배포되었다.

'Food Irradiation'이라는 제목이 붙고 TextWare Plus 검색소프트웨어에서 운영되는 디스크는 1991년 6월에 마스터가 만들어졌고, 탐색과 검색을 위한 확장된 서지 레코드를 사용하였다. 평가 디스크에 포함된 원문 문서처럼 서지 레코드들은 연관된 페이지 이미지들에 링크를 갖고 있다. 이것은 검색을 위하여 서지 레코드를 사용한 NATDP의 첫 실험이었다. 이 디스크상의 5,100 페이지는 NAL에 소장된 대규모 컬렉션으로부터 왔고, 소스문헌들은 도서관에 시급한 공간을 해결하기 위하여 다른 기관에 기증되었다.

NATDP 평가연구 최종보고서는 전자출판물의 지속적인 제작을 추천했지만, 나중의 제품은 이미지 파일만을 포함할 것을 명기하였다: 원문파일 전환은 너무 시간소요가 많고 비싸다고 여겨졌다. NAL은 즉시 추가적인 전자출판물들에 착수하였다. 새 프로그램용으로 최신 검색소프트웨어, 메릴랜드주 락빌의 Personal Library

Software가 개발한 Windows Personal Librarian을 허가를 냈다. 1992년 5월에 나온 첫 디스크는 1907년부터 1924년까지의 *Agronomy Journal*, v.1-16까지를 포함하였다. 데이터베이스는 서지 레코드와 페이지 이미지로 구성되었고, American Society of Agronomy의 협조로 제작되었다.

1992년 말에 배포된 다음 디스크는 'Aquaculture II'로, 6,500 페이지 이미지를 포함하였다. 다음으로 카버의 논문, 편지 및 그림들의 마이크로필름 세 롤로부터 스캔한 3,500 페이지 이미지를 수록한 '조지 워싱턴 카버' CD-ROM이 나왔다. 이 마이크로필름은 터스키기 대학에서 제작한 67-롤 컬렉션의 부분이다. 이것은 도서관이 마이크로폼으로부터 스캐닝을 한 첫 주요한 프로젝트들 중의 하나이다.

NAL은 다음으로 이미지 데이터베이스들과 VTLS 자동화도서관 시스템과 링크를 활성화하기 위하여 두 개의 VTLS InfoStations를 설치하였다. InfoStation은 하드디스크의 2.4 Gb를 갖는 NEXT 워크스테이션으로 구성되었다. 이미지들은 CD-ROM으로부터 선택적으로NEXT 하드디스크 저장장치로 다운로드되었다. 이용자들은 VTLS 시스템에서 저자, 서명, 주제 혹은 기타 접근점으로 서지 데이터베이스를 탐색하고, 이미지 파일이 서지 레코드를 위하여 존재하는 것을 결정하고, 윈도우 디스플레이의 아래에 있는 'Retrieve Multimedia' 버튼을 클릭함으로써 인포스테이션 화면으로 이미지를 검색한다. 이용자들은 이미지 파일을 끝냈을 때, 서지 레코드 화면으로 돌아간다. 인터

페이스는 매끄럽게 되었다 -- 이용자가 다른 시스템에 접근하기 위하여 한 시스템을 나갈 필요가 없고, 자기의 위치를 놓치지 않는다. 그러나, 한 가지 단점은 이용자가 초기의 간략서지 레코드 디스플레이를 참조하여 이미지가 서지 레코드와 연결되어 있는지 알수 없다는 것이다; 이용자는 반드시 전문서지 레코드를 불러야만 한다.

(4) 피바디 대학 교육도서관(밴더빌트 대학교) (Peabody College Education Library (Vanderbilt University))

피바디 대학 교육도서관의 클리퍼 프로젝트는 이용자가 역동적인 참조 파일을 손쉽게 검색하게 해주는 맥에 기반한 광학저장 시스템이다. 클리퍼는 매킨토시를 파일 서버와 워크스테이션으로서 사용하였다. Micro Dynamics, Ltd.로부터의 MARS 소프트웨어는 이용자 각각을 위하여 조정된 인터페이스를 통해 문서의 저장과 검색을 관리하였다. 1988년에 설치된 시스템은 주요한 다중산업 이미징 벤더가 도서관에 설치한 몇 안되는 터키 시스템들 중의 하나이다. 프로젝트는 Apple Academic Development Donation Program으로부터의 하드웨어 그랜트에 의하여 부분적으로 지원되었다.

이미지들 -- 12,000 정도 --이 10개의 파일 캐비닛으로부터 스캔되었다: 신문 클리핑, 신문기사, 팜플렛, 연구보고서, 리포트 등. 많은 정보들이 테네시주 내쉬빌 혹은 남부 주들로부터 온 시사적이고 지역적인 것이다. 차트, 그래프, 도표 및 통계형태로 된 많은 통계적 정보들도 포함하였다. 또한

클리퍼 이전에는 단일 주제표목으로 개별적인 파일 폴더를 찾음으로써 이런 유형의 정보를 검색할 수 있었다.

클리퍼의 범위는 ERIC나 PsycLIT보다 훨씬 좁혀지고 훨씬 집중되었다. 이것은 주제에 대한 포괄적인 정보를 포함하지 않는다. 예를 들어, 체형(體刑)이라는 주제에 관심이 있었다면, 클리퍼는 메트로나 테네시주의 체형의 사용에 관하여 초점을 둔 기사들을 수록할 수도 있다. 그러나, 이용자는 아직도 이 주제를 완전히 연구하기 위해서는 다른 색인들이나 데이터베이스들을 탐색할 필요가 있다.

클리퍼의 주요한 요소들은:

- 입력 워크스테이션: 애플스캐너와 매킨토시 IIx로 구성. 문서들은 스캔되고, 이미지들은 하드디스크에 임시적으로 저장되고, 각 문서에 대한 기술적 정보는 워크스테이션에 입력된다.
- 광학 서버: 파이오니어 5.25 인치 Double WORM Drive가 부착된 매킨토시 IIx. 광학서버는 입력 워크스테이션의 하드드라이브로부터 WORM 디스크로 아카이브되는 스캔된 이미지의 저장과 검색을 관리한다.
- 디렉토리서버: WORM 디스크에 저장된 스캔된 이미지에 관한 기술정보를 저장하는 매킨토시 SE/30. 저장된 이미지에 대한 색인 역할을 한다.
- 검색 워크스테이션: 19인치 시그마 디자인 모니터와 검색된 이미지 프린트용 Laser Writer SC를 갖는 매킨토시 IIx. 워크스테이션은 참고자료실에 설치되었다.

- MARS(Multi-user archival and retrieval system): 매킨토시 네트워크에 문서이미지를 저장, 탐색 및 검색하도록 해주는 소프트웨어.
- 도서관 외부에서 사무실과 방들에서 클리퍼를 이용할 수 있게 해주는, 클리퍼시스템을 캠퍼스네트워크에 연결하는 Ethernet LAN.

(5) AT&T

1989년 이래로 AT&T Information Services Network는 내부의 기술메모를 일차적으로 400dpi로 스캔하고, WORM(write-once-read-many) 광학디스크에 이미지를 저장하고 있다. 수만의 문서들이 스캔되었고, 새로운 문서들이 계속해서 스캔된다. 일차적으로, 카피에 대한 수요는 중심지역에서 400dpi 프린터로 문서를 인쇄함으로써 충족되고, 회사메일을 통해서 보내진다. 비록 이 시스템이 -- 사무원이 파일링 캐비닛에서 원문서를 찾고 제록스카피를 만들었던 -- 그전의 방법에 비하여 많이 개선된 것이지만, 아직도 요청된 카피가 요청자에게 닿기까지는 며칠의 지체가 있다.

AT&T 네트워크 대역폭이 456 Mbps까지로 증대될 때, 이미지서버로부터 데스크탑 워크스테이션으로 이미지를 이동하는데의 모든 압박이 제거되었다. 이미지들은 현재는 회사네트워크를 거쳐 검색할 수 있다.

(6) 보울더 공공도서관

보울더 공공도서관 (Boulder Public Library: BPL)은 1992년에, CARL System

Public Access Catalog에 대한 확장으로, 온라인 목록에 사진 이미지 접근을 설치하였다. 보올더 공공도서관 기금에서 부분적으로 지원을 받아, 사진 이미지 접근 프로젝트 (Photo Image Access Project)는 시각자료들에 대한 새로운 유형의 접근을 제공하였다. 원래 BPL의 역사적 사진 컬렉션과 함께 사용되기로 디자인되어, 이 과정은 나중에 다른 시각자료 컬렉션과도 적용이 되었다.

과정은 사진 이미지를 BPL의 서지 데이터베이스에서 해당하는 MARC 레코드로 링크해주고, PAC 이용자들이 PC에 필요할 때마다, 이미지를 디스플레이하는 것을 허용하였다. BPL에서 프로젝트 직원은 컬렉션으로부터 스캔되고 BPL 시스템의 Bibliographic Maintenance를 거쳐 해당하는 MARC 레코드의 서지식별번호 (BID: bibliographic identification number)를 결정하도록 사진들을 선정하였다. CARL에 의하여 프로젝트를 위하여 개발된 차트에 따라 스캐닝 해상도를 결정한 후에, 직원들이 사진을 스캔하고, 고유한 파일에 저장하고, 파일을 보조기억장치로 옮기고, 알맞은 BID로 이름을 붙여 주었다. 이미지는 최소 32 회색조를 갖는 640x480화소로 압축된 형태로 저장되었다. 비록 컬러가 스캔하기에 더 느리고 더 많은 저장용량을 필요로 하긴 하지만, 이 과정은 컬러와 흑백사진들을 둘 다 지원한다.

사진 이미지를 보기 위하여, 이용자는 보통 때와 같이 PAC를 탐색한다. 이미지 디스플레이는 CARL Systems 대형 소프

트웨어와 VGA 모니터를 갖는 PC 워크스테이션을 필요로 한다. 그것에 이미지가 링크되어 있는 서지 레코드가 검색되었을 때, "이미지화된" 온라인 메시지는 짧은 디스플레이의 부분으로서 나타난다. 이용자는 그리고 나서 전문 서지 레코드를 선택하고, 추가적인 프롬프트에 응답한 후에, PAC 화면에 이미지를 디스플레이한다.

BPL은 사진 이미지를 자기 디스크에 지역적으로 저장하고, BPL의 지역 네트워크에 연결된 PAC 터미널에 이용할 수 있게 만들었다.

스캐닝 과정은 Microtek 300z 스캐닝 기구와 Picture Publisher 소프트웨어를 사용하였다. 스캐닝 소프트웨어는 마이크로소프트 윈도우 환경에서 운영된다. 스캐닝 스테이션은 기억용량 8 Mb와 200 Mb의 디스크 및 마우스를 갖는, 33 MHz 80386 PC-호환기종을 필요로 한다.

CARL Systems과 BPL Foundation은 공동으로 이미징 과정과 링킹 소프트웨어를 CARL Systems 도서관들과 기타 관심있는 기관들에게 매매하였다.

4.2 최근의 사례들

다음의 프로그램들은 최근의 몇 년안에 시작되었다:

(1) 시카고 공공도서관 (Chicago Public Library)

CPL은 1997년 역사적 사진, 고문서 및 지도들의 막대한 컬렉션의 이미지를 촬영하도록 그랜트를 받았다. 목표는 보존보다는 흥미있는 컬렉션을 웹상에 실어주는 것

이었다. 소스문서는 모두 흑백이었고, Agfa DocuScan 플랫폼 스캐너 - 36-비트 컬러와 600x1,200 dpi가 가능한 스캐너 -- 에서 300 dpi로 스캔되었다. 아카이벌 TIFF 카피가 제작되었고, 웹 접근용 JPEG도 만들어졌다.

CPL의 디지털 컬렉션의 흥미있는 측면은 이야기를 구성하도록 주제별로 배열되었는 것이다. 실제로, 컬렉션은 그것이 전체 컬렉션의 단순한 이미지만이 아니라 는 것을 강조하려 디지털 전람회라고 불리워졌다. 전시된 것들 중에는 오대호의 대형 선박들, 시카고의 항만 등이 있었다.

(2) 버지니아 도서관(Library of Virginia)

버지니아 도서관은 주정부 뿐만 아니라 주의 도서관들과 주민들에게 참고 및 연구 지원을 제공하는 주립에이전시로서 미국 내에서 가장 큰 이미지 컬렉션들 중의 하나를 만들어냈다. 2000년 말 현재 웹사이트에 법원기록, 고문서, 지도 및 사진들을 포함하는 이백이십만 이상의 이미지들이 있었다. 여러 공공도서관들의 컬렉션들도 포함되어 있다.

버지니아 도서관은 자체의 이미지 촬영을 할 뿐만 아니라, 여러 서비스 부서들을 이용하였다. 이 도서관은 300내지 400 dpi로 이미지를 찍기를 선호했지만, 지도와 고문서들용으로는 종종 600 dpi로 갔다. 자료들 중 소수만이 컬러로 되어있어, 600 이상의 고해상도로 찍는 것은 필요하다고 보지 않았다. Fujitsu ScanPartner 600 플랫폼 스캐너, Fujitsu 9096 플러스 및 제본도서들을 보호하기

위한 크래들을 갖는 기계인 Zeustschel 5000를 사용하였다. 첫 스캐너는 웹으로 직접 로드하기 위하여 PDF 포맷으로 이미지를 찍었고; 나머지 두 기계는 TIFF로 이미지를 잡아주었다. 비록 이미지 서버가 VTLS에 의하여 공급되지만, VTLS 자동화 도서관 시스템의 MARC 형식 서지 레코드들로부터의 링크는 856태그를 통해서 한다. 따라서, 이미지 서버나 서지 데이터베이스 서버는 서로에게 영향을 주지 않고 교체될 수 있다.

(3) 펜실베이니아 대학(University of Pennsylvania)

펜실베이니아 대학의 Schoenberg Center for Electronic Text and Image는 대학도서관 이미징 프로그램들 중 가장 의욕적인 것의 하나였다. 센터는 대학의 컬렉션으로부터 원문, 문서, 사진, 지도 및 다른 자료들의 팩시밀에 대한 웹 접근을 학계에 제공하도록 설립되었다. 대부분의 이미지 촬영은 세 대의 HP 4C 스캐너, HP ScanJet AEJ 스캐너, Polaroid SprintScan 35사이드 스캐너, Kontron ProgRes 3012 planetary 스캐너 및 Phase One StudioKit 카메라를 포함하는 다양한 장비들을 사용하여 내부에서 이루어졌다. 향후 이년간 세명의 전문 직원과 다섯명의 학생 프로젝트 기능원의 노력의 대부분은 가상의 Shakespeareana의 기획된 주요 archive인 "셰익스피어와 영국의 문예부흥"; 및 Pollard & Redgrave and Wing으로부터 선정된 200건의 본문을 찍기로 UMI와의 계약에 집중할 것이다. 이 스캔은 궁극적으로는

UMI에 의하여 그들의 보다 큰 마이크로 필름 시리즈 중의 하나를 디지털화하는 것에 근거한 제품으로 통합될 것이다. 더 작은 현재 진행중인 프로젝트들은 기업 연차보고서들과 19 세기 및 20세기 초반

의 미국 상인들의 샘플 책자들을 지속적으로 찍는 것을 포함한다. 소스 문서의 넓은 범위를 고려하여, 촬영은 300내지 600dpi 및 종종 더 높은 해상도로 이루어진다.