

Appendix

PRODUCT DESCRIPTION

This appendix gives a detailed product description of Readiris.

Recognition

Connectionist technology
omnifont feature extraction
self-learning routines derived from a proprietary neural network technique
contextual and linguistic analysis
powerful learning module

Linguistic support

52 languages (American English, British English, Afrikaans, Albanian, Basque, Brazilian, Bulgarian, Byelorussian, Catalan, Corse, Croatian, Czech, Danish, Dutch, Estonian, Finnish, French, Frisian, German, Greek, Hungarian, Icelandic, Irish (Gaelic), Italian, Lithuanian, Luxemburgh, Macedonian, Malay, Norwegian, Nynorsk, Polish, Portuguese, Romanian, Russian, Scottish (Gaelic), Serbian, Slovak, Slovenian, Spanish, Swahili, Swedish, Tagalog, Turkish and Ukrainian)
6 mixed language modes (Bulgarian-English, Byelorussian-English, Greek-English, Macedonian-English, Russian-English and Ukrainian-English)
linguistic modules using knowledge about syllables and a 100,000 word lexicon per language

**Character sets**

all American and European character sets, including the Central-European, Cyrillic (“Russian”), Greek and Baltic alphabets

use of mixed characters sets for recognition of “Western” words (proper names, brand names etc.) in Cyrillic (“Russian”) and Greek documents

numeric mode for recognition of tables of figures

trainable on any special symbol: mathematic and scientific symbols, dingbats etc.

Printing quality

typeset, typed, laser printed, inkjet printed, letter quality dot matrix (“25 pin” or “NLQ”) and draft dot matrix (“9 pin”) printed documents

lineskew tolerance 7 degrees

Fonts

virtually any proportional and fixed (“monospaced”) typeface

normal, bold, italic and underlined tpestyles

detection and recreation of font type (serif-sans serif, proportional-fixed), tpestyle (bold, italic, underline) and point size of original document in “autoformatting” mode

character sizes 6 to 72 point (0.08 to 1" or 0.21 to 2.54 cm)

drop letter (“drop caps”) recognition

User interface

toolbar based operation

automatic mode allows “one-click” OCR

pop-up style user verification

stand-alone operation or call from within Windows applications through “Connect” capability
DDE command set available to integrate OCR function

Preview

image preview with 5 zoom levels and 3 rotate commands
highly accurate, skew-tolerant automatic page analysis discriminates text zones, tables and graphics
detection of white text on a black background with automatic inversion at recognition time and image inversion of all scanned pages
adjustable manual windowing of relevant zones using rectangular and polygonal windows
storage of zoning templates for future use

Text output

Windows clipboard and all leading wordprocessors and text formats
paragraph detection ensures wordwrap
“autoformatting” feature recreates facsimile copy of original document through *word formatting* - font type (serif-sans serif, proportional-fixed), typestyle (bold, italic, underline) and point size -, *paragraph formatting* - alignment (left, justified) and indentation - and *page formatting* - placement of text blocks, columns and graphics

Table output

Windows clipboard and generic table format, including Excel and Lotus 1-2-3



detection and recreation of tables in spreadsheets and wordprocessors

Graphic output

black-and-white graphics in (compressed) TIFF and Paintbrush (PCX) format

graphics included in RTF file when “autoformatting” is applied

Scanner support

supports wide range of flatbed, sheetfed and hand-held scanners and “all-in-one” MFPs (“multifunctional peripherals”)

compliant with Twain scanner standard

supports scanning platforms Visioneer PaperPort, HP Document Assistant, HP JetSuite, Microtek PageSuite and Primax PaperEase

Graphic input

black-and-white images in TIFF (uncompressed, packbits, Group 3 and Group 4 compressed), Paintbrush (PCX) and Windows bitmap (BMP) format

allows direct recognition of fax files

“drag and drop” of images onto the Readiris image zone

System requirements

a 486 based Intel PC with 16 MB RAM and 30 MB free disk space. A Pentium based PC with 32 MB RAM is recommended under Windows NT 4.0 and to use the “Connect” capability

the Windows 98, 95 or Windows NT 4.0 operating system

MacOS version available

Linguistic versions

English, Czech, Dutch, French, German, Greek, Italian, Polish, Russian and Spanish