

# 워터마크와 OCR 을 이용한 전자 캐비닛 시스템

김성민\*, 하은용\*

\*안양대학교 컴퓨터공학과

e-mail : [ksm41777@gmail.com](mailto:ksm41777@gmail.com), [eyha@anyang.ac.kr](mailto:eyha@anyang.ac.kr)

## Electronic Cabinet System using Water Mark and OCR

\*Seong-Min Kim, \*Eun-Yong Ha

\*Dept. of Computer Engineering, An-Yang University

### 요 약

오늘날 컴퓨터 및 통신 기술의 급속한 발달로 인해 디지털 시대가 도래 하면서 전자 서비스가 보급 되었지만 전자 출력물들은 프린터로 출력하여 문서 창고 안에 보관되어 문서의 진위 여부 확인이나 법적인 문제가 발생시에 종이 문서 창고에 의지할 수 밖에 없었다. 이렇게 창고 안에 보관하는 것도 문서의 양이 비약적으로 많아져 운영하는데 막대한 비용이 발생되고 있다. 한편 기존 문서를 전자문서로 변환하면서 데이터 누락되거나 데이터가 변질되거나 유실되는 문제가 빈번하게 발생되었다. 본 논문에서는 기존 문서를 전자 문서로 변환하고, 이후 문서 수정이 발생할 때에는 버전 관리를 통해 안정성을 높인 워터마크와 OCR 기술을 적용한 전자 캐비닛 시스템을 설계 구현한다.

### 1. 서론

오늘날 컴퓨터 및 통신 기술의 급속한 발달로 인해 디지털 시대가 도래 하면서 전자 서비스가 보급 되었지만 전자 출력물들은 프린터로 출력하여 문서 창고 안에 보관되어 문서의 진위 여부 확인이나 법적인 문제가 발생시에 종이 문서 창고에 의지할 수 밖에 없었다. 이렇게 창고 안에 문서를 보관하는 것도 문서의 양이 비약적으로 많아져 운영하는데 막대한 비용이 발생되고 있다.

국내 4 대 보험사가 연간 발생하는 종이 문서는 자그마치 1 억 3600 만 장에 달하며 그 종이의 높이만 무려 13Km 63 빌딩의 54 배 높이에 해당하는 분량이다. 어떤 은행은 1 년에 사용하는 문서의 양은 2 억 4000 만 장으로 4 대 보험사를 합한 것보다 70%나 많다고 한다. 이렇게 각종 계약서 및 서류들을 법적으로 1~5 년씩 또는 영구 보존해야 하는 문서들이 창고에 가득 쌓여만 가고 있다.

이러한 문제로 인해 정부는 2007 년 11 월 종이 문서를 대체하기 위해 전자 문서를 인정해 주는 법을 시행하였으며 기존의 종이 문서를 스캔하여 전자 문서화 한 것도 효력 발생이 가능하게 되었다. 하지만 기존에 발생된 종이 문서를 전자 문서로 변경하면서 데이터가 누락되고 운영 중에 데이터 베이스에 저장된 데이터가 변질되는 문제로 인해 전자 문서가 변질되는 문제가 발생되었다.

본 논문은 이러한 전자 문서의 변질을 막기 위해 기존 종이 문서를 스캔한 후 정보를 추출하여 데이터 베이스에 저장한 후 스캔한 문서는 원본 상태 그대로 전자 캐비닛 시스템에 저장하고, 이후 문서의 수정이 발생할 경우에는 버전 관리를 통해 안정성을 높인 시

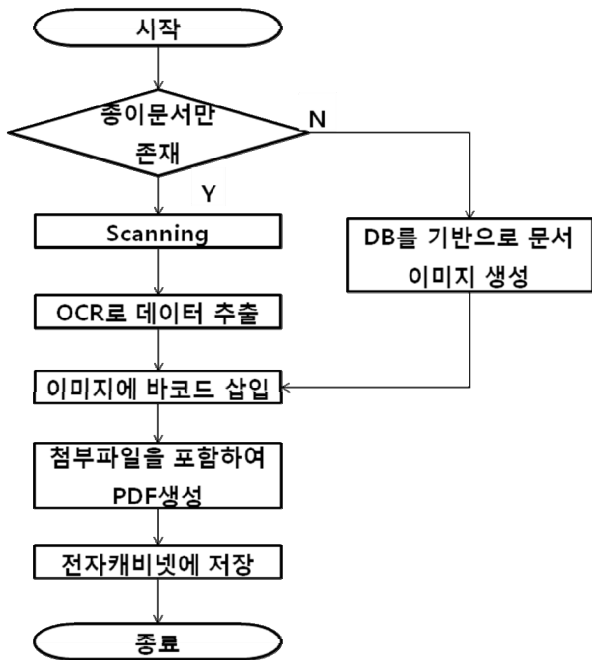
스템을 설계 구현한다.

본 논문의 2 장에서는 기존 문서를 전자문서로 만들어 전자캐비닛에 저장 보관하는 시스템 흐름에 대해 설명하고, 3 장에서는 전자캐비닛의 설계 및 구현 결과에 대해 설명하고, 마지막으로 결론을 맺겠다.

### 2. 전자 캐비닛 시스템 개요

전자문서화 과정을 두 경우로 나누어서 처리한다. 첫째, 전자 서비스가 도입되기 이전에 종이 문서만 발급되어 DB 에 데이터가 존재하지 않는 경우, 둘째 전자 서비스 도입 이후 DB 에 문서에 대한 자료가 존재하는 경우이다. 두 가지의 경우를 모두 고려해서 운영중인 전자 서비스에 영향을 주지 않고 독립적으로 전자 문서만을 관리하는 전자 캐비닛 시스템을 운영하여 어떠한 업무 환경에서도 도입이 가능하도록 설계를 하였다. 추후에 전자 문서의 세밀한 관리 즉 전자 결재, 버전 관리, 문서 폐기 등의 관리는 EDMS(Electronic Document Management System)를 도입하여 확장이 가능하도록 설계하였다.

그림 1 은 전자 캐비닛 시스템의 전자 문서화 과정을 보여준다. 종이 문서만 존재할 경우에는 OCR 로 데이터를 추출하여 이미지에 바코드를 삽입한 후 첨부파일을 포함한 PDF 파일을 생성해서 전자캐비닛에 보관하고 기존 DB 에 정보가 있는 경우에는 DB 를 기반으로 이미지를 생성한 후 동일한 방식으로 처리하여 저장한다.



(그림 1) 전자 문서화 과정

### 3. 시스템 설계 및 구현

본 장에서 전자 캐비닛 시스템을 구성하는 요소들에 대한 설계 및 구현에 대해서 설명한다.

#### 3.1 종이 문서 처리

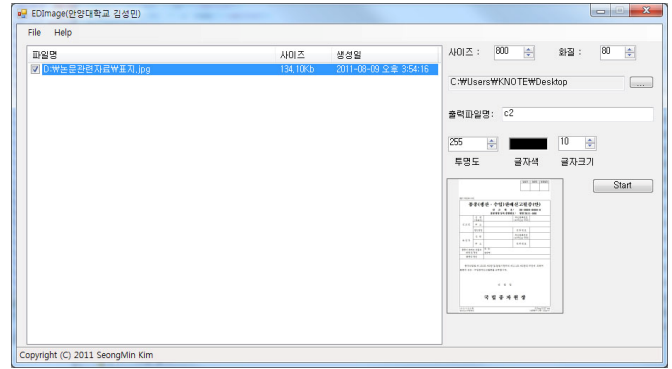
전자 서비스가 보급되기 전의 자료로 종이 문서만 존재하고 DB 에 저장되지 않은 자료는 스캔을 통해 이미지 파일로 생성한 후 그림 3 과 같이 각각의 Rectangle 영역으로 전자 서비스 상의 키워드 검색과 기본 정보 확인을 위해 OCR 을 통해 정보를 추출하여 DB 에 저장한다.

그리고 전자 문서가 오프라인 상으로 발급될 시에 오프라인 상의 검색의 효율성을 위해 바코드를 하단에 삽입한 후 문서 포맷의 표준화를 위해 PDF 파일로 변환한다.

#### 3.2 데이터가 존재하는 문서 처리

전자 서비스가 보급된 이후의 문서로 DB 에서 데이터를 읽어온 후, 그림 2 와 같이 문서 양식 폼과 데이터를 조합하여 하나의 전자 문서 이미지를 생성한다. 이렇게 생성된 전자 문서 이미지의 하단에 바코드를 삽입하고 문서와 관련된 다양한 첨부 파일들(doc, xls, jpg, hwp 등)을 전자 문서와 함께 문서 포맷의 표준화를 위해 PDF 파일로 변환한다.

그림 2 와 같이 새롭게 생성될 전자문서를 미리보기 기능을 통하여 전자문서가 문제 없이 생성이 되는지 생성 전에 확인이 가능하도록 기능을 구현하였으며 문제가 없는 것으로 판단되어 문서 생성 버튼을 클릭하면 하나의 전자문서가 완성하게 된다.



(그림 2) DB 에 존재하는 문서 처리 구현 화면

#### 3.3 OCR 구현

그림 3 의 각 영역은 샘플 문서에서 영역을 추출하는 어플리케이션을 통하여 마우스의 Drag & Drop 의 간단한 조작으로 추출할 정보 Rectangle 좌표와 키보드를 통하여 입력한 ID 값의 조합으로 XML 파일로 생성이 된다.

그림 4 와 같이 스캔을 통해 종이문서에서 이미지 파일로 변환된 전자 문서에 이전 단계에서 정의한 XML 파일을 통하여 OCR 영역으로 지정될 영역 좌표를 적용하여 정보를 추출한 후, XML 에 지정된 ID 값에 해당하는 DB 의 테이블 컬럼에 맞추어 저장하게 된다. 이 정보는 전자 문서의 검색 키워드나 기본 정보로 사용된다.

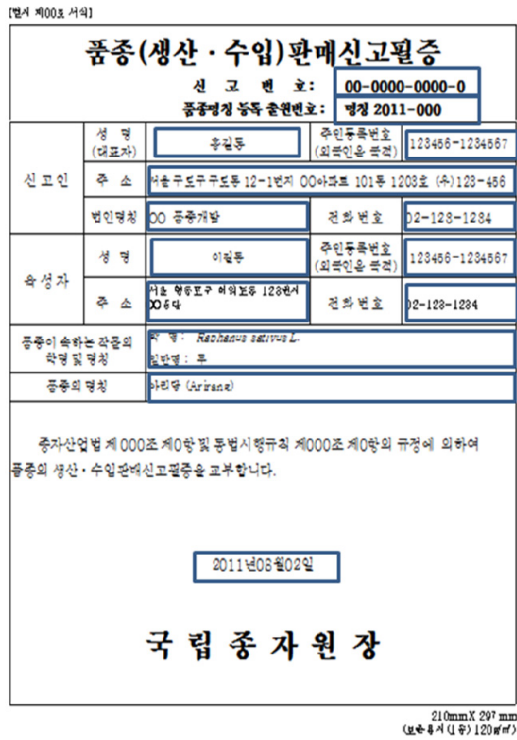
그림 5 은 프로그램으로 구현한 OCR 화면이며 각 영역에 해당하는 정보를 추출하여 오른쪽 화면에 정보가 표시된다.

추출한 정보가 이상이 없을 때에는 DB 저장 기능을 통하여 ID 값에 해당하는 테이블의 컬럼에 각 각 저장되지만 추출한 정보가 맞지 않을 때에는 수동으로 키보드 입력을 통하여 수정한 정보가 DB 에 저장된다.

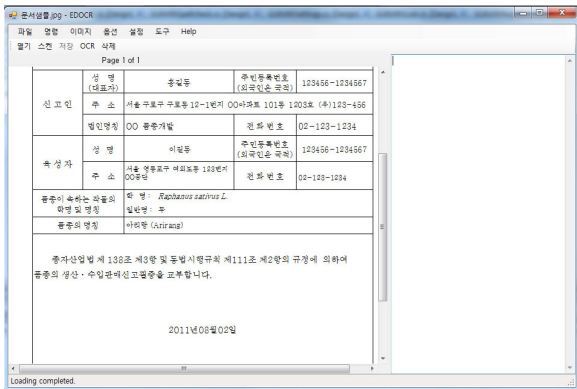
```

<?xml version="1.0" encoding="utf-8"?>
<Grid id="form1">
  <Contents>
    <colinfo x="4.7" y="3.5" width="4.9" height="0.9" id="name"/>
    <colinfo x="12.3" y="3.5" width="3.4" height="0.9" id="rsdtNo"/>
    <colinfo x="4.7" y="4.5" width="11" height="0.9" id="addr"/>
    <colinfo x="4.7" y="6.4" width="4.9" height="0.9" id="name2"/>
    <colinfo x="12.3" y="6.5" width="3.4" height="0.9" id="rsdtNo2"/>
    <colinfo x="4.7" y="7.5" width="4.9" height="0.9" id="addr2"/>
    <colinfo x="12.3" y="7.5" width="3.4" height="0.9" id="tel"/>
  </Contents>
</Grid>
  
```

(그림 3) OCR 에서 사용할 영역 XML



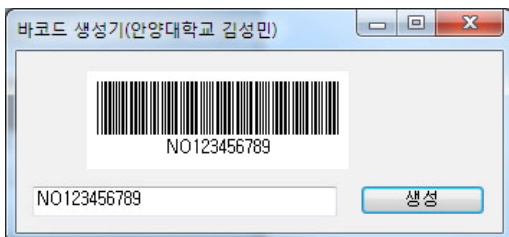
(그림 4) OCR 영역 좌표 화면



(그림 5) OCR 구현 화면

### 3.4 바코드 구현

전자 문서가 오프라인 상으로 발급(출력)될 시에 오프라인 상의 검색의 효율성을 위해 그림 6 와 같이 문서의 고유일련번호를 바코드로 생성하여 전자문서 하단에 삽입한다. 출력된 문서는 전자 문서 관리 시스템에서 바코드 스캐너를 통해 손쉽게 검색이 가능하다.



(그림 6) 바코드

### 4. 결론 및 향후 연구

본 논문은 DB 에 존재하지 않는 종이 문서를 전자화하면서 문서의 누락과 변질을 최소화하기 위해 스캔을 통해 이미지 원본 형태로 유지하여 저장하고 OCR 을 이용하여 정보를 추출하였으며, 바코드를 삽입하여 오프라인상에 출력되었을 때 바코드 스캐너를 통해 손쉬운 검색이 가능하도록 구현하였다. 이렇게 생성된 이미지는 PDF 로 변환한 뒤 전자 캐비닛에 저장하게 되는 과정을 거치게 된다. 그리고 DB 에 존재하여 새롭게 발생할 전자 문서의 경우에는 Water Mark 를 통해 전자 문서 이미지를 생성한 후 마찬가지로 바코드를 삽입하고 다양한 포맷의 첨부 파일을 하나의 PDF 파일로 변환하여 전자 캐비닛 시스템에 저장하는 방식을 제안하였다. 향후 연구로는 데이터가 전자 캐비닛 시스템에 저장되는 것에 확장하여 전자 서명을 통한 인증시스템을 이용하여 보안적인 부분을 더욱 강화하고 EDMS 를 도입하여 문서에 대한 세밀한 관리가 가능하도록 할 것이다. 그리고 전자 문서를 생성, 발급, 수정, 폐기에 따른 권한을 체크하여 문서 보안인 DRM 을 추가하여 문서의 안전성과 무결성의 측면을 강화할 계획이다.

### 참고문헌

- [1] 한국전자문서산업협회, “전자문서산업 정의 및 분류체계”, 2010
- [2] 김형자, “전자문서 종이 없는 세상 만든다”, 사이언스 테크놀러지, 2009
- [3] 정현애, “전자정부 구현을 위한 전자문서관리시스템(EDMS)에 관한 연구”, 2004
- [4] 공준식, “전자문서관리시스템(EDMS) 구축의 핵심 성공요인에 관한 연구”, 2005
- [5] 박대철, “Visual C#.NET 과 함께하는 영상처리 프로그램설계”, 한티미디어, 2011