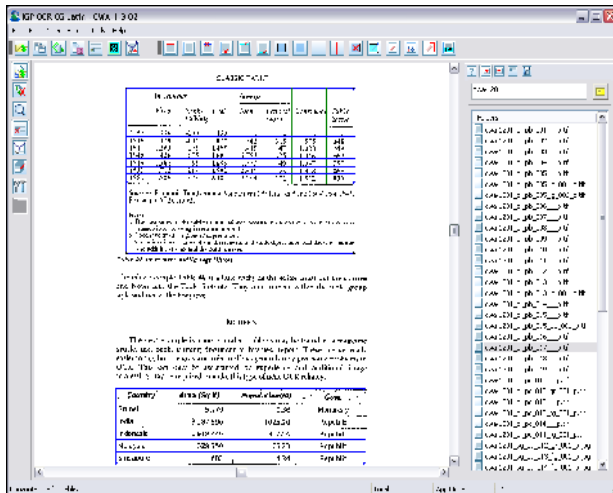


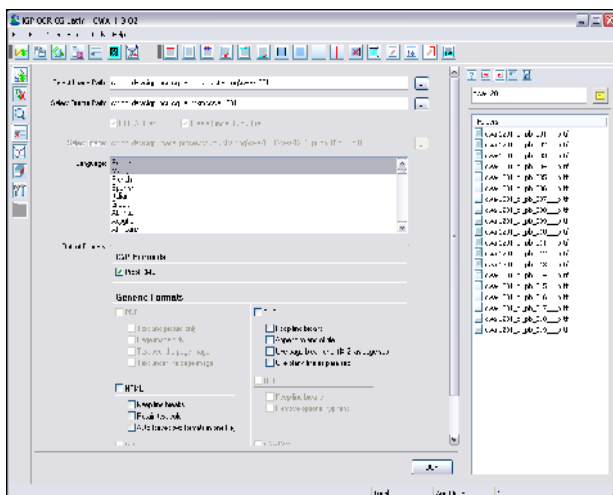
### OCR+. The next level of text processing

#### Main Features

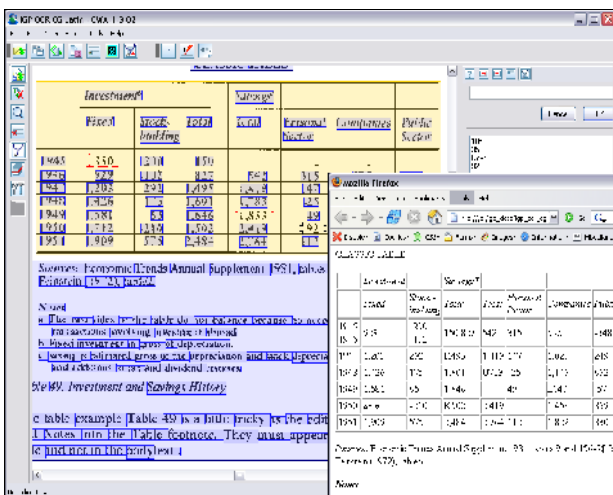
- ◆ OCR Latin, Greek and Cyrillic text
- ◆ All the tools to get the best performance from standard OCR technology
- ◆ FixIt Mode allows preparation of image before committing to OCR for optimized page layout
- ◆ Document Doctor Mode allows evaluation and repair of OCR output results, especially missing text
- ◆ OCR Mode sets languages for diacritic processing
- ◆ Instant HTML view of OCR result for quality confidence and defect analysis



FixIt mode allows the editor to apply guidelines to improve OCR layout perception, and to remove unrequired text



In OCR Mode language selection can be made before committing to OCR to increase accuracy of special character perception.



Document Doctor Mode allows you to see and instantly key missing content in position directly on the image. HTML View allows you to evaluate the OCR layout perception quality.

#### Overview

IGP:OCR:CGLatin takes the industry standard Fine Reader\* OCR output to new levels of accuracy. It is designed for critical OCR'ing of complex layout and text content, as well as simple "OCR-Now" operations. It handles all languages that use the Cyrillic, Greek and Latin alphabets.

OCR:CGLatin can be used as a stand-alone application or as a workflow node in an Information Processing Grid.

#### Moving Past OCR defects

OCR defects are of many types and can occur for many reasons, from disturbances on the page image, to over-kerned text. Whatever the reason, you want to spend as little time as possible finding and fixing the defects.

IGP:OCR:CGLatin lets the user prepare the content for proofing operations, analysing string constructions as each page is OCR'ed to speed later processing. It also optimizes layout and applies basic auto-tags to speed up later processing. The graphical approach to content deconstruction makes it easy to use, even by "non-editor" novices.

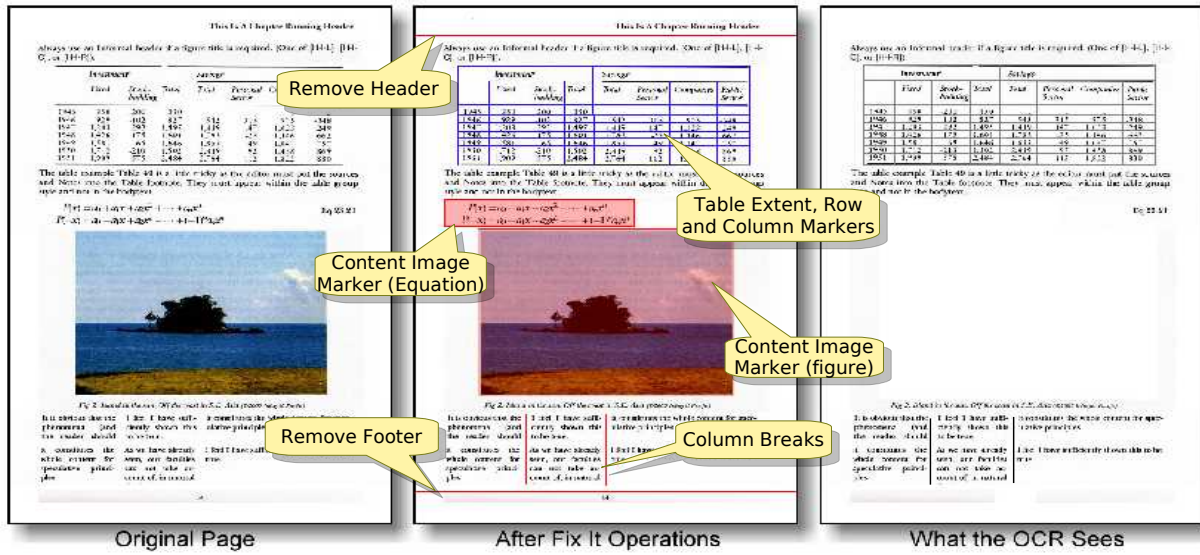
#### More than Just OCR

The Fix It mode allows the user to mark-up documents with complex layout to compensate for OCR limitations. Including re-interpret the block layout, examine and correct auto-paragraphing, and reconstruct and correct table layouts and formatting.

The Document Doctor mode allows the user to find and replace any text the OCR may have missed for any reason, including directly inserting missing text on the image.

IGP:OCR:CGLatin not only speeds up proofing, but also drastically cuts XML Encoding time in subsequent workflow steps.

You can also use IGP:OCR:CGLatin in an ad-hoc workflow mode. If you need to just quickly OCR half a page for metadata extraction, or just extract a few blocks, it is as easy as the full workflow model. It is designed to be used the way you need it.



### User friendly

OCR: CGLatin makes the intimidating job of OCR'ing and deconstructing complex academic, mixed language and scientific texts easy for any level of operator with the minimum training investment.

\*Fine Reader is a registered Trade Mark of ABBYY Software. Standard product pricing includes Fine Reader licensing for 10,000 pages per month per seat.

### Application Features

#### Fix It Features

- Create a special temporary OCR page image edited to ensure only the items that need to be OCR'ed on a page don't have to be edited out with later text processes
- Mark and clear any part of the document for no-OCR output such as images
- Clear headers and footers page by page or globally (if image processing centering is carried out accurately)
- Insert column guide rules
- Interactively create or repair page block structures such as tables, images and column structures
- Create content images with Image cut or copy, and assign a content type value to them

#### OCR Features

- OCR all the languages Fine Reader can handle
- Output single file per image, or a continuous file from range or directory of images.
- Create text, HTML PLUS Infogrid Proof Ready XML.
- Output XML as UTF-8 and UTF-16
- Standard seat licensed for 10,000 pages per month (A4 area)
- Watched folder operation allows automatic OCR of any image file or directory of files that appears in the folder. Works in Local and Profiles modes.

#### Document Doctor Features

- View overlays of OCR block structures on the image, right down to the character and word level.
- Directly select and edit word strings in OCR blocks

- Create new blocks over any content the OCR missed and key in the correct text
- Use HTML view to instantly see if the FixIt and OCR applications have correctly interpreted page layout into the correct sequence.

#### IGP:OCR CGLatin Toolkit

- IGP:OCR CGLatin Toolkit application is supplied free with every IGP:OCR CGLatin application
- Create project configuration files, profiles and local setup files.
- Directly create workflow configuration files.
- Click-thru configuration makes creating new projects fast and easy

#### Enterprise Features

- Can be used in Local (stand-alone), Profiles or Grid modes
- Connects to IGP:Distribkutor server through IGD:Grid Connector when installed and allows separation of content on a project basis.
- See the IGP:Distributor Fact Sheet for More information.

### Software and Hardware Requirements

#### Operating System Environment

- Operating Systems Linux (Kernel 2.6) or Windows (XP Home or Professional) • Recommended window environment for Linux – KDE 2.0. or a user friendly installation of Linux such as Xandros or Ubuntu • Full Linux Installation CD-ROM provided on request

#### Recommended Hardware

- 32/64 bit Processor with better than 2GHz processor speed
  - 512Mbyte RAM minimum. 1GByte RAM highly recommended
  - 30Mbyte common application software library space on hard disk
  - 500Kbyte application space on hard disk

**Infogrid Pacific Pte Ltd**  
 141 Cecil St #06-05  
 Tung Ann Assn. Bldg.  
 Singapore 069541

Tel: +65 62243798  
 Fax: +65 62243797

email: sales@infogridpacific.com  
 www.infogridpacific.com

