



iDRS

IRISpdf 5

عربي



PRESS RELEASE

Brussels, March 27th 2006

I.R.I.S. introduces new Arabic and Farsi OCR technology and releases Middle-East versions for three I.R.I.S. products: Readiris Pro Corporate Edition, IRISpdf and iDRS.

Brussels, March 27th 2006 – I.R.I.S. Group, a publicly traded company (Euronext: IRI), market leader in Intelligent Document Recognition (IDR), Electronic Document Management (EDM) and Optical Character Recognition (OCR) announced the immediate release of a new Arabic and Farsi OCR engine that will be incorporated in a new Middle-East version of several products. This Middle-East version also includes Hebrew recognition.

Arabic is the official language of twenty two countries representing more than 280 million inhabitants. It is ranked amongst the top ten languages in the world in terms of number of speakers.

Farsi is a different language spoken in Iran, and based on a variation of the Arabic script.

I.R.I.S. will release Middle-East version for three of its best-selling products

- **Readiris Pro 11 Corporate Middle-East:**

Readiris Pro 11 is an “all-in-one” document scanning and OCR solution for Small and Medium businesses. Amongst its variety of features are:

- Powerful OCR and image compression capabilities.
- Generation of very compact fully-searchable PDF files.
- Support of professional color-duplex scanners (Fujitsu, Kodak, Bell&Howell, etc...)
- Document separation and document indexing.

- Powerful table recognition technology for the encoding of price lists, contact lists, etc..
- Powerful background detection technology that allows to rescan and republish product brochures or documents containing text printed on an artistic background.
- Circulation of documents through e-mail
- Simultaneous scan and recognition of up to 10 business cards simultaneously placed on a flatbed scanner and export of the contacts to outlook or other contact management databases.

- **IRISPdf 5.0 Server Middle-East:**

IRISPdf 5.0 is an advanced production OCR solution that allows to build powerful OCR and image compression servers that will convert volumes of scanned documents or images into editable or searchable electronic documents. This product is the ideal front-end for any document management system that supports full-text search and that needs to import compact fully-searchable files containing the images and the full-text index.

- **iDRS Middle-East:**

The iDRS imaging and OCR toolkit of I.R.I.S. offers different capabilities in the image processing and the automatic document reading areas. Its simple and complete C++ interface will help you to integrate easily one or more of the latest I.R.I.S. technologies. With its ActiveX component, you can quickly and easily make a complete application including I.R.I.S.' recognition engines.

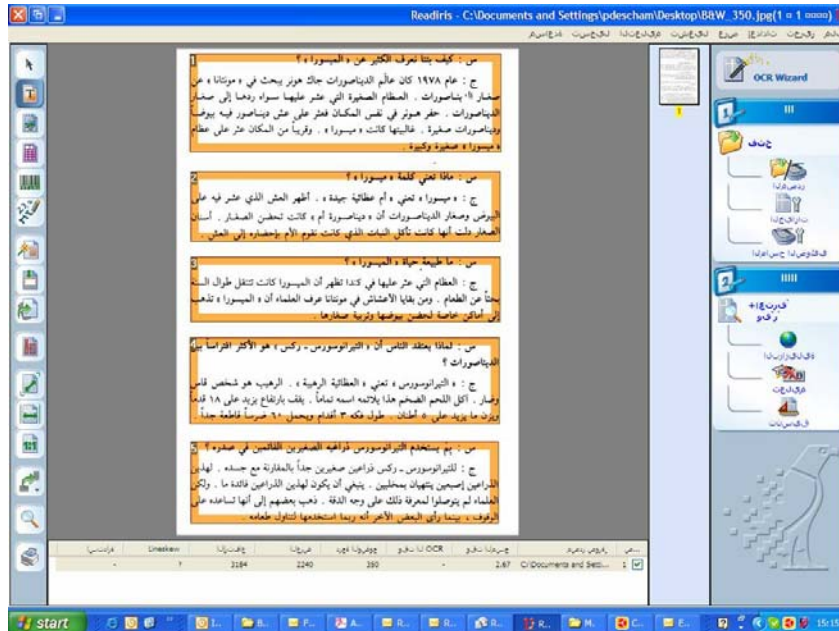
Pierre De Muelenaere, President & CEO of I.R.I.S. Group, comments this new launch: *"The addition of Arabic and Farsi OCR into three of our best selling products will allow us to develop our business in the Middle-East market and to forge new partnerships in the coming years. We are offering a comprehensive product line that can fit the needs of Small and Medium Businesses, Very Large Accounts and Systems Integrators. With these three products, we are making available a very broad set of functionalities, which will allow our partners to solve a wide variety of problems."*

Recognition of Hebrew was available earlier from I.R.I.S., but only as a single add-on. It has now been included in this specific Middle-East version.

Major Features of Readiris Pro 11 Corporate Middle-East

Intuitive User Interface in either Arabic, Farsi or Hebrew

Readiris Pro's logical interface is the easiest to use of any OCR software on the market. The status window gives you important information about your scanned documents (source, resolution, times, image processing, etc).



Input Formats
 Readiris reads most popular image formats as TIF, JPEG, BMP, etc. It can also open and recognize the highly compressed JPEG 2000 and DjVu files.

Efficient Batch OCR

Batch OCR executes the recognition on all pre-scanned images in a specific folder.

Productive Watched folder

Readiris Pro 11 Corporate Edition can set up a “watched folder”. Readiris systematically executes the recognition on any image file dropped in this folder. You can leave the OCR software running day after day... Acquire new documents and they will be recognized promptly.

Powerful Bar Code Recognition for indexing and separation

Separate your documents by blank pages followed by a page containing bar codes. More than 20 bar codes are automatically recognized (even the PDF 417 barcode) and the embedded information is saved in an XML index.

Impressive Foreign Language Support

Recognize up to 124 foreign languages including Arabic !

Exhaustive List of Output Formats

We let you reproduce your documents in an extensive number of applications such as Word, Acrobat, Internet Explorer, Netscape, StarOffice, and many others.

Advanced PDF Generation

Exceptional tools allow you to create PDFs with optimized file size:

- Generate the PDF type that best suits your needs: image over text, text over image, text or image only
- Optimize the PDF size for an easier archiving and sharing: reduce the image resolution or set the JPEG file quality. You can

even save your PDF using JPEG 2000 compressed format to get the smallest file size. Perfect for sharing and archiving!

- Directly create a PDF without recognizing the scanned document with the "Save full page as image" function

Digital Signature of PDF files

This new output format gives you the security that your document, once created, won't change in the future. You can then store it or send it to a workflow knowing it will remain unchanged during the whole process.

Splendid Color Output Files

Colored text, backgrounds, graphics are fully reproduced with Readiris Pro. Recreate the look and feel of your originals with great touch and accuracy.

System Requirements

Readiris Pro 11 Corporate Edition

- A 486 based Intel PC or compatible. A Pentium based PC is recommended.
- 64 MB RAM. 128 MB RAM is recommended to process greyscale and color images.
- 120 MB free disk space. 105 MB of disk space suffices when you leave the sample files on the CD-ROM.
- The Windows XP, Windows ME, Windows 2000, Windows 98 or Windows NT 4.0 operating system.
- A monitor with a 1024 x 768 resolution.

Price and Availability

Readiris Pro 11 Corporate Edition – Middle East will be available in the distribution, in retail channels and through Value-Added Resellers (VARs) as from March 25th at the price of 1.239€ (Suggested Retail Price).

IRISPdf 5.0 Server Middle-East will be available through Value-Added Resellers (VARs) as from March 25th at the price of 3.980€.

iDRS including the Middle-East module will be available directly from I.R.I.S. The price of the Software Development Kit is depending on the number of modules needed (+ Runtime Licenses).

List of features (Readiris Pro 11 Corporate Edition, IRISPdf 5.0 Server, iDRS)

	Readiris Pro 11 Corporate Edition	IRISPdf 5.0 Server	iDRS
	"All-in-one" document scanning and OCR solution	Powerful production OCR for the conversion of high volumes of documents	Modular toolkit to easily integrate I.R.I.S.' technology into your own application
	For: SOHO and SMB's	For: SMB's, large companies, institutions and organisations	For: VARs, integrators, developers and scanner manufacturers
Recognition of ligatures	X	X	X
English-Arabic Recognition in the same text zone	X	X	X

Omnifont Arabic recognition	X	X	X
Recognition of italic characters	X	X	X
Arabic user interface	X		
Maintains Color of Text in the Output File	X	X	X
Maintains Background Color in the Output File	X	X	X
Opens and Recognizes JPEG 2000 Files	X	X	X
Opens and Recognizes DjVu Files	X	X	X
Best speed & accuracy combination	X	X	X
Barcode Recognition	X	X	X
Efficient "Batch OCR"	X	X	X
Productive "Watched Folder"	X	X	X
Many output formats (RTF, HTML, WordML, etc.)	X	X	X
Four types of PDF output	X	X	X
Easy Batch Separation in Documents	X	X	X
Powerful Document Indexing Based on Barcode Reading	X	X	X
XML Index Generation	X	X	X
Direct Saving of Scanned Documents in JPEG 2000	X	X	X
One-click Transfer to Your Clipboard	X		
Maintenance Program with Free Releases and High Quality Support	X	X	X
Support of Duplex Scanners	X		
HTML Output	X	X	X
WordML Output	X	X	X
Up to 124 Languages Recognized	X	X	X
RTF Output for OpenOffice and StarOffice	X	X	X
Sends the Recognized Document Directly to Your e-mail Application	X		

For detailed list of features of Readiris Pro 11 Corporate Edition and IRISpdf 5.0, please refer to the respective press releases.

About Arabic Recognition

Introduction

The Arabic handwriting is an evolution of the Aramaic handwriting, which is used since the 4th century. The Aramaic handwriting has less consonants than the Arabic, and several letters have therefore been created during the 7th century by adding points (called diacritic signs) to the existing letters, in order to avoid ambiguities. Other signs, like diacritic points indicating short vowels, were also introduced but are merely used for the writing of the Koran.

The Arabic alphabet has 28 letters. Additional letters are used for the handwriting of foreign words, containing sounds that aren't familiar in Arabic. The Arabic alphabet also comprises 10 digits, which form is different from the digits used in Europe and, nevertheless, qualified as Arabic numerals.

Arabic is a Semitic language, spoken by more than 280 million people. Other languages are using the Arabic alphabet like, for example, Farsi (Iran), Urdu (Pakistan) or Pashto (Afghanistan).

Example of a text written in Arabic:

يولد جميع الناس أحراراً متساوين في الكرامة والحقوق. وقد وهبوا
عقلاً وضميراً وعليهم أن يعامل بعضهم بعضاً بروح الإخاء.

Translation:

All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood.

(Universal Declaration of Human Rights. Article 1)

Difficulties

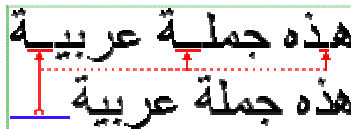
The challenges related to the recognition of texts printed in Arabic are of different nature:

- Arabic texts are written right-to-left. Nevertheless, the numbers are written left-to-right and so are the proper nouns (English, French), often written with the Latin alphabet.
- The text is cursive: letters within a word are joined to one another with a baseline.
- Each letter of the Arabic alphabet has four forms: initial, medial, final or isolated. Nevertheless, for several letters, the medial and final forms are missing. In this case, the letter takes the initial (or isolated) form and the previous letter in the word takes the final (or isolated) form.

- Moreover, Arabic text contains a great amount of special signs, called ligatures, which replace two, or even three consecutive letters. For example, when the letter LAM is followed by the letter ALEF, these two letters are almost always replaced by the LAM-ALEF ligature. While the LAM-ALEF ligature is almost universal, many ligatures are only optional. The choice, whether he wants to use them or not, relies completely on the author.
- Arabic texts are often justified, in order to align the right and left borders of the text columns. In « Roman » printing (French, English...), you can realize this by stretching out the spaces between the words or the characters, in order to get the wanted length. In Arabic printing, however, it's the baselines joining the characters that are being stretched out. Those lengthened baselines are called « kashidas ».

Example:

The following figure shows 2 identical paragraphs of Arabic text. The blue line in the second line (not justified) shows the length used for _____ and spread amongst the elongations in the first line (justified) as shown by the red underlining.



- The speckles due to the scanning or the bad quality of the document can be misinterpreted as diacritic signs.

Arabic Alphabet:

IPA	Value	Name	Final	Medial	Initial	Isolated	IPA	Value	Name	Final	Medial	Initial	Isolated
[ʔ]	ʔ	qāf	ض	ض	ض	ض	[ʔ]	'(a)	alif	ا	-	-	ا
[t]	t	tā'	ط	ط	ط	ط	[b]	b	bā'	ب	ب	ب	ب
[z]	z	zā'	ظ	ظ	ظ	ظ	[t]	t	tā'	ت	ت	ت	ت
[ʕ]	ʕ	'ayn	ع	ع	ع	ع	[θ]	th	thā'	ث	ث	ث	ث
[ɣ]	gh	ghayn	غ	غ	غ	غ	[ʒ]	j	jīm	ج	ج	ج	ج
[f]	f	fā'	ف	ف	ف	ف	[ħ]	h	hā'	ح	ح	ح	ح
[q]	q	qāf	ق	ق	ق	ق	[x]	kh	khā'	خ	خ	خ	خ
[k]	k	kāf	ك	ك	ك	ك	[d]	d	dāl	د	-	-	د
[l]	l	lām	ل	ل	ل	ل	[ð]	dh	dhāl	ذ	-	-	ذ
[m]	m	mīm	م	م	م	م	[r]	r	rā'	ر	-	-	ر
[n]	n	nūn	ن	ن	ن	ن	[z]	z	zāy	ز	-	-	ز
[h]	h	hā'	ه	ه	ه	ه	[s]	s	sīn	س	س	س	س
[w]	w	wāw	و	-	-	و	[ʃ]	š	shīn	ش	ش	ش	ش
[j]	y	yā'	ي	ي	ي	ي	[s]	ṣ	ṣād	ص	ص	ص	ص

About I.R.I.S.

Image Recognition Integrated Systems (I.R.I.S.), a Belgian company founded in 1987 and listed on Euronext Brussels (IRI), is a leader in the "Document to Knowledge" market, and provides extremely high-quality solutions for converting paper documents into electronic formats for archiving, storing, managing and sharing digital information.

In 2004, the revenue of I.R.I.S. Group was 46.695.918 €. I.R.I.S. employs almost 250 employees based in Louvain-la-Neuve and Brussels (Belgium), Orly (France), Luxembourg and Delray Beach (USA).



I.R.I.S. has been awarded "Enterprise of the Year 2002" and has received the "2003 ICT Award" and the "2003 BeLAIIM" Award" for its projects.



For more information on our company, our solutions or our products, visit the I.R.I.S. website at <http://www.irislink.com>.

To get more information

I.R.I.S. Europe Headquarters
 Press contact: Sarah Dheedene
 Tel: +32 (0) 10 48 75 13
 Fax: +32 (0) 10 45 34 43
 E-mail: sarah.dheedene@irislink.com