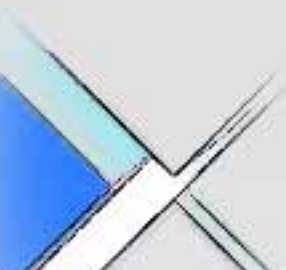




# Vincent

OCR | ICR | OMR | barcode solutions  
for automated acquisition of data  
and document images





## Vincent

OCR|ICR|OMR|barcode solutions for automated acquisition of data and document images

Computers enable the management of a great deal of data.

Yet, a high percentage of information is still created and distributed on paper. Processing this information requires a large amount of data acquisition work, which proves to be slow, costly, repetitive and affected by a high percentage of error.

*Vincent* has been designed to provide a solution to automatic data entry and document storage process needs.

*Vincent* is an integrated solution for digital acquisition and automatic recognition, using completely software-based modules, or dedicated hardware with standard PCs or LAN systems.

*Vincent* includes the following modules:

- **Acquisition:** interface module for the most common scanners available on the market using SCSI, ISIS, TWAIN, KOFAX, DUNORD and other drivers.
- **Recognition:** module for recognition of document images acquired via scanner, using high-performance OCR and OMR engines.
- **Correction:** module for controlling and validating recognized data.
- **Export:** module for transferring recognized data to Host Computers.
- **Application editor:** module for automatically generating or modifying applications.

## Advantages

*Vincent* is the ideal instrument and the starting point for all operations involved in the document processing management and in the automation of business processes using Workflow systems.

Among the most important advantages offered by *Vincent*, you can:

- Lower, or even eliminate, costly manual preparation and correction processes for digitally acquired paper documents.
- Optimize scanner performances.
- Reduce processing times and costs of resources.
- Improve the quality and accuracy of information acquired via scanner, reducing margins for error, thanks to the image pre-processing functionalities and Intelligent Character Recognition, as well as automatic correction.

*Vincent* is an open and scalable system, it can be easily integrated with ERP and other and document management applications, optimizing information and business processes and removing inefficiency.

With *Vincent* you can improve your company productivity and the services for your customers and suppliers, taking advantage of the document management system features such as automatic indexing and quick information retrieval.

### Data management using any database

*Vincent* manages the acquisition of images and data, as well as comparative tables, using the most common DBMS on the market, such as Oracle™, Informix™, SQL Server™, Access™, Paradox™ or others.

Besides providing data management security, this allows the creation and the management of databases to improve the recognition and the correction of acquired data, independently from the OCR engine.

The use of ODBC and OLE-DB drivers enables online comparisons with files saved in the Host Database, making it easy access data on legacy systems.

### Using form design, including non-optical forms

*Vincent* gives you the opportunity to avoid using optical forms (in green, red, or blue).

All the functions for removing the image noise and for correcting the image rotation enable the removal of lines, dots and grids (in black), thus reconstructing any characters possibly cut by these lines.

### Acquiring images: integration of the pre-processing tool

With this module you can manage the scanner (both with a video interface and a compression card, or with SCSI) and send the acquired images to the character recognition engine. Image quality can be improved via the special pre-processing functions integrated into the Image Enhancement tool.

*Vincent* also allows you to scan multi-page documents.

Acquired images can be matched with a range of support data, entered by the operator before scanning.

You can use preset fields that can be saved in the record layout (for example: date, operator's name, account balancing).

### Recognition module

The recognition engine analyzes the portion of image containing information and transforms it into ASCII format data. Several different recognition engines are available, both in terms of the characters to be recognized (handwritten, typewritten, checkboxes, barcodes) and recognition speed. There is also an option for using multiple recognition engines simultaneously. Further, a range of tools can be activated to process, with excellent results,



### Independence from the recognition engine

Optionally, *Vincent* can use the most specialized OCR engines on the market, such as, for example, Mitek™, Caere™, Abbyy™, CGK™, RE™ and others.

It also provides an OMR engine, developed entirely by *siav*, specialized in recognizing checkboxes. Users can select the recognition performance in terms of characters per second, ranging from 5 char./sec. up to 2,400 char./sec.

### Module for creating and modifying applications

With *Vincent* you can create and/or modify a character recognition application quickly and easily, as well as acquire information from paper documents. More complex applications can be created quickly without writing any line of code.

### Automated workflow management

The document workflow is completely automated.

The processing status of each group of acquired documents is stored using special database tables.

*Vincent* can be assigned the task of identifying which operations must be performed on each batch (recognition, correction, transfer). Even the selection of the position of acquired documents can be passed on to *Vincent*. Simply you can insert the sheets into the scanner and acquire them, without the need to establish the batch number, the document progressive number, etc.

even documents that are not suited to automatic recognition, such as, for example, the realignment of images through reference points (targets, logos, etc.).

### Customizing character recognition

The fields to be read can be customized to obtain the best results: by defining the features of the characters in the field to be read (typewritten, handwritten, OCR A and B, FARRINGTON 7B, CMC7, BARCODE); by indicating the type of character (alphanumeric, alphabetic, numeric, marked field); by recognizing single characters (for example, when acquiring you can specify that the first 6 characters are letters, the following two are digits followed by a letter, etc.); by identifying the character set found in the field.

### Module for correcting and validating the recognized data

The correction and validation module integrates any non-interpreted characters. The use of ICR systems and the sophisticated automatic data control/correction functions make corrections faster and more effective. Thanks to the connection to external databases via OLE-DB and ODBC, to the Thesauri and selfincreasing archives created by *Vincent*, you can compare, by similarity or analysis, the context of a partially recognized character with those already stored (Full Text and Fuzzy Logic), and you can automatically complete the character string, increasing the quality of recognition up to 100%. Correction speed is quite high.

## Reacquiring documents

If the correction phase reveals that a document has been incorrectly acquired (e.g.: black borders, excessive rotation, too dark/light, etc.), the document can be re-scanned directly from the correction station, immediately replacing the underlying image, or marking the image as invalid and postponing its reacquisition to a later time (if, for example, no scanners are currently connected).

## System configuration

*Vincent* allows maximum flexibility. In fact, you can define the configuration of all of its components, according to your needs, and in particular:

scanners with characteristics, speed and automatic feeder capacity that meet your needs | higher speed recognition units (multiprocessor cards or dedicated hardware) | correction stations consisting of simple personal computers, the number of which can be increased later.

## Data export module

Once the data validation phase is complete, you can format them according to the record layout set for the destination data management system, and then transfer the data to this system. Images and acquired data are pre-processed in order to be imported into *Galileo*, or into any other program enabled to the import of ASCII formatted data and TIFF images. The stored data, appropriately indexed, are used as search keys for retrieving the saved documents. You can process data and images to be transferred to the archival system, which takes care of the automatic indexing of the search keys and stores the images in CD ROMs, optical disks or juke boxes. It is also possible to export data directly from *Vincent* to an external database (both in Windows or in the Host system) via OLE DB and ODBC, reducing time-consuming data management and allowing real-time access to information.

## System Administration

With the administration module you can access the system tables, view the list of documents used by each application, remove some or all no longer used documents, and define the general features of an application (work directory, workflow).

You can also view statistics regarding document processing as well as create customized reports using Crystal Report.

## Control libraries

*Vincent* provides over 150 predefined control routines, which can be called directly by the user.

Some examples are listed below:

automatic insertion of the most widely used check digits (module 10, 97, etc.) | automatic control on codes and VAT numbers | automatic balancing on import fields (e.g.: partial and/or total sums) | threshold value settings for considering a character correctly acquired | mandatory fields input | date validity control | automatic conversion of marked fields into alphabetic or numeric characters | access to internal or external databases (via OLE DB/ODBC) to check the validity of the acquired data | use of fuzzy searching to perform automatic corrections on acquired fields, etc.

You can also customize the applications using Visual Basic Scripting.

A development kit is available to developers to create autonomously customized control libraries for particular applications.

## Technical specifications

Supported platforms	Windows 95, 98, Windows NT, Windows 2000 (recommended); Internet Explorer 4.0 or greater
Server requirements	Pentium III; 256 MB RAM; SCSI disk system needed for larger applications; specific solutions may require higher-performance systems
Client requirements	Pentium III; 128 MB RAM
Standalone installation	Pentium III; 128 MB RAM
Supported scanners	All SCSI scanners with TWAIN interface (Canon, Ricoh, Fujitsu, InoTec, Kodak, etc.), scanners with Kofax interface; Scantron optical readers
Database	Export and logical control of the data access to any DBMS that supports OLE DB/ODBC (SQL Server, Oracle, etc.)
Network protocols	TCP/IP; the server must have a static address

