

OCR SCANNING

Graeme Simpson

OCR Scanning (1)

- Optical Character Recognition - OCR
- Requires OCR program
 - usually comes bundled with scanner
 - separate program such as OmniPage
 - option in scanners software such as HP's PrecisionScan Pro
- Creates an output file of text not a “picture” of the text

OCR Scanning (2)

- Output file is much smaller, e.g. the size of a sample page of text saved as:

File type	Size bytes	File type	Size bytes
Text (TXT)	3,473	Picture (BMP)	995,198
Word (DOC)	22,016	Picture (TIFF)	3,520,182

- Text output, allows
 - editing of text
 - inclusion in word processor file
 - inclusion in Genealogy Program text fields

OCR Scanning (3)

- Output may be formatted
 - e.g. columns, pictures etc.
- OCR Software may automatically start word processor and open text
- May also be able to have text read back through PCs speakers

OCR Scanning (4)

- OCR Scanning process
 - scan page
 - select area wanted,
 - maybe a number of text areas “framed”
 - may also include pictures
 - recognize/process text
 - correct any unrecognizable text/characters
 - save file

OCR Scanning (5)

- May also process files containing text
 - already have a “picture” file of text
 - i.e. *.TIFF file (size 3,520,182 bytes)
 - some programs also process PDF files
 - i.e. *.PDF file (size 12,455 bytes)
- Don't use for capturing pictures
 - better quality/control from picture scanning