

# RECOGNITA NEWS

Vol.1.No.1.

March 1989



On behalf of the 700 member team of our Institute let me cordially welcome the readers of

RECOGNITA NEWS published for the first time now.

You certainly know that this publication was born on our partners' suggestion and is destined to provide information to our end-users and commercial partners on the professional results concerning Recognita and other related matters.

I think that the market response to the launching of new enhancement of our OCR program Recognita well represents that the use of computer support in the offices and in printing activities is gaining momentum and spreading rapidly to wide extent.

On this basis we can suppose that both the theoretical basis of Recognita and the actual development carried out with regard to our end-users' experiences proved to be right.

We are bound by our results, so the development of new products is carried out ceaselessly.

Important results have been recently achieved in the development of programs to be applied in the field of office-automation, archiving and image-processing. I am sure that these products will give you as much satisfaction and advantage as Recognita did.

Let me take the opportunity to express my sincere hope, that we can further meet our customers' demands.

Dr. Zsolt Náray  
Director-General of the SZKI

## New RECOGNITA Version on Market

From the beginning of 1989, SZKI is shipping a new RECOGNITA version, Version 1.30. The major improvements of the new version are as follows:

RECOGNITA Version 1.30 is running on IBM PS/2, and 80386 computers under MS-DOS, too. VGA or compatible graphics card in CGA mode and corresponding monitor are supported. RECOGNITA's line of supported scanners now include AVR, COMPUSCAN, DEST PSSCAN 2000, SIEMENS HighScan, IBM PageScanner, and TELIC AGD-300 scanners, too. (See page 4 for the supported scanners.) The Version 1.30 of RECOGNITA lets you read multiple pages if an automated document feeder is connected to your scanner. It supports CompuScan, Hewlett Packard, MICROTEK and RICOH ADF devices.

There is a menu driven setup program on the RECOGNITA diskette, so you need not to edit the RECOGN.SYS configuration file, any more. You need to answer a few questions to get RECOGNITA ready to run on your computer configuration. Having answered the questions your RECOGNITA is installed, you can use it without any delay.

A set of different parameters e.g. text formatting options, rejection symbol, missing symbol etc. can be set from the main menu (item Options) instead of editing the RECOGN.SYS configuration file, as in the previous version.

"Multiple pages" is a new item in the main menu. According to the default setting, it is active after starting the RECOGNITA. This is a useful possibility if you want to read large quantities of docu-

ments. If you have an automatic document feeder, RECOGNITA reads one page after the other without user intervention. If you have no automatic document feeder you must feed every page in the scanner, when RECOGNITA's message appears in the upper right corner of the screen.

The recognized text can be stored in text files, every page in separate files or all pages in one file. If you want to read one page only, you must reset "Multiple pages" or click on "Stop" on the screen.

Improvements in the recognition method give you the ability to read and recognize documents with mixed Latin and Greek letters. It is expressly designed to help the creation of bilingual newsletters, dictionaries or any printed papers using a combination of Latin and Greek characters

If you teach RECOGNITA in learning mode, the pretrained RECOGN.TRE classification tree does not change. The new information is built in into the USER.TRE file or into another file, if you have changed the name of the user tree. So you have the possibility to erase the unwanted modifications made in the classification tree in the training mode.

RECOGNITA Version 1.30 users estimate it very effective, it is useful to upgrade the previous version of RECOGNITA to the 1.30 one. SZKI promotes this action with supplying the Version 1.30 of RECOGNITA to registered customers for an update price. If you want to get a new RECOGNITA version on update price, you must be registered RECOGNITA user. Having mailed the USER'S REGISTRATION CARD to your dealer, you become a registered RECOGNITA user.

## RECOGNITA V1.30 Tutorial Package

The Tutorial Package contains two programs: TUTOR and DEMO. TUTOR is a self-running tutorial program offering an intensive introduction to the basics of OCR in general as well as to the RECOGNITA Optical Character Recognition program.

TUTOR is a useful tool for managers before they decide which OCR program should be obtained, and for RECOGNITA users if they want to get detailed knowledge of the application possibility of their OCR program.

You do not need a scanner to run TUTOR, all you need is an IBM PC with 640 Kbytes of RAM and a hard disk.

TUTOR is installed by the installation program like RECOGNITA, but you are not asked to define the scanner type used.

RECOGNITA Tutorial Package is delivered on 1 high density diskette or 3 low density diskettes.

The Tutorial Package contains -- besides the TUTOR and DEMO programs a short User's Guide to RECOGNITA. We suggest you should print the file named MANUAL.DOC via the

### PRINT MANUAL.DOC

command, and to read the printed matter before starting the TUTOR program.

"TUTOR is SZKI's contribution for the

present and future RECOGNITA users. TUTOR's first version is related to RECOGNITA Version 1.30, but this is not the last version of TUTOR.

Parallel to RECOGNITA, TUTOR is developed too, to give our users the possibility to learn the new facilities of RECOGNITA" — said Mrs. Anna Nikodémusz, member of SZKI's RECOGNITA team.

SZKI distributes the Tutorial Package free of charge, and allows the duplication of the package provided that you indicate the copyright notice on the diskette label:

COPYRIGHT (c) SZKI, 1989.

## My Point of View

Three questions to one Hungarian user and an international distributor.

1. What do you currently use Recognita for?
2. What is your opinion of the Recognita?
3. What do you need?

### András Virág (Olajterv, Budapest)

1. Our company uses Recognita to read the everyday incoming letters to be processed and distributed in our PC-based local area network. Old and repeatedly usable project materials are also read and updated from time to time.

2. The Recognita gives us adequate comfort of use and service. Reading of printouts made by laser printers is practically error-free. In the overwhelming majority of cases we use old typewriters. We acknowledge that the usability of the program will be improved by replacing our typewriters and with the use of good-quality carbon tapes. We have had occasion to see that even in the case of relatively high-quality daisy wheel printers, text density is not homogeneous enough to give acceptable reliability, unless new ribbons are used.

3. We need a Recognita in which characters printed by a matrix printer can be recognized.

### Werner Okrina (Artaker, Vienna)

1. In our experience Recognita is mostly used by printing offices, publishing houses, universities, documentation and translation offices and manufacturing companies (for example Philips Data).

2. We are finding the increasing satisfaction of the customers. In 1988 RECOGNITA gained a nearly 15 % market share of the European OCR market. RECOGNITA has a big success in the German speaking country, where it's market share is over 30 %.

3. There is a need for a Recognita built under Microsoft WINDOWS (2.x., -286 and -386).

## What is Coming Next?

SZKI is continuously developing RECOGNITA to offer enhanced facilities to prospective buyers and registered RECOGNITA users.

"The most important point of the function of OCR software is the recognition process.

Now RECOGNITA read some bad quality documents, but if we want to enlarge its application area, we must assure that it should be able to read documents of very bad quality, too. RECOGNITA's recognition process must be improved, though it assures a very low error rate in the recognition process for the moment, too.

We plan to improve the reading quality

of documents containing frames, underlining, vertical line separation. These improvements eliminate the need of cutting out frames or vertical lines with windows. Improved recognition of multi-column texts will also be provided.

There should not be limitations concerning the number of windows, either.

The program selectivity should be increased, namely the character set may be limited during the recognition process, e.g. it should be able to read numbers or lower case or upper case letters only. Another task is the improvement of reading letters with accent.

Our target is to provide automatic sep-

aration of graphics and texts, so the user need not struggle with creating windows. Further, we are making efforts to integrate RECOGNITA under MS-WINDOWS, and to make the most of it for RECOGNITA" — said Emőke Kovács, leader of SZKI's RECOGNITA team.

Certainly, RECOGNITA's new version which will run under WINDOWS can be considered as a new OCR system based on the former RECOGNITA.

The majority of RECOGNITA users will have the intention to upgrade their old program version for the new WINDOWS version. SZKI will offer for its registered RECOGNITA users upgrade possibilities.

## Desktop OCR Takes off

OCR is actually very important, because many companies have typewritten or printed information, and it makes sense to input it by scanning, rather than re-keying. This presents the user with software as well as hardware considerations. A potential customer needs to see a scanner working with appropriate software.

Looking at the international peripheral market: last year's favourite was the desktop scanner. There are now several products on the market with a resolution from 200 dpi up to 800 dpi and in a price range from less than 1000 \$ up to more than 10000 \$.

There are two approaches to mainstream scanning hardware: the flatbed and the rotary. Generally, rotary scanners are less expensive than the flatbed models, but they are also less versatile. Alignment is particularly important with optical character recognition because several OCR programs flatly refuse to recognize even a familiar typewriter face like Courier 12 unless it is straight on the page. That is why rotary scanners are scarcely used for OCR purposes, only with OCR programs which are misalignment tolerant. RECOGNITA - SZKI's OCR program - tolerates misalignment of paper in the scanner, so it can be used with rotary scanners.

Most scanners come with, or offer as an option, software for controlling the scanning process. Some companies also provide

or offer "paint" programs, which can be used to drive the scanner or just to edit the scanned images. Many image-scanning packages run under Microsoft Windows.

In most cases, an automated software installation routine copies the programs to the hard disk drive and asks all the necessary questions about the hardware.

However, users will need one to two megabytes of hard disk space to accommodate the software, and probably several more megabytes for the large data files created.

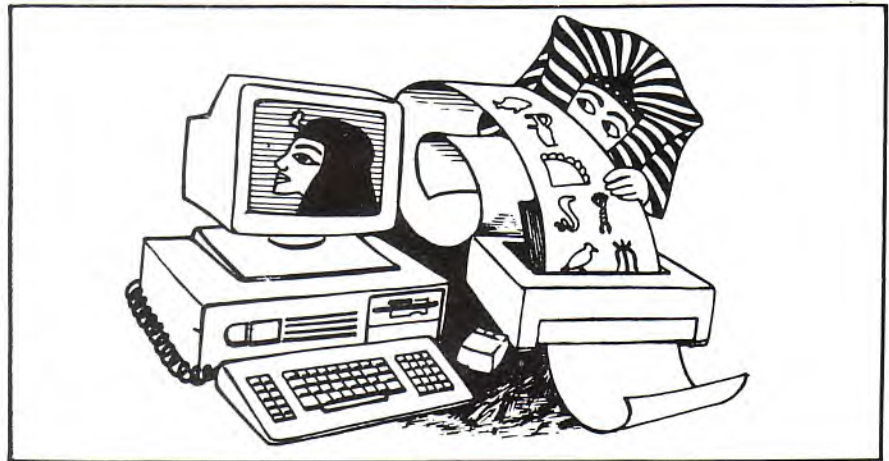
Optical character recognition (OCR) programs for personal computers were very expensive - until last year. In 1988, however,

a lot of OCR programs appeared which can scan characters for documents into text files stored on disk with remarkable accuracy and at a reasonable price.

The OCR programs should read an A4 page with less 5 errors in less than 1 minutes to be attractive for IBM PC users.

Perhaps the ultimate challenge for an OCR system is reading in typeset text from magazines, books and other printed media.

As scanners and OCR software grow in sophistication, more products will be able to read typeset text. Meanwhile, the increased flexibility and improved accuracy of OCR systems make the scanning of text not only possible, but quite plausible.



## Better Understanding OCR

### Font-independent OCR system

Early OCR technology required that the text for scanning be typed in specific fonts. Nowadays OCR systems can read several type-styles commonly used in the office. This font-independence is one of the most important requirements in OCR systems. What does it mean for the user if an OCR system is font-independent? It means both that

- the user need not tell the system what

font(s) it is reading;

- the different fonts and sizes may be mixed on a page or even within paragraphs.

One of the factors that make the recognition more difficult and so the possibility of error may increase, are the accents.

### The accents

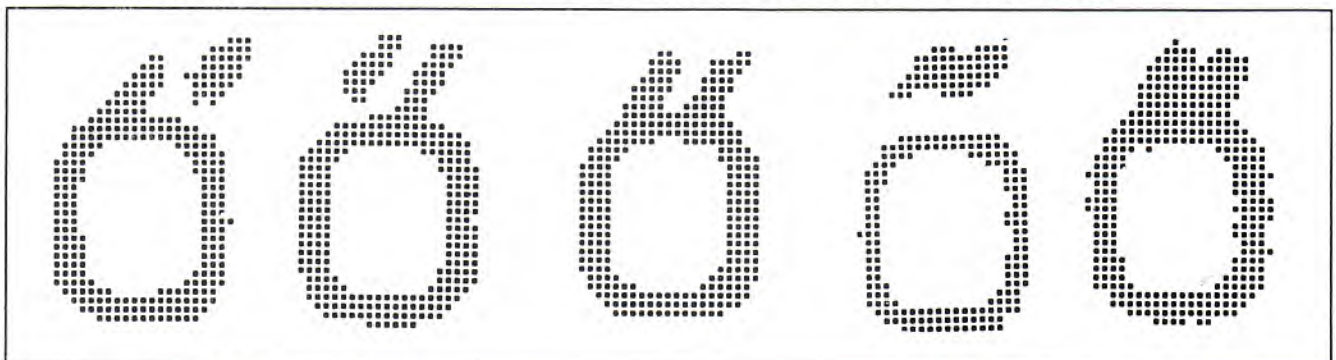
There is no problem if the accent doesn't touch the letter itself. Problems occur if the accents and the letters run into each other,

or if the accent is very small.

### For example: "ö" or "ü"

In the case of the letter "ö" there are the following possibilities:

- one of the dots touches the letter
- the other one touches the letter
- both dots touch the letter
- the two dots run into each other and
- the two dots run into each other and touch the letter.



# RECOGNITA Meeting in Budapest

A two-day RECOGNITA Meeting will be organized to distributors and dealers in Budapest, on 29-30th May.

The aim is to inform the participants not only about the up-to-date development results of RECOGNITA, but to give a better technical understanding of the basics of OCR. Sales technics and other problems will be discussed too. Further on partici-

pants will have the possibility to test the up-to-date version of RECOGNITA. SZKI provides the instructors and the equipment needed.

Although two days are a short period, but it shall be very effective, since you can gather a lot of information from the members of RECOGNITA's development and sales team.

There is no charge for the two-day meeting in Budapest.

If you are announcing your participation before the end of April, SZKI will make your hotel reservation in Budapest. Distributors and OEM partners will be informed about the detailed schedule of the meeting, and we do also encourage retailers to contact them for more information.

## Shaking hands

List of scanners supported by RECOGNITA V.1.30:

- AGFA-Focus
- Advanced Vision Research
- Canon IX-12, IX-12F
- CompuScan PCS
- Dest PC Scan 2000
- HP Scanjet
- IBM Page Scanner 3119 (PS/2 Modell 50 and upwards)
- MICROTEK MS-300, MS-300A, MS-300C, MSF-300C, MSF-300G
- Panasonic FX-RS505
- RICOH IS30M2, IS30M4, SS30, IS10, IS11
- SIEMENS HighScan 300, 400, 600, 800
- TELIC AGD-300

List of supported ADF's:

- Hewlett Packard
- Microtek
- RICOH
- CompuScan

## Introduction

RECOGNITA NEWS is SZKI's periodical publication for RECOGNITA distributors, dealers and users. RECOGNITA NEWS informs you about the development, marketing and applications of RECOGNITA, further on about OCR knowledge. Users' letter are welcome by RECOGNITA NEWS.

You can get RECOGNITA NEWS from your RECOGNITA dealer or directly from SZKI, if you are a registered RECOGNITA user.

For direct mail write RECOGNITA NEWS, Donáti-u. 35-45, BUDAPEST, I. HUNGARY, H-1015. Don't forget, you must be a registered RECOGNITA user. Users become registered, if they send their Registration Card via their dealers to SZKI.

## Worth reading

Optical Character Recognition and Image Scanning

**Faulkner Electronic Office** (Office Automation) Vol.2. April 1988

Black, L. - Chalmers, D.

Product Comparison: Image Scanners

**Info World**, Vol.10. Issue 28. July 11, 1988 p.41, 44-47, 50-52

O'Malley, Ch.

Putting Scanners...

**Personal Computing**, Vol 12, No.6, June 1988. p.114.

Dalton, R.

The Growing Sophistication of Desktop Scanners

**Lotus**, Vol. 4. No. 7, July 1988.p. 11-12.

Lónyai L.

RECOGNITA - Felismert lehetőség

**Mikrovilág**, 4.évf. 8.sz. 1988. ápr. 27. p.24.

Glenny, M

Finding new ways to pay old debts

**The Guardian**, May 5, 1988

Gergely, S.M.

Wie der Computer lesen lernt

**Trend** 5/88 p.268.

Felismerések

**CWI Szoftver**, 1988/5 p.10.

Optical Character Recognition Program

**Byte**, 13/6, June 1988.

p.88A-18

Recognita liest 120 Seiten pro Stunde  
**PC Magazin**, 8. Juni 1988

Lesesoftware

**Computer-Forum**, 7/88, Juli 1988. p.60.

Copony, H-P.

RECOGNITA - Intelligent und preiswert

**PC Magazin**, No. 28, 6. Juli 1988. p.61.

Eine Technologie sucht ihre Anwendung

**Computerwelt**, No. 14, 22. Juli 1988. p.10.

Grassberger, R.

Die Monotonie sinkt, die Freude steigt: Der Computer kann lesen

**Kurier**, 26. Juli 1988. p.11.

RECOGNITA - Texterfassung mit Komfort

**Personal Computer**, No.8, August 1988. p.52.

Recognita-OCR-Software

**Mikro-Report** 4, Aug./Sept. 1988 p.44.

Scanner-software mit mehr Leistung  
**Computerwoche**, 9. Sept. 1988. p.19.

CHIP-TEST: Recognita

Software lernt lesen

**Chip**, No.12., Dec.1988. p.2.

Többet látó lapolvasó

**CWI Software**, 1988.No. 9. p.2.