



문서 인식 프로그램 '아르미 5.0'

합산컴퓨터(대표 김인광)에서 개발한 아르미 5.0은 한글을 비롯해 한자, 일어, 숫자용 문서 인식 프로그램으로 책, 잡지, 신문 등과 같은 인쇄된 자료를 스캔하여 텍스트로 자동 입력시켜 주는 제품이다.

아르미에서는 다단, 그림이 섞여 있는 양식, 세로쓰기 등이 사용된 잡지를 비롯해 소설, 신문의 사설이나 칼럼 등 대부분 문서를 분리하여 인식할 수 있다. 또한 인식 속도도 주간지의 글씨로만 된 경우도 인식 시간은 10초 이내에 처리할 만큼 빠르며 인식률도 제목 같은 큰 글씨를 제외하더라도 기사 부분

의 경우에도 정확하게 인식한다.

아르미는 문서의 자동 해석, 빠른 인식 속도, 정확한 인식이라는 인식 프로그램의 기본 사항을 충족시켜 주는 제품으로 인식이 끝난 후엔 설정된 검색기가 자동으로 철자를 검색한다. 또한 오인식 가능성이 있는 글자의 경우엔 파란색으로 표시되고, 교정 마법사를 실행시키면 후보 단어들을 리스트에서 마우스로 선택해 가며 교정할 수 있는 특징이 있다.

아울러 이런 일련의 작업이 끝나면, 이 결과를 곧바로 아래아한글이나 MS-워드에서 불러 쓸 수 있다. 아래아한글로 출력하려면 출력 설정에서

TXT파일에서 HWP파일로 변경해주고, 외부 편집기에 아래아한글 실행 파일을 입력해야 한다.

이밖에 교정시에 원문 이미지와 대조할 수 있도록 눈 모양의 아이콘을 뜨

고 닫게 하는 기능과 TIFF, PCX, GIF, BMP, 2DC 등 여러 가지 이미지 파일 포맷을 지원한다.

설치 및 구동

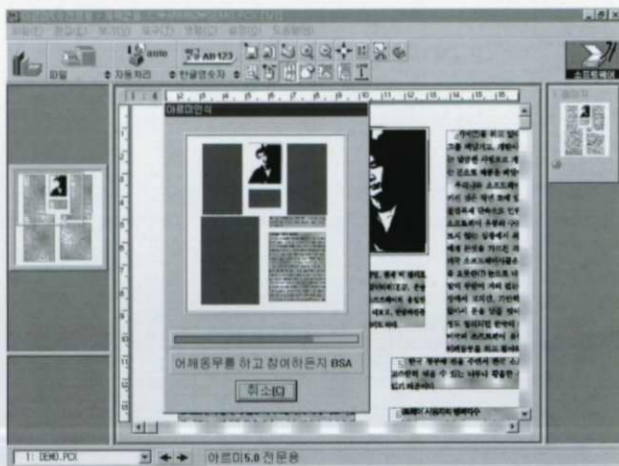
아르미 5.0의 설치를 위해서는 우선 CD를 넣고 인스톨 디렉토리 아래의 setup.exe를 실행시키면 된다. 설치 중에 OLE 관련 DLL의 복사 때문에, 기존 DLL이 설치된 사용자의 경우엔 덮어쓰기를 할 것인지를 물어오는데, 이때는 취소하면 된다. 설치가 끝나면 시스템은 재부팅된다.

다른 윈도우 프로그램과 마찬가지로 아르미도 메뉴와 아이콘의 사용은 동일하다. 그리고 메뉴의 온라인 도움말과 아이콘으로 마우스가 지나갈 경우에 풍선 도움말이 있어 처음 접하는 사용자도 쉽게 사용할 수 있다.

사용자 창(Client Area)은 이미지/문서창, 전체보기창, 확대보기창, 작업표시창 등으로 전체 작업의 진행 정도 및 이미지 인식 때 필요한 정보들이 일목요연하게 나타난다.

사용방법

우선 문서인식을 하려면 원본 인쇄



(그림) 아르미 실행화면

물을 스캐닝해서 이미지 파일로 만들어야 하며 문서인식에서 이미지를 인식하기 좋게 스캐닝 하려면 해상도와 밝기를 잘 맞추어야 한다. 해상도와 밝기는 문서인식을 잘하게 하기 위한 가장 중요한 요소이다. 하지만 각 스캐너마다 이를 조절하는 방법이 다르고 기준이 다르므로 일률적으로 설명할 수는 없지만 일반적으로 해상도의 경우엔 거의 무조건 300dpi로 맞추는 게 좋다.

영문의 경우엔 200dpi로 맞추어 사용하는 경우도 있으나 한글/한자의 경우엔 글자의 획이 복잡하기 때문에 잡지나 신문, 책 등 보통의 경우엔 해상도를 300dpi로 맞춘다. 하지만 전화번호나 사진같이 작은 글씨는 400dpi에서 600dpi 까지 높여 주어야 하고, 어린이 동화책같이 큰 글씨의 경우엔 200dpi로 맞추어 주는 게 좋다.

아르미에서는 스캐닝하는 방법을 여러 가지로 선택할 수 있도록 해 놓았다. 우선 초보자의 경우엔 아르미 상자를 통한 스캐닝이 좋다. 아르미 상자를 열어서 신문류, 잡지류, 갱지류 등 비슷한 유형의 용지 종류만 선택해서 스캐닝하면 어느 정도의 인식률을 얻을 수 있다.

만약 HP 스캐너 사용자는 HP에서 제공하는 기능을 활용할 수 있다. HP Accupage 기능이 그것으로 문서인식을 잘 할 수 있도록 드라이버 수준에서 적당한 밝기로 조절하여 스캐닝해 준다. 초보자의 경우엔 이렇게 스캐닝해서 인식해도 좋다. 하지만 이 기술이 영문에 적합하도록 설계되어 있어

〈표〉 아르미 5.0에 추가된 기능(전문용)

추 가 기 능	내 용
일본어 인식 기능	일본어 히라가나, 가다가나 및 일본어 한자 6,300자를 인식할 수 있고, 편집할 수 있는 기능 추가. 저장은 워드용에는 KS코드로, TXT, RTF에는 S-JIS 코드로 저장함
다중 인식 기능	일괄 인식 작업 중에 편집, 스캔 등의 다른 작업을 동시에 할 수 있는 있음
OLE 기능	OLE Client/Server를 둘 다 지원하여, 아르미를 OLE Client로 설정하여 이미지를 편집해 올 수도 있고, Server로 설정해서 HWP, MS-Word 등에서 OLE 객체로 등록하여 메뉴로 사용할 수 있음
HWP, DOC 파일 포맷	HWP96, 97, DOC 7.0 및 RTF 7.0, 8.0 호환 기능, XLS 파일 등을 추가 지원함
저장 기능	아르미 상자에 의한 스캔 기능 및 미리 보기 기능을 추가하여, 사용자가 쉽게 OCR을 위한 스캐닝을
스캐닝 기능	할 수 있도록 기능을 향상시킴.
한자 인식률 및 인식 속도 향상	한자의 인식률을 높이고 인식 속도도 평균 2배 향상시킴

한글/한자/일어에서는 최적이지 않지만 초보자의 경우엔 아주 유용한 수단 이 된다.

입력하려는 이미지를 스캐닝 받았으면 그 이미지가 이미지 창에 뜨고, 전체 보기창에 레이아웃이 나타나고, 작업 표시창에 작게 나타난다. 여기서 스캐닝 상태를 점검하려면 1:1로 확대하여 글씨 크기가 몽개지거나 끊어짐이 없는지를 한 번 살펴본다. 이는 전체보기 창에서 마우스를 갖다 대면 손 모양의 마우스 커서가 나오면 드래깅하면서 볼 수 있다. 스크롤 바를 이동시키는 것 보다 훨씬 편하고 원하는 위치로 이동시킬 수 있어 좋다.

또한 이미지 상태가 좋지 않다고 판단되면 다시 밝기를 조정해 스캐닝을 받는 게 좋다. 이렇게 이미지 상태를 확인하면서 스캐닝을 잘 하였으면, 인식시키고 교정하고 다른 워드 프로세서나 응용 프로그램에서 활용하는 작업은 훨씬 쉽다고 할 수 있다.

사용환경

- 운영시스템 : 한글 윈도우즈 95/98 또는 한글 윈도우즈 NT-Workstation

- 구동스캐너 : TWAIN을 지원하는 모든 스캐너
- 시스템 : 한글윈도우즈가 실행되는 시스템
- 메인 메모리 : 16MByte 이상 / 권장 32MByte 이상
- 디스크 용량 : 약 70Mbyte / 권장 100Mbyte 이상

제품 가격 및 연락처

- 전문용 : 396,000원(부가세 포함)
- 번들용 : 비매품(각 제품의 번들용으로만 판매)
- 연락처 : 02-786-0291