

Capitolo 2

INFORMAZIONI GUIDATE

Readiris è un pacchetto di OCR che rappresenta lo stato dell'arte, dotato di numerose caratteristiche avanzate. Questo capitolo illustra tutte le principali caratteristiche, informazioni utili e consigli relativi all'uso di Readiris.

AVVIARE IL SOFTWARE

Fare clic sull'applicazione Readiris nel sottomenu "Applicazioni I.R.I.S. - Readiris" o sul collegamento dell'applicazione Readiris sul desktop.



Vengono visualizzati la videata di avvio e la finestra dell'applicazione Readiris. La videata di avvio visualizza la versione e il copyright del software Readiris. Facendo clic con il mouse in qualsiasi punto, la videata scompare.



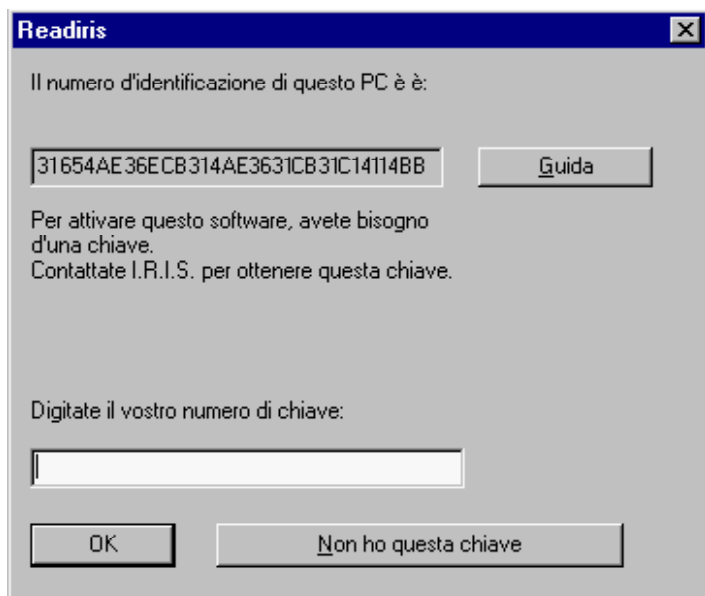
La finestra seguente riguarda la procedura guidata d'OCR; fare click su Cancella per il momento.

IL PRIMO AVVIO

In base al pacchetto software acquistato, per il primo avvio può essere necessaria una procedura particolare: è possibile che venga richiesta la registrazione della licenza.

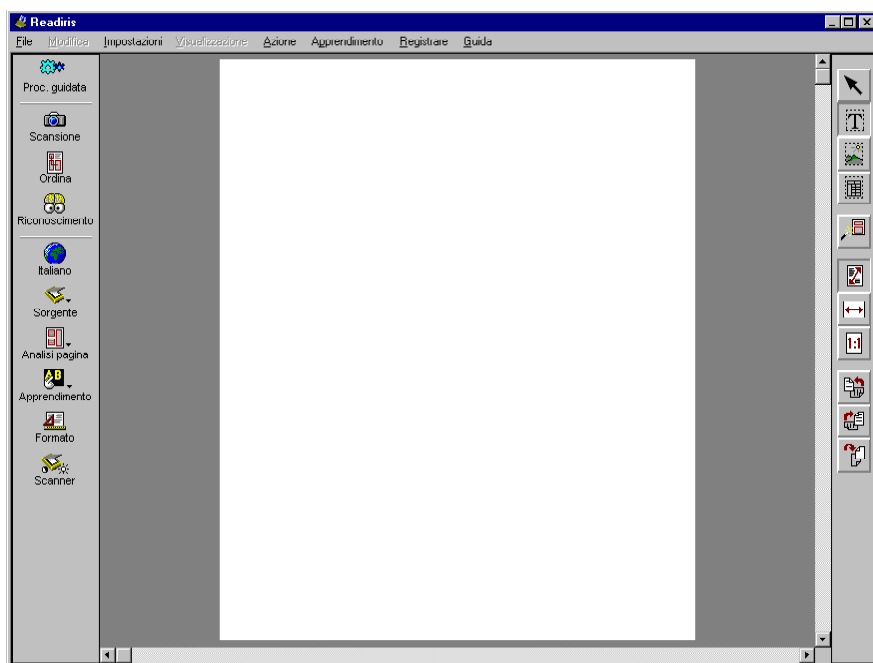
In questo caso, la possibilità di utilizzare Readiris è limitata ad un mese. In seguito alla registrazione, I.R.I.S. vi fornirà un **tasto personalizzabile** che consente di continuare ad utilizzare Readiris dopo il primo mese.

Per creare il tasto personalizzabile è necessario il **numero di identificazione** personale; durante la registrazione della licenza assicurarsi che tale numero sia disponibile o indicato.



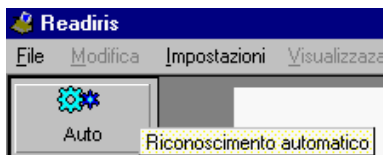
SCOPRIRE L'INTERFACCIA READIRIS

La finestra dell'applicazione Readiris non solo contiene **menu dei comandi**, ma anche due barre di pulsanti che consentono un rapido accesso a tutti i comandi utilizzati di frequente. Inizialmente alcuni menu dei comandi relativi all'anteprima non sono visibili e fino a quando non viene aperta alcuna immagine non sono disponibili.



Lo stesso vale per la **barra degli strumenti immagine** posta sul lato destro della finestra dell'applicazione: questa contiene tutti i comandi necessari durante l'anteprima dell'immagine. La **barra degli strumenti principale** posta sul lato sinistro consente un rapido accesso a tutti i comandi generali utilizzati di frequente.

Per scoprire quale comando corrisponde ad un determinato pulsante, tenere per qualche secondo il puntatore del mouse sul pulsante: una **casella informativa** indicherà all'utente la funzione del comando.



Infine, il riquadro della finestra o la **zona di immagine** è l'area in cui vengono visualizzate le immagini scansionate. Per riconoscerle è possibile portare i file di immagini nella zona di immagine.

UNA PRIMA ESERCITAZIONE

Il modo migliore per acquisire dimestichezza con il funzionamento di Readiris è indubbiamente utilizzarlo. Unitamente al software viene fornita una serie di **immagini prescansionate** che consentono di iniziare persino quando lo scanner non è collegato al computer. Segue la spiegazione.

Il pulsante “Sorgente” della barra dei pulsanti permette di indicare se si intende utilizzare un'immagine scansionata o prescansionata come fonte di immagini. Readiris consente di aprire immagini TIFF (non compresse, pacchetti di bit, compresse di Gruppo 3 e Gruppo 4), immagini Paintbrush (PCX) e bitmap Windows (BMP). Sono supportate solo le immagini in bianco e nero. Questa proprietà è particolarmente utile per convertire i **fax** in file di testo editabili.

Se si desidera aprire un'immagine prescansionata, selezionare il disco e non lo scanner come fonte di immagini, mediante il pulsante “Sorgente” della barra degli strumenti principale.



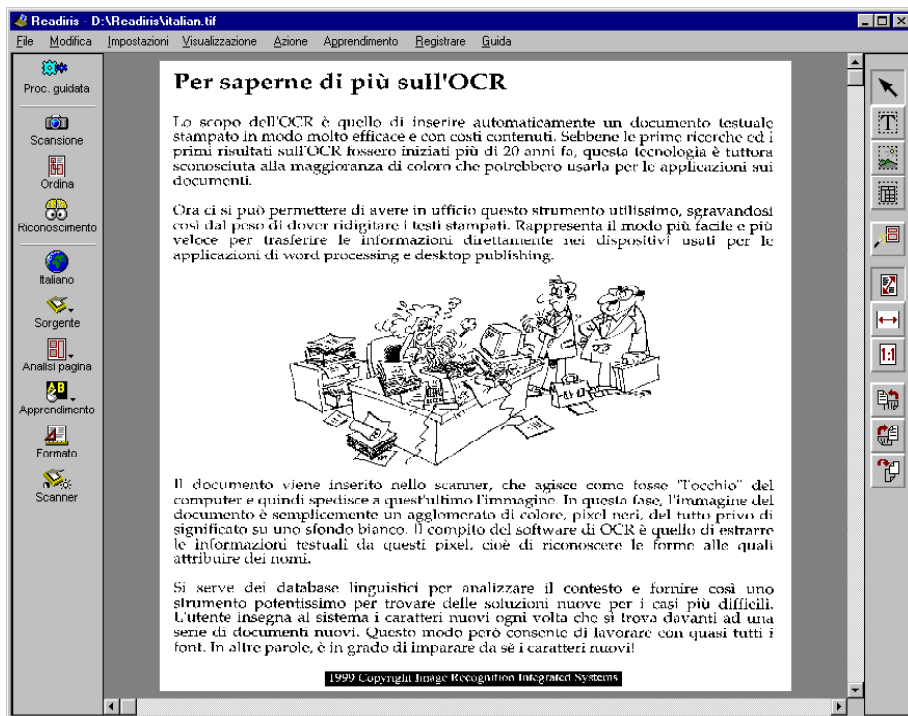
Cliccare quindi sul pulsante “Apri”. (Selezionando il disco come sorgente di immagine, il tasto “Scansione” viene sostituito dal tasto “Apri”, mentre il corrispondente comando “Scansione” del menu “Elaborazione” viene sostituito dal comando “Apri”).



È possibile inoltre selezionare il comando “Apri” dal menu “File” e aprire direttamente un’immagine prescansionata - questa operazione è possibile anche quando la fonte di immagini corrente è lo scanner.



Il sistema chiederà di selezionare un file di immagine. Selezionare il file ITALIAN.TIF della cartella Readiris. L’immagine viene letta dal disco e visualizzata progressivamente nella zona di immagine.



Un terzo modo per aprire immagini prescansionate è costituito dall'utilizzo di **“trascinare selezione”**. Trascinare un'immagine da Esplora Risorse (Explorer) di Windows alla zona di immagine Readiris consente di aprire l'immagine immediatamente.

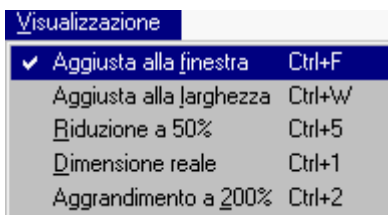
La barra degli strumenti immagine posta sul lato *destro* della finestra dell'applicazione Readiris contiene tutti i comandi necessari durante l'anteprima dell'immagine, relativi all'indicazione delle zone di interesse, alla rotazione dell'immagine, al suo zoom, ecc.



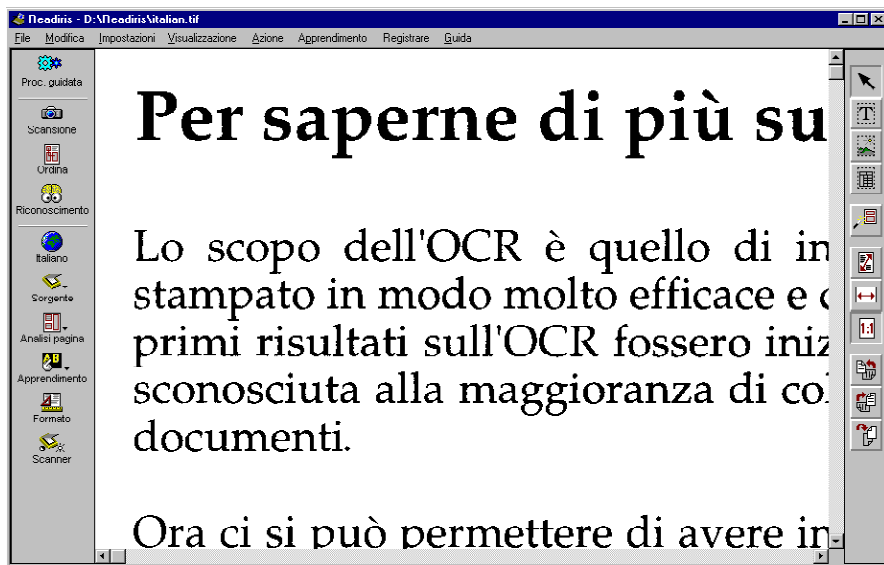
ZOOMARE SULLE IMMAGINI

Readiris è dotato di numerosi comandi che consentono di **zoomare** sull'immagine scansionata, ad es. per verificare la qualità della scansione.

La barra degli strumenti immagine contiene pulsanti che consentono di zoomare fino alle dimensioni reali, di adattare l'immagine all'ampiezza della pagina e di inserire l'intera immagine nella finestra di anteprima. Il menu "Visualizzazione" contiene gli stessi comandi ed è dotato di due ulteriori livelli di zoom: è possibile visualizzare l'immagine al 50% e al 200% delle dimensioni reali. In dimensioni reali, un pixel dello schermo corrisponde ad un pixel dell'immagine. (Per tutti i livelli di zoom sono disponibili tasti di scelta rapida!)



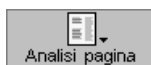
Infine, facendo clic con il pulsante destro del mouse su un'area dell'immagine scansionata è immediatamente possibile ingrandirla. Facendo nuovamente clic con il pulsante destro l'immagine verrà riportata alle dimensioni iniziali.



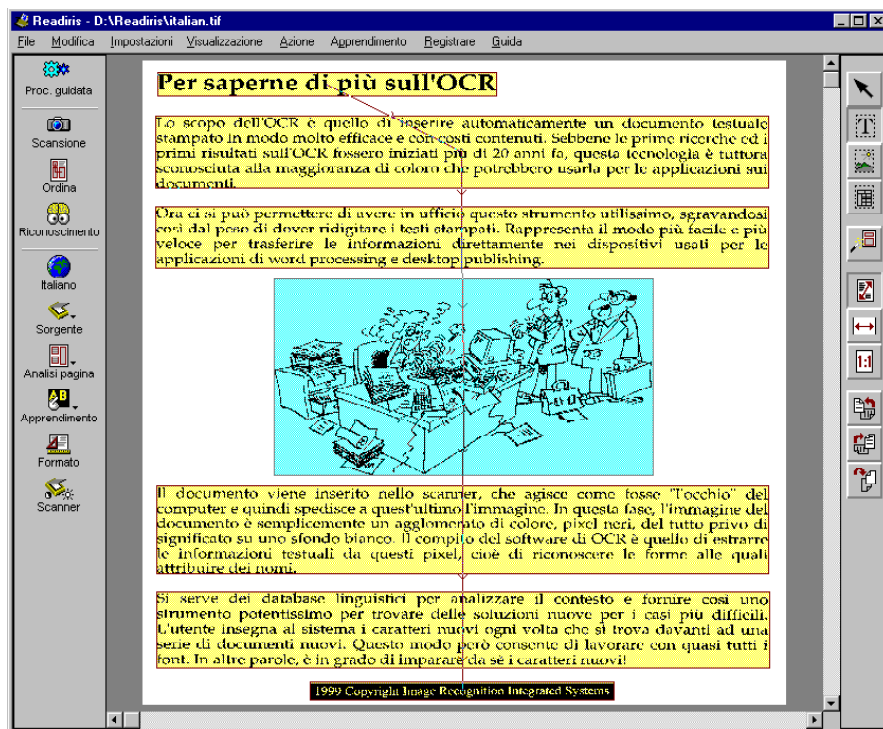
UNO, SCOMPORRE UN'IMMAGINE SCANSIONATA

Una volta scansionata l'immagine, è necessario indicare quali parti si desiderano convertire in testo editabile, tracciando strutture definite "finestre" intorno alle zone di interesse.

In realtà, abilitando l'opzione "Analisi Pagina" della barra degli strumenti principale, Readiris lo farà automaticamente.



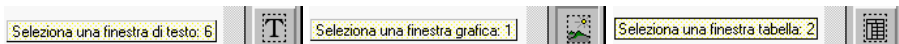
La scomposizione automatica della pagina risulta particolarmente utile quando vengono riconosciuti **testi formati da colonne** e documenti che presentano un complesso layout di pagina, comprendente grafici e tabelle.



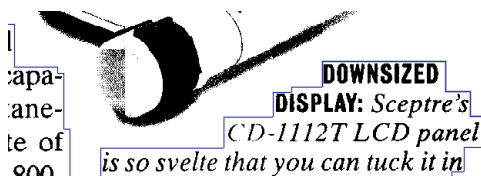
La scomposizione della pagina utilizza **tre tipi di finestra**: testo, grafica e tabella. Readiris riconosce i blocchi di testo, le tabelle e le zone di grafica contenenti foto, illustrazioni, ecc. nella pagina. (Il salvataggio dei grafici e il riconoscimento delle tabelle sarà illustrato più avanti.)

Un **codice colore** indica il tipo di finestra: le zone di testo sono in giallo, i grafici in blu e le tabelle in bordeaux.

Il numero delle finestre è sempre indicato nelle caselle informative degli strumenti “Finestra di Testo”, “Finestra di Grafica” e “Finestra di Tabella”.



L'analisi della pagina è veloce, estremamente accurata ed è in grado di riconoscere righe inclinate: traccia forme complesse, “irregolari”.



L'analisi della pagina individua zone in cui compare **testo bianco su sfondo nero**. Il riconoscimento di tali inserimenti non costituisce alcun problema: mentre l'anteprima visualizza sullo schermo il documento correttamente scansionato, quando si presenta la necessità di riconoscere tali blocchi di testo Readiris “inverte” l'immagine! (Inoltre è possibile creare un'immagine completamente capovolta dallo scanner per elaborare pagine con testo bianco su sfondo nero. Vedere di seguito.)

UNO E MEZZO, SELEZIONE DELLE FINESTRE

Readiris non solo individua i vari blocchi ma li seleziona: le zone vengono selezionate per default a testa in giù, da sinistra a destra, per adattarsi ai documenti organizzati in colonne.

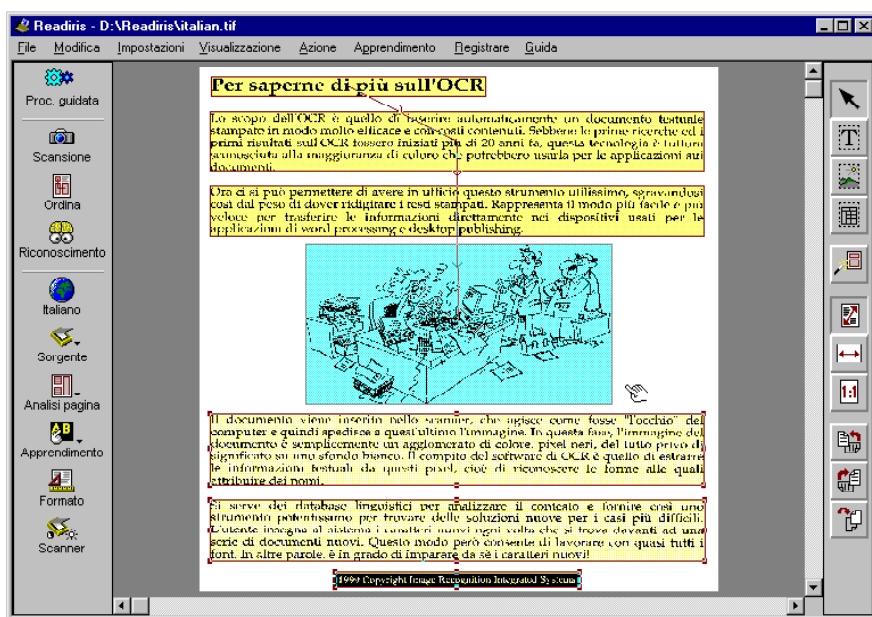
È possibile modificare l'**ordine di selezione**. Cliccare sul tasto “Seleziona” sulla barra strumenti principale. Il cursore del mouse assume la forma del puntatore non appena il modo “Seleziona” viene abilitato.



Cliccare sulle finestre che si desidera includere. Le finestre *non* selezionate vengono semplicemente ignorate ed escluse dal riconoscimento. È facile vedere



quali finestre sono selezionate o meno: le finestre selezionate hanno un colore più intenso, quelle non selezionate un colore più tenue.



L'analisi della pagina è abilitata per default. Qualora si sia involontariamente disabilitata l'analisi della pagina o cancellato alcune finestre e si intenda rieseguire un'analisi della pagina, per fare in modo che Readiris scomponga la pagina corrente è sufficiente fare clic sul pulsante "Analizza Pagina" della barra degli strumenti immagine.

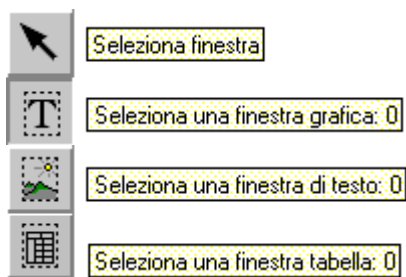


(Selezionate la lingua del documento *prima* d'eseguire l'analisi della pagina quando trattate dei documenti asiatici. Delle routine specifiche sono utilizzate per

queste lingue: lo spazio tra righe è d'abitudine più grande nei documenti asiatici che nei documenti occidentali, il testo è composto di piccoli grafici ("ideogrammi") che potrebbero facilmente essere visti come delle zone grafiche nei documenti occidentali ed il testo si legge d'alto in basso, di destra a sinistra.)

DUE, ORGANIZZARE MANUALMENTE IN FINESTRE UN'IMMAGINE SCANSIONATA

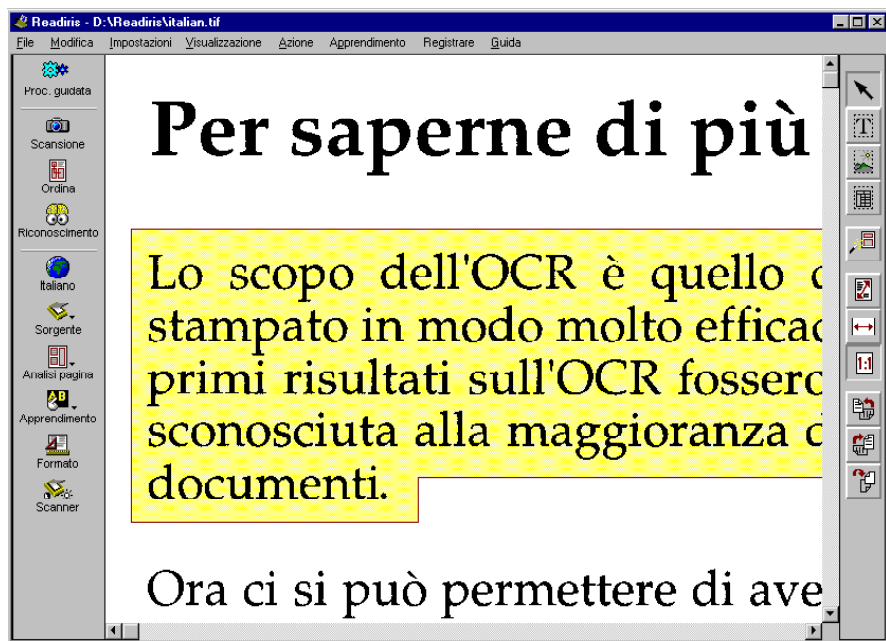
L'analisi della pagina consente di organizzare automaticamente in finestre una pagina scansionata. In alternativa, è possibile dividere manualmente in zone un'immagine mediante gli **strumenti di organizzazione in finestre** di Readiris.



Per creare un **rettangolo** intorno a una zona di interesse, selezionare lo strumento corrispondente nella barra degli strumenti immagine, cliccare con il cursore nell'angolo superiore sinistro della finestra, allungare la finestra spostando il mouse nell'angolo inferiore destro e cliccare nuovamente. (Non sono consentiti lati inferiori a 1 mm perché non potrebbero contenere neppure un singolo carattere.)

Le finestre vengono ordinate automaticamente sulla base dell'ordine di creazione: le frecce indicano la sequenza di ordinamento.

È possibile anche incorniciare i blocchi di testo "irregolari" tracciando intorno ad essi **finestre poligonali**. Unendo zone rettangolari si creano finestre non rettangolari: quando due rettangoli (dello stesso tipo) si intersecano, diventano automaticamente un'unica finestra! È come costruire una casa aggiungendo una stanza dopo l'altra... (Non ha senso creare finestre di tabelle poligonali.)



L'organizzazione in finestre manuale può essere associata con la selezione delle finestre: è possibile disegnare nuove finestre anche quando il “modo seleziona” è attivato. Utilizzare la funzione “seleziona” per includere un numero di finestre individuate e per creare manualmente altre finestre se l'analisi della pagina non ha dato i risultati desiderati. Non appena inizia l'operazione di creazione della finestra nel “modo seleziona” tutte le finestre non selezionate saranno immediatamente cancellate!

Per modificare, spostare ed eliminare finestre, è necessario innanzi tutto **selezionarle**. Per fare ciò, selezionare lo strumento “Selezione Finestra” (lo strumento “freccia”) della barra degli strumenti immagine e fare clic all'interno di

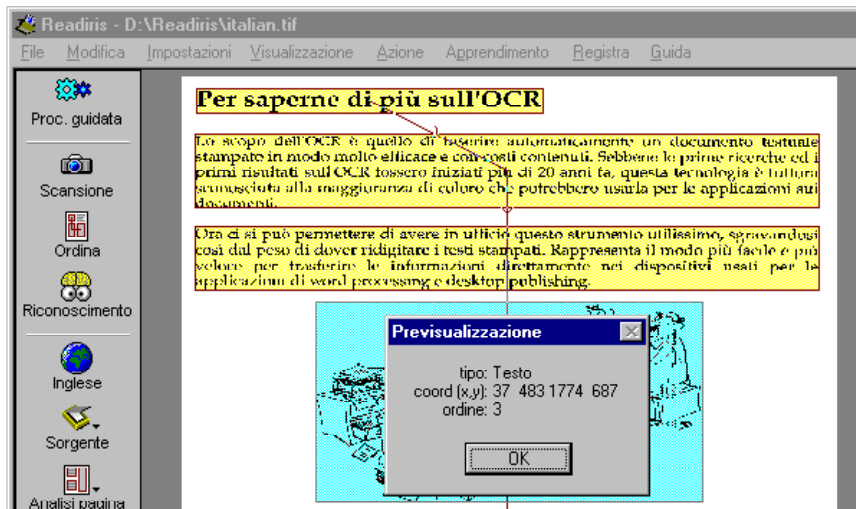
una finestra. Compaiono così marcatori rettangolari in ciascun angolo e al centro dei lati della finestra.



Per saperne di più sull'OCR

Per **delesezionare** finestre, fare clic con il pulsante del mouse in qualsiasi punto. Per selezionare **ulteriori finestre**, tenere premuto il tasto Mai facendo clic su tali finestre supplementari. Per selezionare una finestra e le **finestre incluse** (dello stesso tipo), tenere premuto il tasto Ctrl facendo clic sulla finestra principale.

Premendo il tasto Alt durante la selezione, è possibile visualizzare il tipo, le coordinate dei pixel e la sequenza di ordinamento di una finestra!



Queste sono le operazioni relative alla selezione delle finestre. Per **modificare** una finestra di testo, selezionarla, posizionare il cursore del mouse su un marcatore e trascinare il lato per cambiare le dimensioni della finestra.



Per **spostare** una finestra, è sufficiente selezionarla e trascinarla in un'altra posizione.

Per **eliminare** finestre, selezionare la (le) finestra (finestre) e scegliere il comando “Taglia” o “Cancella” del menu “Modifica”. Il comando “Taglia” taglia la (le) finestra (finestre) trasferendola (trasferendole) in una memoria temporanea interna, mentre “Cancella” elimina definitivamente la (le) finestra (finestre). Quando si incollano finestre, queste vengono inserite nella loro posizione originale ed è necessario trascinarle nella nuova posizione.

Infatti, alle finestre si applicano *tutti* i comandi noti del menu “Modifica”, che consentono di eliminare, tagliare, copiare ed incollare finestre! È possibile utilizzare anche il comando “Ripristina”: nel caso in cui alcune finestre siano state involontariamente eliminate, spostate, ridimensionate, ecc., mediante il comando “Ripristina” sarà possibile cancellare l'ultima operazione.

Modifica	
Ripristina	Alt+BkSp
Taglia	Ctrl+X
Copia	Ctrl+C
Incolla	Ctrl+V
Cancella	Delete
Seleziona Tutto	Ctrl+A

Inoltre per tutti i comandi sono disponibili tasti di scelta rapida! Segue un esempio: per eliminare tutte le finestre esistenti, è possibile scegliere il comando “Seleziona Tutto” del menu “Modifica” o il relativo tasto di scelta rapida Ctrl+A e fare clic sul comando “Cancella” del menu “Modifica” o sul relativo tasto di scelta rapida Elimina. A questo punto è possibile ricreare il layout necessario. Per ripristinare il layout precedente, è possibile scegliere “Ripristina” o il relativo tasto di scelta rapida Alt+Backspace.

TRE, SALVARE MODELLI DI ORGANIZZAZIONE IN FINESTRE

Per impighi successivi, i layout di organizzazione in finestre ottenuti possono essere salvati come **modelli di divisione in zone** mediante il comando “Salva Finestre” del menu “File” e caricati in memoria mediante il comando “Carica Finestre”.



Qualora si renda necessario riconoscere documenti con un layout simile, ad es. una relazione di 50 pagine in cui, per ovvie ragioni, sarebbe necessario escludere l'intestazione e il piè di pagina, è possibile utilizzare un singolo modello per dividere in zone tutte le 50 pagine.

Quando si carica un modello in memoria, l'analisi della pagina viene disabilitata automaticamente e il modello di divisione in zone resta attivo fino a quando dalla barra degli strumenti principale non si riabilita l'analisi della pagina.

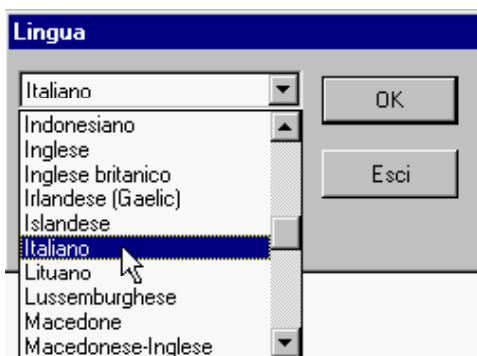
READIRIS PERMETTE DI VIAGGIARE IN TUTTO IL MONDO

Ammesso che le finestre siano definite correttamente, ora il testo è quasi pronto per il riconoscimento dei caratteri. “Quasi” perché la lingua e le impostazioni del documento non sono ancora state verificate.

L'impostazione della lingua viene eseguita dalla barra degli strumenti principale.



Per modificare la lingua del documento, fare clic sul pulsante “Lingua”.



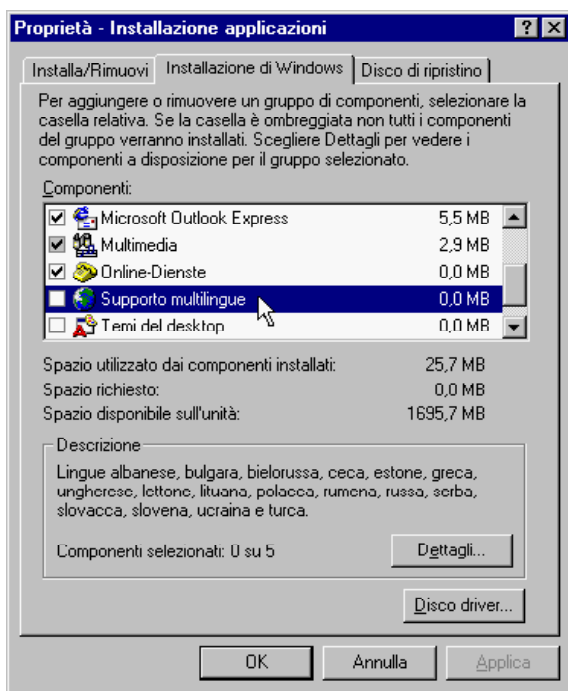
Per accedervi direttamente, premere un tasto lettera: qualora la lingua attualmente selezionata sia l'inglese e si desideri selezionare per esempio l'ungherese, è possibile cliccare sul tasto "H" della tastiera per passare direttamente alla lingua ungherese. Nel caso in cui numerose lingue abbiano la stessa iniziale, per accedere alle opzioni premere più volte la lettera. Ad esempio: Readiris legge il polacco e il portoghese. Premendo "P" una volta viene selezionato il polacco, premendo "P" una seconda volta viene selezionato il portoghese e premendo "P" una terza volta, si torna al polacco. (Per passare a un'altra lettera, ad esempio T, premere il tasto di ritorno prima di digitare il carattere "T").

Readiris non si limita all'inglese ma è in grado di gestire fino a **55 lingue!** Supporta infatti tutte le lingue americane ed europee, incluse le lingue dell'Europa centrale, il greco, il turco, il cirillico ("russo") e le lingue caltiche.

Potete leggere opzionalmente dei **documenti asiatici**: il modulo supplementario "OCR asiatico aggiuntivo" offre il riconoscimento del giapponese e del cinese semplificato. Il cinese tradizionale non è supportato. (Il cinese semplificato è usato nella maggior parte della Cina, allora che il cinese tradizionale è usato dalle 5 grande comunità - chiamate "Big 5" - ossia Hong Kong, Taiwan, Singapore etc.)

Inoltre l'inglese britannico viene distinto dall'inglese americano - o si dovrebbe dire "internazionale"?

Per visualizzare i caratteri delle lingue dell'Europa centrale, del greco, del turco, del cirillico e delle lingue baltiche, è necessaria la configurazione Windows appropriata. Per consentire al sistema di Windows di gestire queste lingue, può essere necessario installare il modulo di Windows "Supporto Multilingue". Per verificare che questo modulo di Windows sia installato nel PC, selezionare l'icona "Aggiungi/Rimuovi Programmi" del "Pannello di Controllo". In base al pacchetto software acquistato, è possibile che il CD-ROM Readiris contenga informazioni dettagliate sulle modalità di installazione del modulo di Windows.





Per vedere e modificare dei documenti asiatici, dovete installare una versione giapponese o cinese del sistema operativo Windows. Alternativamente, potete utilizzare un software “d’emulazione” specializzato (del tipo UnionWay AsianSuite o TwinBridge AsianBridge) su una versione di Windows occidentale per rappresentare correttamente gli ideogrammi di queste lingue asiatiche.

È indispensabile selezionare la lingua del documento appropriata. Basandosi sulla lingua selezionata, il software sa quali simboli riconoscere. Il supporto multilingue assicura il riconoscimento corretto di caratteri “esotici”, quali ad es. ç, ß, ñ, γ e ø.

Inoltre, per convalidare i risultati ottenuti, il software fa largo uso di **database linguistici**. Si supponga sia necessario leggere la parola “presidente”, dove una macchia d’inchiostro rende la “r” simile alla “f”. Consultando il vocabolario d’italiano, Readiris scoprirà autonomamente di dover leggere la parola “presidente” e che non ha alcun senso riconoscere il simbolo “f”. È chiaro che questa **tecnica di “autoapprendimento”** è strettamente legata al contesto linguistico.

I criteri di linguistica offrono un utile aiuto per risolvere **casi ambigui**, come nel caso della lettera “O” che potrebbe essere confusa con lo “0”. Un altro esempio tipico è costituito dalla lettera “l” e dal numero ‘1’ aventi forma identica in molti font - si pensi ai testi prodotti dalle vecchie macchine da scrivere! Il contesto linguistico aiuta a determinare se si tratta di una “l” o di un ‘1’.

La figura sotto riportata mostra varie forme di ‘1’ e “l”. Le forme poste sulla prima riga risultano ambigue così come quelle poste sulla seconda riga, tuttavia i criteri di linguistica consentono di distinguerle. Qualora il contesto non sia sufficiente, interviene l’utente.

193 1950s. 1hr
Well, Rossellini

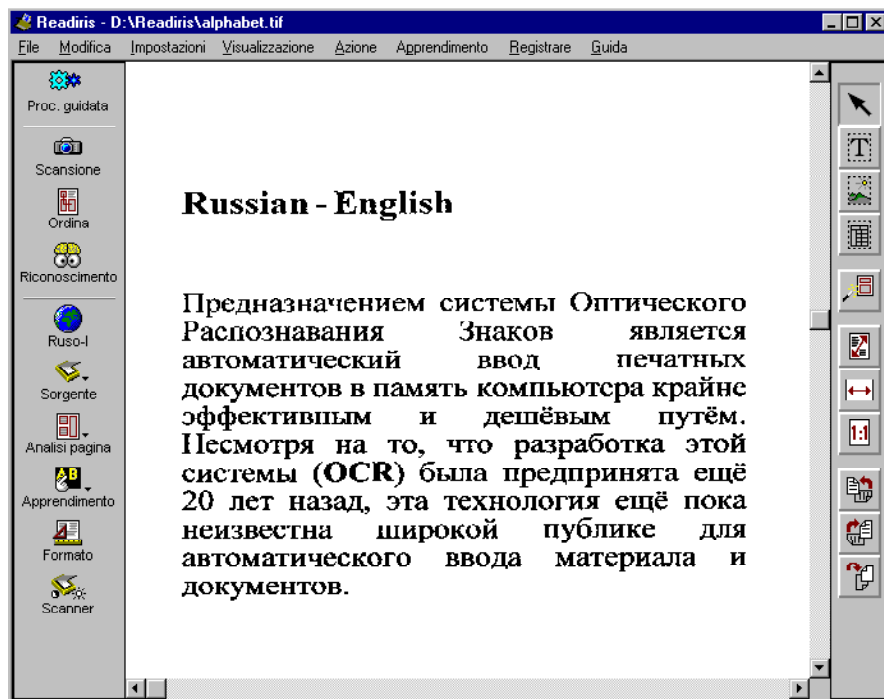
SE NECESSARIO, READIRIS CAMBIA LE LINGUE

Ma Readiris fa molto di più: Readiris è in grado di commutare tra più lingue all'interno di una frase senza alcun aiuto da parte dell'utente! Quando parole "occidentali" compaiono in documenti in lingua greca, cirillica o asiatica - molti nomi propri non traslitterabili, marche, ecc. vengono scritti con comuni simboli occidentali - Readiris è in grado di adattare automaticamente l'alfabeto corretto. In altre parole, è possibile attivare un **alfabeto misto** di caratteri greci, cirillici o asiatici con caratteri occidentali.

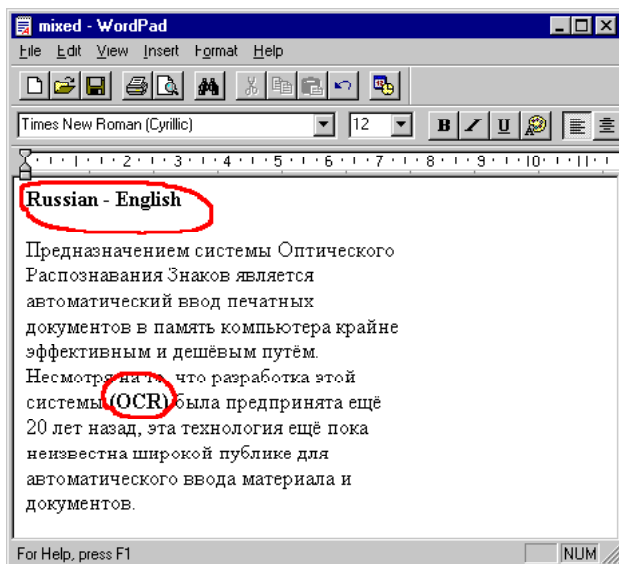
Assicurarsi di selezionare "Greco-Inglese" o la lingua cirillica appropriata - ad es. "Bielorusso-Inglese". In altre parole: non tentare di selezionare solo "Greco" o "Bielorusso" come lingua del documento sperando che i simboli occidentali risultino chiari!



Segue l'esempio di un testo russo contenente alcune parole inglesi - nel caso in cui si desideri tentare autonomamente, aprire il file di immagine ALPHABET.TIF!



Quando viene aperto in un programma di videoscrittura, il risultato finale è il seguente - per visualizzare correttamente il testo russo, può essere necessario selezionare un **font** cirillico.



Per **combinare tra loro altre lingue**, è sufficiente selezionare la lingua che presenta il set di caratteri più esteso. In caso di un documento nel quale ad es. la traduzione francese è posta accanto al testo inglese, per garantire che caratteri quali ç, é e ù vengano riconosciuti correttamente, è necessario selezionare la lingua francese.

DEFINIRE LE CARATTERISTICHE DEL DOCUMENTO

Una volta impostata la lingua, si passerà a specificare le altre caratteristiche del documento, ad es. il tipo di font e il passo dei caratteri, impostando con precisione il riconoscimento del documento stesso. (Questi comandi non si applica ai documenti asiatici.) Segue ora la spiegazione.



Innanzitutto, indicare se la stampa dei documenti di cui si sta effettuando il riconoscimento è “normale” o a **matrice di punti** mediante il comando “Tipo di Font” del menu “Impostazioni”.



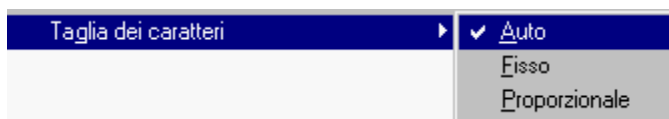
È di fondamentale importanza che questo parametro venga impostato correttamente: i simboli a matrice di punti “bozza” o “9 pin” sono composti da punti isolati, separati e per riconoscerli sono necessarie procedure di riconoscimento estremamente specializzate.

ape-descended life

La stampa a matrice di punti “qualità corrispondenza”, chiamata anche matrice di punti “25 pin”, richiede l’impostazione normale, proprio come le **qualità di stampa** di materiale tipografico, dattiloscritto, stampato su laser e a getto d’inchiostro. Ovviamente “Normale” è il valore di default.

Il tipo di font è indicato nella casella strumenti del tasto “Riconosci”; se non viene aggiunto alcun messaggio nella casella strumenti, verrà attivata la qualità di stampa “normale”. Quando nella casella strumenti compare il messaggio “Matrice di punti” viene attivato il modo corrispondente.

È possibile impostare il **passo dei caratteri** mediante il comando “Taglia dei Caratteri” del menu “Impostazioni”.



Con font *fissi* o “monospaziati”, tutti i simboli del font sono della stessa larghezza. In questa frase ad esempio, una “i” e una “w” occupano in senso orizzontale lo stesso spazio sulla riga. Si pensi ai documenti prodotti con una macchina da scrivere, in cui il carrello lascia lo stesso spazio per ogni simbolo dattilografato.

Con passo *proporzionale* si intende che la larghezza di un carattere dipende dalla sua forma. Simboli come *m* e *w* sono più larghi, occupano più spazio in senso orizzontale su una riga rispetto ai caratteri stretti *l* o *j*. In genere, tutti i libri, riviste e giornali sono stampati con passo proporzionale.

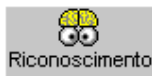
Courier, un font fisso
Letter Gothic, un font fisso

Century Schoolbook, un font proporzionale
Arial, un font proporzionale

La soluzione più semplice consiste nel lasciare ogni volta questa opzione impostata sul valore di default “Auto”, in tal modo Readiris individuerà automaticamente il passo dei caratteri.

READIRIS DIVENTA OGNI VOLTA PIÙ INTELLIGENTE!

Una volta selezionata la lingua ed impostate le caratteristiche del documento, è possibile fare clic sul pulsante “Riconoscimento”.



La procedura di OCR viene indicata sullo schermo. Per interrompere il riconoscimento del testo, fare clic sul pulsante “Stop”.



Al termine del riconoscimento, Readiris passa alla fase di apprendimento interattivo in cui l'apprendimento è abilitato dal pulsante “Apprendimento” della barra degli strumenti principale. L'apprendimento interattivo è abilitato per default.

(L'apprendimento interattivo non si applica in documenti asiatici: l'apprendimento non ha senso per queste lingue che usano migliaia di simboli differenti - e di più dovete essere capace di introdurre l'ideogrammi, cosa non facile con delle tastiere occidentali!)

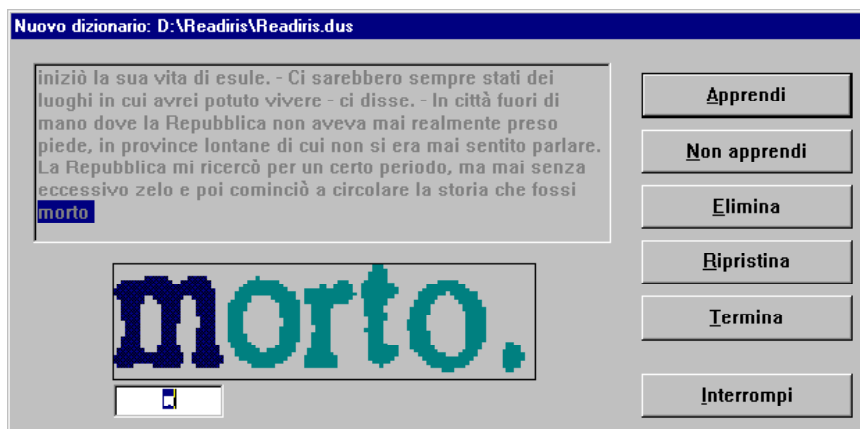


L'addestramento ai font può migliorare notevolmente la precisione del sistema di riconoscimento. Quando l'utente prova a leggere caratteri distorti e illeggibili, quali si trovano in documenti reali o forme di font stilizzate che Readiris non riconosce in modo ottimale, l'addestramento consente di superare questo temporaneo “insuccesso”.

L'apprendimento mediato dall'utente viene anche utilizzato per addestrare il sistema a **simboli speciali** che Readiris non è in grado di riconoscere, quali i simboli matematici e scientifici e i segni grafici decorativi. Ad esempio, è possibile

addestrare Readiris a riconoscere il simbolo “ π ” come “pi” o il segno grafico “☎” come “Tel”. (Tuttavia, la lista dei simboli riconosciuti non può essere ampliata con i simboli “ π ” e “☎”!)

Il testo riconosciuto viene visualizzato progressivamente e il sistema si arresta su caratteri dubbi, o - se si tratta di caratteri in contatto (“legature”) - su stringhe di caratteri dubbi. I caratteri dubbi vengono evidenziati e presentati sempre nel loro contesto. I caratteri che non sono stati riconosciuti vengono rappresentati con una tilde (il simbolo “~”).



Verificare innanzitutto di aver attivato il dizionario di font e la modalità di dizionario appropriati - sempre indicati nel titolo della finestra di apprendimento. In caso contrario, fare clic sul pulsante “Interrompi” - viene così rivisualizzata l’immagine del documento con la divisione in zone iniziale -, abilitare il dizionario di font o la modalità di dizionario appropriati ed eseguire nuovamente l’OCR. (Seguirà una breve spiegazione sul funzionamento dei dizionari dei font.)

Se necessario, sostituire la forma errata o sconosciuta con un carattere (o stringa di caratteri) e fare clic su uno dei seguenti pulsanti.



Apprendi

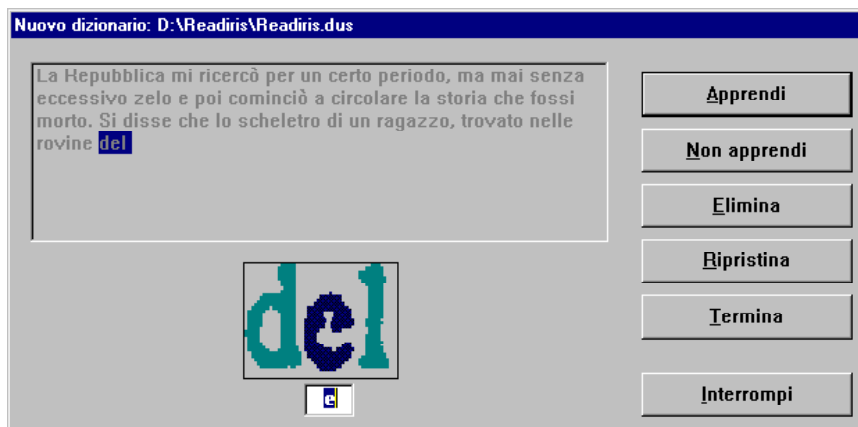
Approvare la soluzione proposta o correggerla. Il programma salva questo carattere dubbio nel dizionario di font come “certo”, definitivo. Per un riconoscimento successivo non verrà più richiesto l’intervento dell’utente, in quanto la forma viene considerata definitivamente appresa.

Nell’esempio sopra indicato, il sistema si ferma su un carattere non pulito e facendo clic su “Apprendi” si accetta una forma che non può essere confusa con altri caratteri.

Non Apprendi

Approvare la soluzione proposta o correggerla. La differenza rispetto al pulsante “Apprendi” è data dal fatto che il simbolo appreso viene classificato nel dizionario come “**incerto**”. Per un riconoscimento successivo il sistema proporrà la soluzione appresa, ma richiederà un’ulteriore conferma.

Questo pulsante viene utilizzato per simboli che potrebbero essere confusi con altri: una “e” illeggibile che potrebbe essere scambiata per una “c”, una “t” danneggiata che somiglia molto ad una “r”, ecc.



Nel caso della “e” sopra indicata, danneggiata a tal punto da somigliare al simbolo “c”, l’utente dovrebbe fare clic sul pulsante “Non Apprendi”, in modo tale da non confonderla con il simbolo “c”.

Elimina

La forma visualizzata viene eliminata dall’output. Questo pulsante viene utilizzato per ignorare elementi di disturbo presenti sui documenti - chiazze, macchie di caffè, ecc. - che potrebbero essere riconosciute come punti, virgole ecc. - e per cancellare ogni altro simbolo non desiderato.

Ripristina

Questo comando consente di tornare a correggere gli errori, annullando le ultime nove decisioni.

Termina

Il processo di apprendimento viene interrotto, ma l’OCR prosegue in modalità automatica. Tutte le decisioni prese da questo momento in poi dal sistema vengono accettate senza la convalida dell’utente.

Fare clic su questo pulsante quando il riconoscimento è estremamente preciso e non richiede una revisione dettagliata.

Non confondere il pulsante “Termina” con il pulsante “Interrompi”: con “Interrompi” non si ottiene alcun output e si ricomincia tutto da capo, con “Termina” si crea un testo che però non viene controllato dettagliatamente!

IL RUOLO DEI DIZIONARI DEI FONT

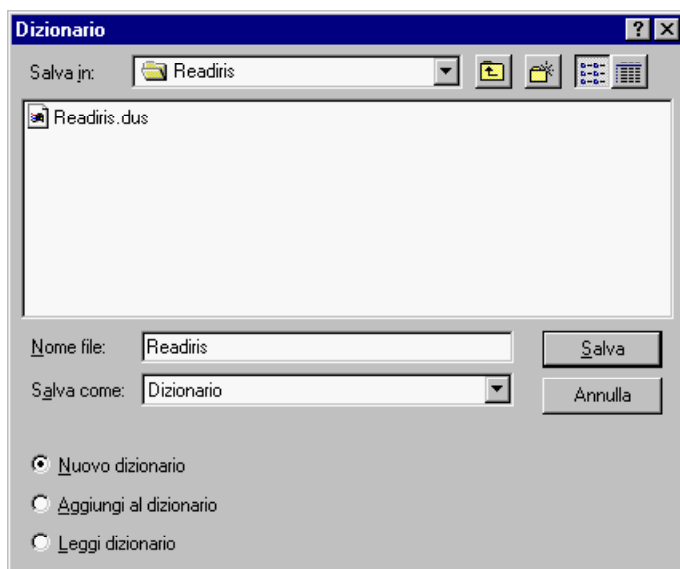
I risultati di ciascuna sessione di addestramento vengono conservati temporaneamente nella memoria del computer ma, per impieghi successivi, è possibile e necessario memorizzarli in file definiti “dizionari”.

Quando si desidera riconoscere documenti simili, è necessario caricare in memoria questi dizionari dei font per utilizzare l’intelligenza supplementare che



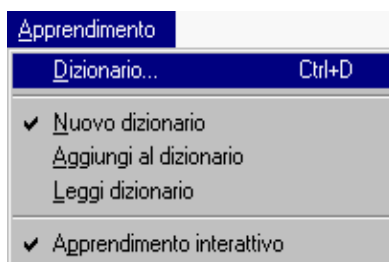
contengono. In questo modo Readiris tiene conto dell'intelligenza memorizzata in queste biblioteche dei font. Si potrebbe affermare che ad ogni impiego Readiris diventa più intelligente!

Come è possibile questo? Il funzionamento dei dizionari dei font è controllato dal menu "Apprendimento": selezionare un dizionario con il comando "Dizionario" e determinarne la modalità di funzionamento.



I **dizionari** dei font sono limitati a 500 forme, pertanto si consiglia di creare dizionari separati per applicazioni specifiche, ad es. per tipo di documento. Per motivi di chiarezza, si consiglia di assegnare ai font dei dizionari nomi significativi, ad es. RELAZION.DUS, PALATINO.DUS, ecc. I dizionari hanno per default l'estensione *.DUS. Quando il dizionario è pieno, l'addestramento non ha più alcun effetto: i risultati dell'apprendimento non vengono più conservati in memoria o scritti in un dizionario.

È possibile impostare la modalità di dizionario all'interno del comando "Dizionario" o direttamente dal menu "Apprendimento". Sono disponibili tre modalità di dizionario: nuovo, aggiungi e leggi.



Selezionando "Nuovo Dizionario", i risultati del processo di apprendimento verranno salvati in un *nuovo* dizionario. (Qualora si selezioni un dizionario esistente, i suoi contenuti verranno cancellati.)

Con la modalità "aggiungi" i risultati del processo di apprendimento verranno salvati in un dizionario *esistente*: il riconoscimento si avvale di un'intelligenza supplementare già presente nel dizionario, a cui vengono aggiunte nuove forme di font. In altre parole, questa opzione consente di creare man mano un dizionario di font.

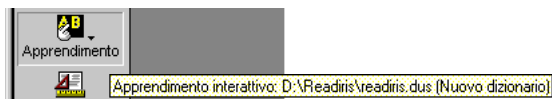
(Quando si assegna un nome di file ad un nuovo dizionario e si attiva la modalità aggiungi, viene creato un dizionario di font vuoto che deve essere completato.)

Con l'ultima opzione, "Leggi Dizionario", il dizionario funziona nella modalità di sola lettura: si utilizza il dizionario senza aggiungervi *nuove* forme di font.

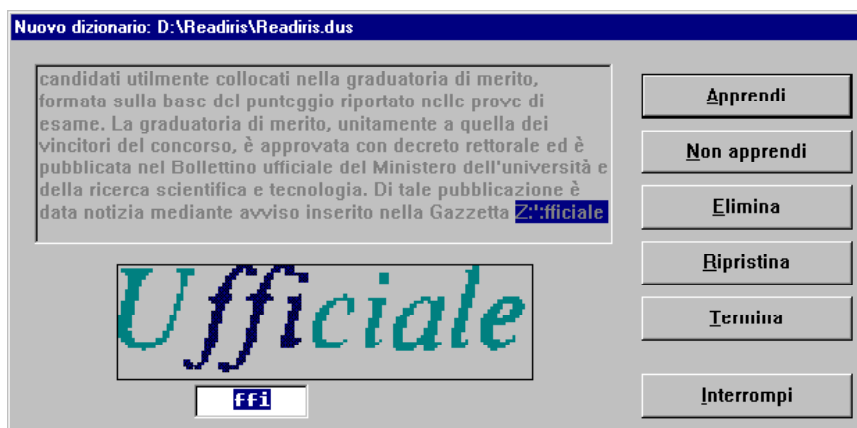
Selezionare la modalità nuovo quando si esegue il riconoscimento di una sola pagina. Per riconoscere numerose pagine dello stesso tipo - pagine con gli stessi font e la stessa qualità di stampa - selezionare la modalità nuovo per la prima pagina, la modalità aggiungi per alcune pagine successive e la modalità leggi per la parte restante del (di) documento (documenti). Quando vengono riconosciuti documenti costituiti da più pagine, la scansione di una nuova pagina porta automaticamente il dizionario utente nella modalità aggiungi.



La casella informativa del pulsante “Apprendimento” indica sempre quale dizionario di font è attualmente attivo e in quale modalità opera.



Quando si passa all’apprendimento interattivo, nel titolo della finestra vengono indicati il dizionario e la relativa modalità di funzionamento; qualora essi siano errati, fare clic sul pulsante “Interrompi” e ricominciare.



INVIARE DIRETTAMENTE IL RISULTATO ALL'APPLICAZIONE

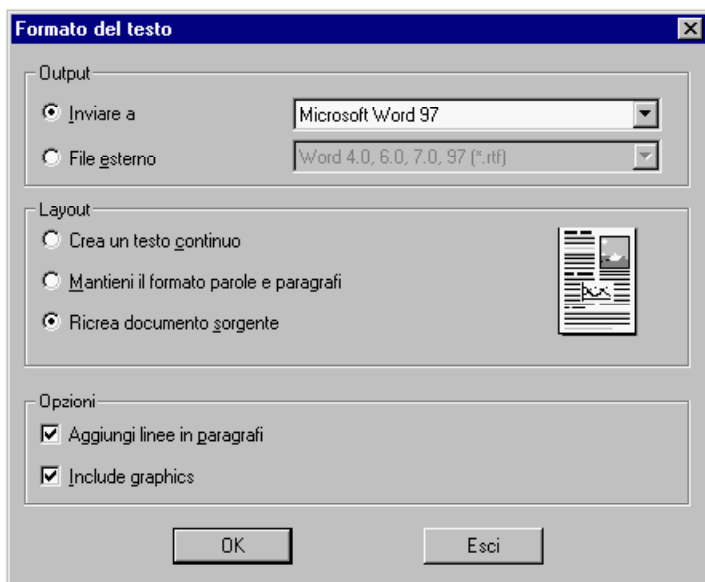
L’addestramento interattivo pone termine al riconoscimento dei caratteri. Dal momento che Microsoft Word 97, agisce per default come target di output, il sistema di videoscrittura viene avviato automaticamente al termine del riconoscimento (se necessario). Il testo riconosciuto viene inserito.

L'immagine scansionata viene nuovamente visualizzata con la divisione in zone creata, per essere disponibile per ulteriori elaborazioni, restando visualizzata fino a quando non si esegue la scansione di un'altra pagina.

In questo modo si è convertito un documento cartaceo in un file editabile, impiegando tempi da 10 a 25 volte inferiori rispetto a quelli necessari per la ribattitura manuale! Procedere e confrontare il testo con l'immagine posta all'interno della finestra di Readiris.

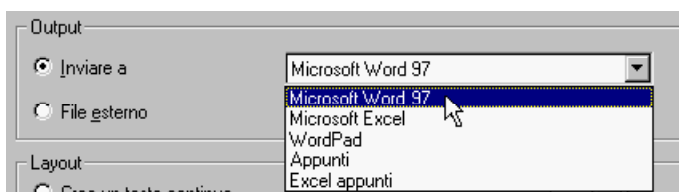
Readiris dispone di tre metodi diversi per salvare il risultato OCR: inviare direttamente il documento riconosciuto a una applicazione target, salvare il risultato in un file esterno e copiare il risultato negli appunti di Windows.

Utilizzando il tasto "Formato" sulla barra degli strumenti principale (o il comando "Formato testo" nel menu "Impostazioni") viene selezionato il **target output**.





La funzione “Invia a” consente un collegamento diretto OCR fra lo scanner e le applicazioni Windows: **inviare** i documenti scansionati direttamente al sistema di videoscrittura o al foglio elettronico. Readiris esporta i documenti riconosciuti direttamente in Microsoft Word 97, Microsoft Excel e l'accessorio di Windows WordPad.



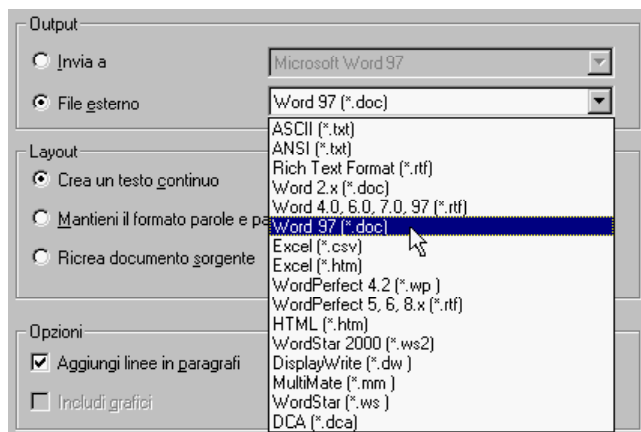
Al termine del riconoscimento, viene avviata l'applicazione target e aperto il documento riconosciuto all'interno di un nuovo file di testo o foglio di lavoro.

Attendere durante il caricamento
Microsoft Word 97...

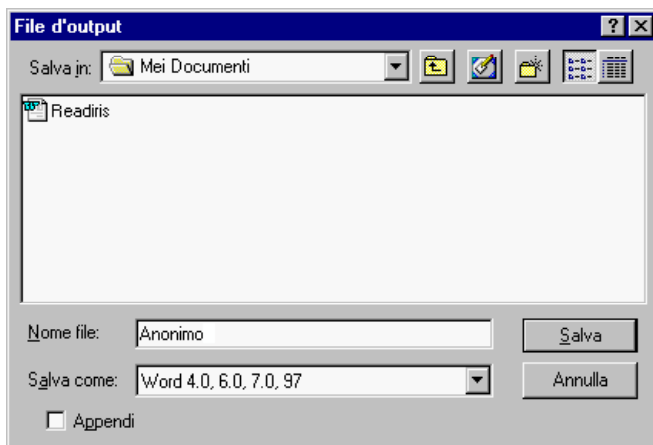
Non dimenticare che l'opzione “Invia a” consente anche di copiare il testo riconosciuto negli **appunti** di Windows. Non vi è quindi alcuna necessità di esportare il risultato o di salvarlo in un file esterno!

SALVARE I RISULTATI IN UN FILE DI TESTO

È possibile scrivere il risultato OCR in un file “esterno”. Readiris supporta un'ampia gamma di formati file comprendenti tutti i sistemi di videoscrittura e fogli elettronici più noti - Microsoft Word (DOC), RTF e HTML, ecc.



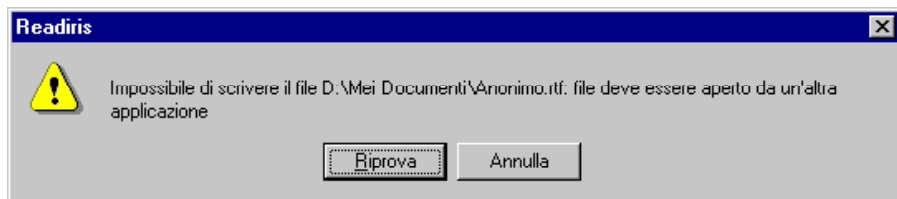
Selezionando l'opzione "File esterno", vi viene chiesto di salvare il testo riconosciuto al termine della fase di riconoscimento.





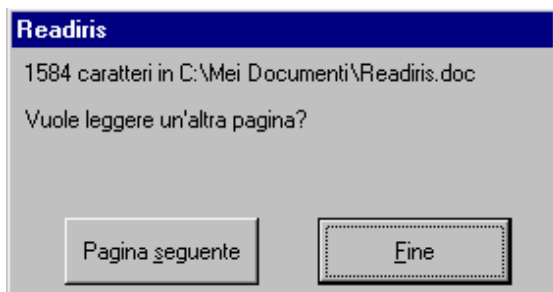
È possibile creare un nuovo file e aggiungere testo a un file esistente. Per questo motivo esiste l'opzione "Aggiungi". È possibile, ad esempio, aggiungere due pagine di testo scansionate a un documento da voi creato. Aggiungendo testo a un file esistente, assicurarsi che non sia attualmente aperto, perché non sarebbe possibile lavorarvi all'interno.

È possibile creare un nuovo file e aggiungere un file di testo esistente - questa è la funzione del pulsante "Appendi". Ad esempio, è possibile aggiungere due pagine di testo scansionate ad un testo scritto personalmente. Qualora si aggiunga un file esistente, assicurarsi che al momento non sia aperto, poiché in tal caso non sarebbe possibile scrivervi!



RICONOSCIMENTO DI PAGINE MULTIPLE

Cliccando su "Salva" per salvare il testo, il sistema suggerisce di leggere un'altra pagina o di non leggerla. (Nel caso in cui il testo riconosciuto venga inviato a una applicazione target o agli appunti, ovviamente non viene richiesto di salvare il testo.)



In questa fase vengono fornite alcune brevi **statistiche** relative al numero dei caratteri appena letti.

È ora possibile aprire il testo riconosciuto con il programma di videoscrittura o editor di testi, importarlo nel software di desktop publishing o in qualsiasi applicazione basata su testo.

L'esercitazione su una singola pagina termina qui. Ma come salvare il testo di pagine successive? O in altre parole: come elaborare documenti costituiti da molteplici pagine? In realtà è molto semplice: quando il sistema suggerisce di leggere un'altra pagina, fare clic su "Pagina Seguente". In questo modo, nel caso in cui lo scanner funzioni come fonte di immagini, verrà immediatamente scansionata un'altra pagina. Successivamente, il file di output ed il dizionario di font vengono posti automaticamente in modalità "aggiungi", così da potere aggiungere il testo e continuare comodamente l'addestramento ai font.

È possibile procedere con il riconoscimento di altre pagine: il nuovo testo viene aggiunto al file di testo ogni qualvolta viene effettuato il salvataggio dopo il riconoscimento.

ORGANIZZAZIONE DELL'OUTPUT DI TESTO

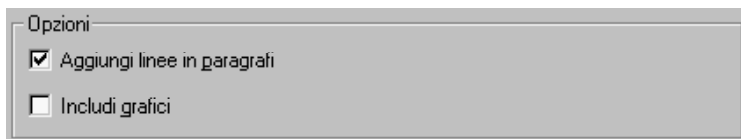
Salvare o esportare il testo significa molto più che selezionare un metodo di output o assegnare un nome di file al file di output. È possibile anche selezionare



un formato file e determinare l'aspetto del testo riconosciuto. In breve, prima di lanciare l'esecuzione è necessario decidere dove si desidera prendere il testo.

Alcune opzioni del tasto “Formato” consentono di modificare l'aspetto dell'output di testo.

Il **flusso di testo** del documento di output è direttamente influenzato dall'opzione “Aggiungi Linee in Paragrafi” del pulsante “Formato”.



Per fare in modo che Readiris individui i paragrafi, mantenere questa opzione abilitata: Readiris applicherà così il normale **ritorno a capo automatico** tipico dei programmi di videoscrittura, altrimenti dopo ciascuna riga viene aggiunto un ritorno del carrello e le parole separate da trattino rimangono tali! L'individuazione dei paragrafi è abilitata per default.

Segue un esempio chiarificativo. Quando le prime tre righe di una colonna sono: “Il nuovo presi-”, “dente salutava dal balcone.” e “Sua moglie lo aveva raggiunto.”, l'individuazione del paragrafo fornisce il seguente risultato: “Il nuovo **presidente** salutava dal balcone. **Sua** moglie lo aveva raggiunto.” Le parti separate da trattino della parola “presidente” sono state “re-incollate” ed è stato aggiunto uno spazio alla fine della frase, creando un testo fluente.

Se l'individuazione del paragrafo *non* fosse stata abilitata, sarebbe stato mantenuto il layout originale, con un ritorno del carrello aggiunto alla fine di ogni riga.

(Il pulsante “Formato” contiene alcune opzioni di formattazione non ancora trattate - l'argomento verrà trattato tra breve.)

INSTALLAZIONE DELLO SCANNER

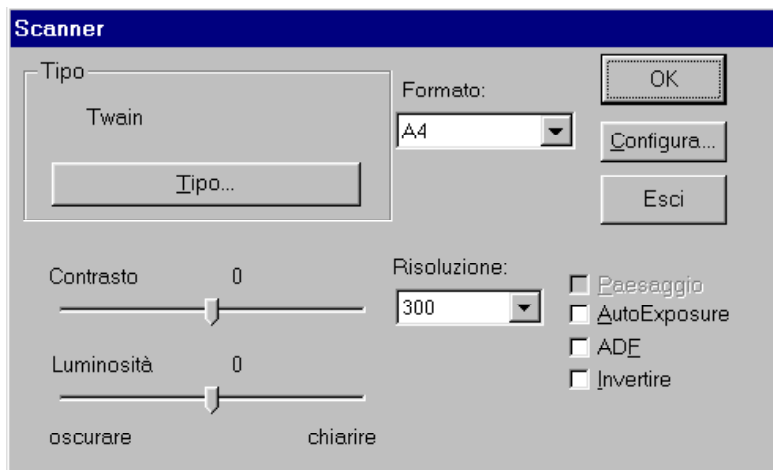
Si proceda ora all'installazione dello scanner. Si presume che l'hardware dello scanner ed i driver necessari siano installati correttamente.

Nel caso in cui la licenza del software Readiris sia allegata ad un modello di scanner, questa operazione può non essere necessaria in quanto lo scanner potrebbe già essere installato sotto Readiris.

Fare clic sul pulsante “Scanner” della barra degli strumenti principale.



Per determinare il **modello dello scanner** in dotazione, fare clic sul pulsante “Tipo”.

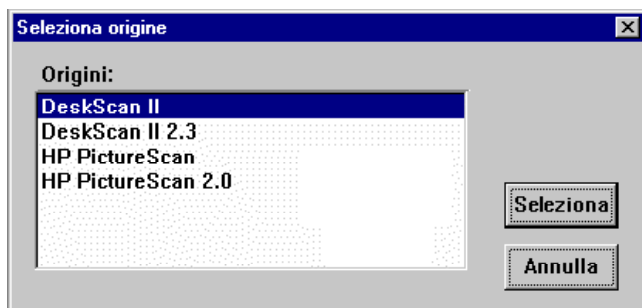


Quando si seleziona “<Image>” (immagine) come “scanner”, le immagini prescansionate funzionano sempre come fonte di immagini - non è neppure



necessario selezionare il disco come fonte di immagini sulla barra degli strumenti principale con il pulsante “Sorgente”.

Il pulsante “Configura” è disponibile solo quando lo scanner ne permette l'utilizzo. Tale pulsante consente l'accesso a parametri di scansione avanzati; con gli scanner Twain, facendo clic sul pulsante “Configura” è possibile selezionare l'origine Twain. (È possibile utilizzare anche il comando “Seleziona Origine” del menu “File”).



È possibile inoltre che lo scanner sia dotato di una **piattaforma di scansione**, quale Visioneer PaperPort o HP Document Assistant. In base al pacchetto software acquistato, è possibile disporre di manuali specializzati relativi alle piattaforme di scansione sul CD-ROM Readiris.

Una volta selezionato lo scanner, la stessa finestra consente di impostare la risoluzione di scansione, il formato e l'orientamento della pagina, la luminosità ed il contrasto, e può permettere d'indicare se sarà utilizzato il sistema di lettura automatica di documenti dello scanner o no. Con scanner Twain compatibili, tutti i parametri di scansione vengono spesso impostati all'interno dell'interfaccia utente.

Selezionare una **risoluzione** di 300 dpi per applicazioni normali, utilizzare una risoluzione superiore di 400 dpi per stampe di dimensioni ridotte (inferiori a 10 punti) e quando la qualità del documento è estremamente bassa.

Readiris legge **corpi del carattere** compresi fra 6 e 72 punti (da 0,21 a 2,54 cm).

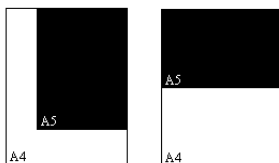
E più

72 punti

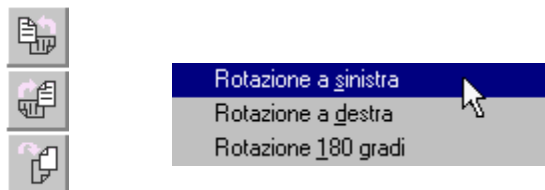
Readiris riconosce inoltre “**capolettere**”, ampie maiuscole che coprono più righe. (Naturalmente queste non possono essere di dimensioni superiori a 72 punti!)

Readiris legge
letterine (“drop
letters”) che
coprono diverse
linee e le indirizza alla
prima riga.

Abilitando l’opzione “Paesaggio”, la pagina selezionata sarà orientata in senso orizzontale, anziché verticale. L’orientamento della pagina viene applicato ai formati di pagina ridotti: con uno scanner da tavolo A4 è possibile, ad es., eseguire la scansione di pagine A5 (le cui dimensioni sono dimezzate) in formato verticale o orizzontale, ma ovviamente è possibile eseguire la scansione di una superficie piena A4 solo in una direzione!



Per regolare l’orientamento della pagina, è possibile utilizzare gli strumenti di **rotazione** della barra degli strumenti immagine. (I comandi corrispondenti si trovano nel menu “Visualizzazione”).



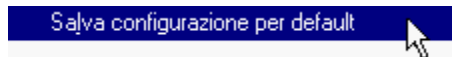
Sono disponibili tre sensi di rotazione: verso sinistra, verso destra e capovolta. La rotazione richiede alcuni secondi poichè è l'immagine stessa ad essere aggiornata e non solo la visualizzazione sullo schermo.

Impostare la **luminosità** e, se disponibile, il **contrasto**. L'impostazione del contrasto è disponibile solo per alcuni scanner. Trascinare la barra di scorrimento per modificare la luminosità o il contrasto del documento; i valori selezionati vengono rappresentati numericamente.

Finalmente, l'opzione "Invertire" permette di generare delle **immagini "invertite"** - attivare quest'opzione per analizzare delle pagine intere con testo bianco su sfondo nero.

SALVATAGGIO DI IMPOSTAZIONI DI DEFAULT

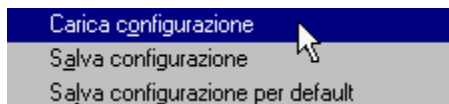
Impostare correttamente tutti i parametri di scansione e fare clic sul comando "Salva Configurazione per Default" del menu "File" per salvare le impostazioni correnti come impostazioni di default per impieghi successivi.



I file di impostazione non solo contengono le **impostazioni** dello scanner, ma consentono di determinare anche l'eventuale impiego dell'apprendimento interattivo, il dizionario di font utilizzato (e la relativa modalità), la lingua e il tipo di font utilizzati nel documento (ad es. un font normale, proporzionale), il modo di output e le opzioni di layout utilizzate - ad esempio il testo inviato a WordPad - ecc. In breve, *tutte* le impostazioni di funzionamento di Readiris vengono memorizzate nei file di impostazione.

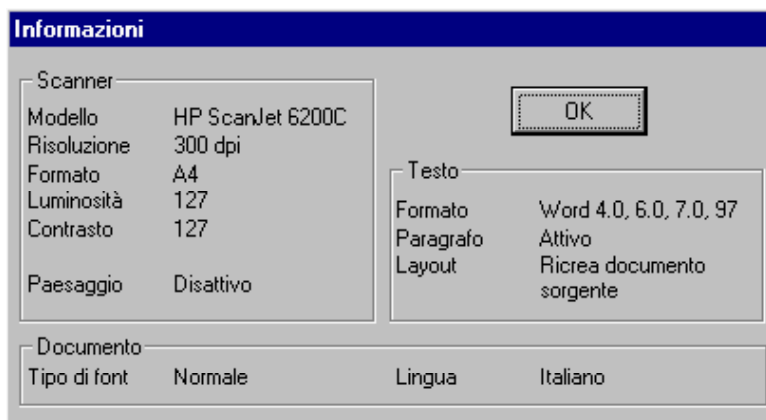
SALVATAGGIO DI IMPOSTAZIONI SPECIFICHE

Ad ogni avvio si utilizzeranno ovviamente le impostazioni di default, ma è possibile anche salvare impostazioni specifiche per non dover ridefinire i parametri di funzionamento. I comandi “Salva Configurazione” e “Carica Configurazione” del menu “File” svolgono tale funzione.



Segue un esempio: qualora si renda spesso necessario il riconoscimento ottico dei caratteri di documenti in lingua italiana che presentano un layout specifico, si consiglia di creare un file di impostazioni per questo tipo di documento. Selezionare quindi “italiano” come lingua del documento, caricare un modello di divisione in zone specifico per non dover rieseguire ogni volta la stessa divisione dello schermo, disabilitare l’apprendimento ma attivare un dizionario di font nella modalità “leggi”, in quanto si utilizzano sistematicamente gli stessi tipi di caratteri, ecc.

Qualora non si sia certi delle impostazioni correnti, per scoprire quali esse siano non è necessario “precipitarsi” in ogni menu e comando. Per averne una panoramica, è possibile utilizzare il comando “Informazioni” dal menu “File”.





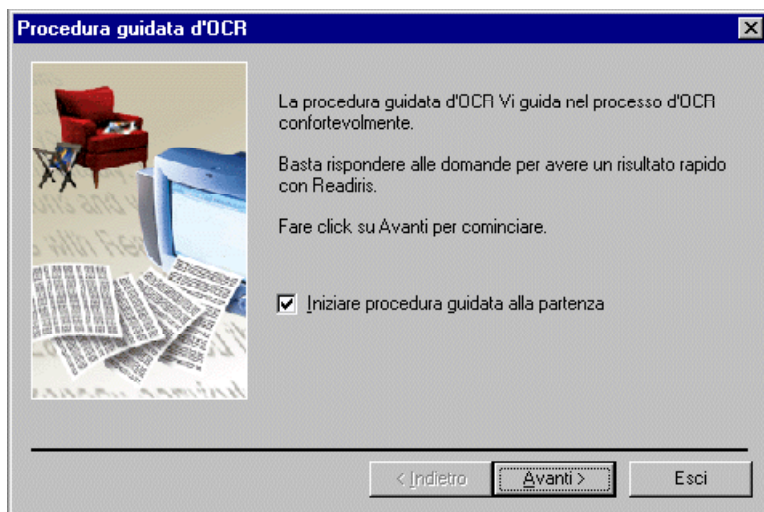
LA PROCEDURA GUIDATA FÀ TUTTO PER VOI!

Una volta installato lo scanner, è possibile iniziare a catturare i documenti. Invece di scorrere tutti i parametri, utilizzare la **procedura guidata d'OCR** (“wizard”), un modo decisamente agevole di riconoscimento delle pagine.

Cliccare il comando “Proc. GUIDATA” sulla barra degli strumenti principale (o selezionare il comando “Procedura GUIDATA” nel menu “Elaborazione”).



L'assistenza vi guida agevolmente attraverso la procedura OCR: basta rispondere a qualche domanda e si otterranno facilmente risultati rapidi con Readiris.



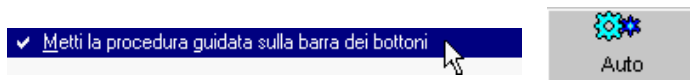
Si noti che la procedura guidata d'OCR inizia ogni volta che usate Readiris; potete evitarlo disattivando l'opzione “Inizia Procedura GUIDATA alla Partenza”

nel primo schermo dell'assistenza alla registrazione (e con l'opzione equivalente nel menu "Impostazioni").

RICONOSCIMENTO AUTOMATICO DELLE PAGINE

La procedura guidata d'OCR è un modo semiautomatico di riconoscimento delle pagine. Anche Readiris consente il riconoscimento automatico *completo*. Invece di scorrere passo dopo passo i parametri, riconosceremo immediatamente una pagina, senza interruzione.

Per sostituire il tasto "Proc. Guidata" sulla barra degli strumenti principale con il tasto "Auto", disattivare l'opzione "Metti Procedura Guidata sulla Barra dei Bottoni" nel menu "Impostazione".



Selezionare lo scanner come sorgente d'immagine utilizzando il tasto "Sorgente" e cliccare sul tasto "Auto" (oppure selezionare il comando "OCR Automatico" nel menu "Elaborazione").



L'**OCR automatico** consente di eseguire consecutivamente la scansione della pagina, di organizzarla in finestre mediante l'analisi della pagina o un modello di divisione in zone e di riconoscerla senza apprendimento interattivo. È sufficiente dare inizio alla scansione e salvare il testo riconosciuto, mentre Readiris si occupa dei passaggi intermedi.

READIRIS RICREA IL LAYOUT DEL DOCUMENTO

Non si deve confondere il riconoscimento automatico, che rende automatico il processo di riconoscimento, con l'autoformattazione! "Autoformattazione"



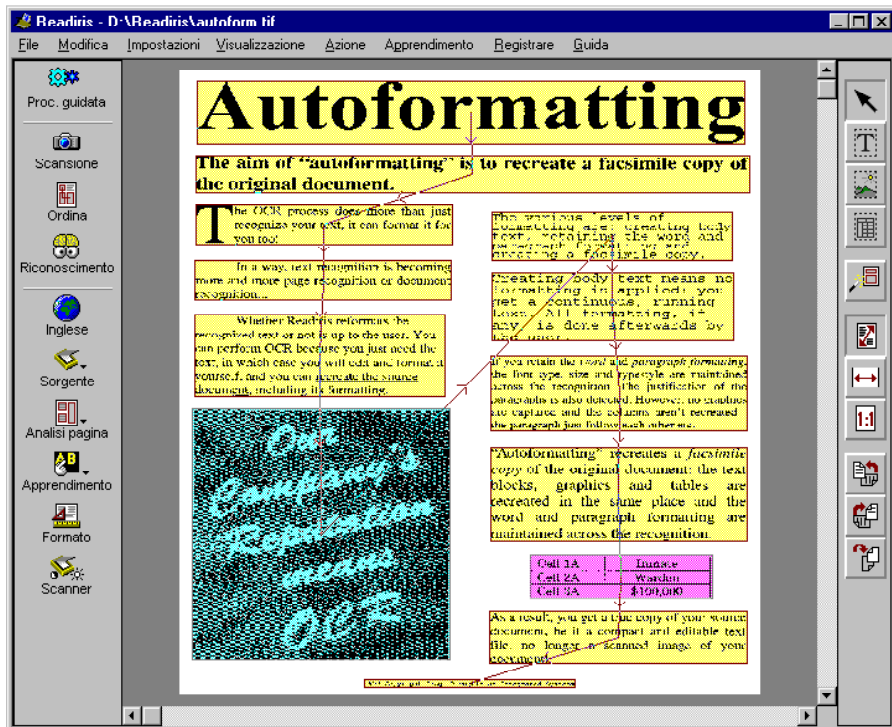
significa che Readiris ricrea una **copia facsimile** del documento scansionato, applicando la formattazione di parola, paragrafo e pagina del documento originale.

Durante il riconoscimento vengono utilizzati tipi di carattere (serif e sans serif, proporzionale e fisso, normale e condensato) simili a quelli del documento di partenza e vengono mantenuti i corpi e gli stili dei caratteri (grassetto, corsivo e sottolineato). Si ricreano le tabulazioni e l'allineamento (a sinistra, centrato, a destra e giustificato) di ciascun blocco di testo. Il posizionamento delle colonne, dei blocchi di testo e degli elementi di grafica è conforme al documento originale.

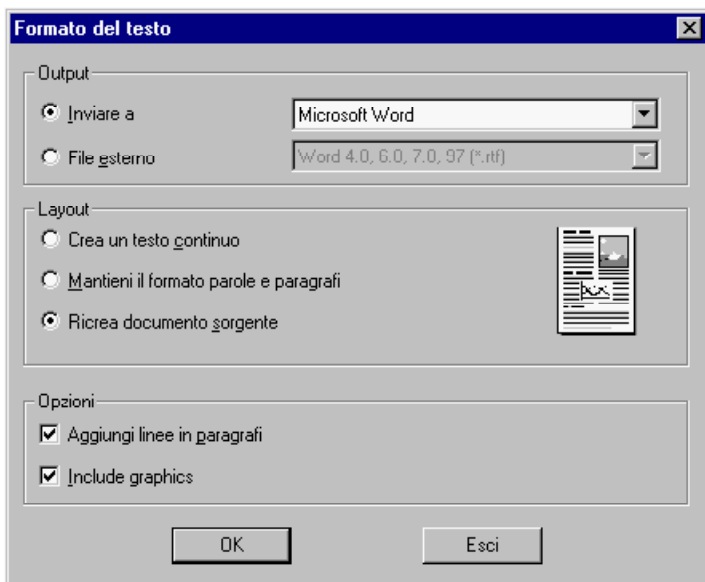
In altre parole, Readiris permette di archiviare una copia conforme di documenti, pur trattandosi ora di file di testo compatto ed editabile e non di un'immagine scansionata!

Pertanto, quando si utilizza "l'autoformattazione", *non* viene applicato l'ordinamento di finestre: è possibile includere ed escludere zone, ma qualsiasi riordinamento delle zone viene semplicemente ignorato!

Segue un esempio. Per comprendere meglio questa caratteristica, aprire l'immagine AUTOFORM.TIF posta nella cartella Readiris.



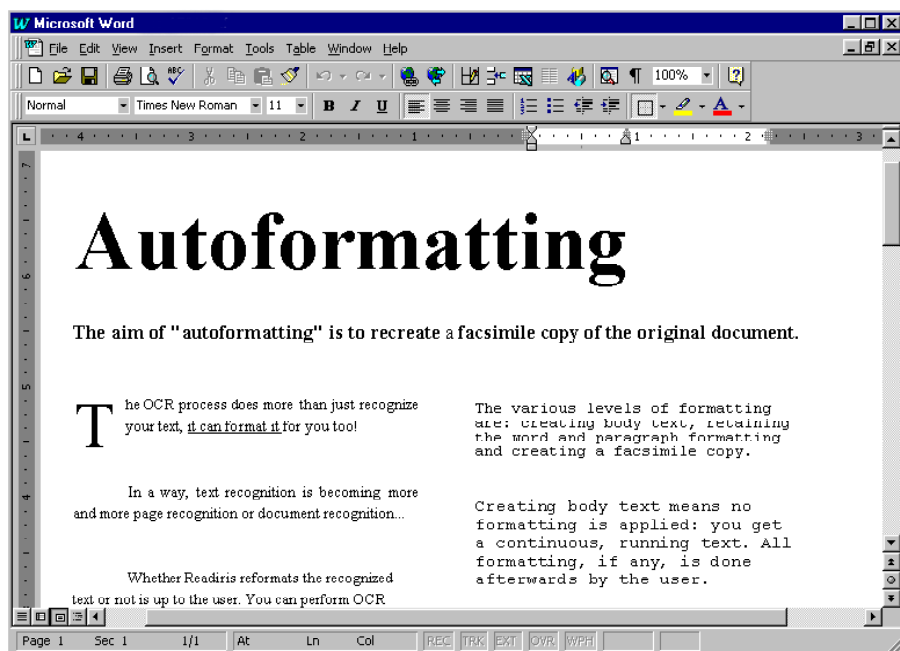
Cliccare sul pulsante "Formato" della barra degli strumenti principale e inviare il risultato OCR a Microsoft Word o selezionare "RTF" (Rich Text Format) o Word (DOC). Selezionare "Ricrea Documento Sorgente" come opzione di layout. (Viene attivata per default l'opzione "Unisce Linee in Paragrafi" per attivare il ritorno a capo all'interno dei paragrafi).



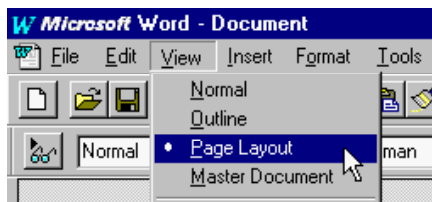
Si osservi che la ricostruzione del layout è limitata al formato Word (DOC), RTF, e ai documenti inviati direttamente a Microsoft Word. Inoltre, i formati Word (DOC) e RTF sono ampiamente usati come formati di testo leggibili con qualsiasi sistema di videoscrittura noto.

Altri formati “poveri” che producono testi “semplici”, quali ASCII, ANSI, ecc. *non* supportano codici di formattazione avanzati e quindi non sono in grado di offrire l’autoformattazione. WordPad è un editor di testi “ridotto”, non un programma di videoscrittura completo di tutte le caratteristiche, in quanto è in grado di aprire i file RTF, ma ignora la maggior parte degli elementi di formattazione quali blocchi di testo, allineamento, ecc.

Quando il testo riconosciuto viene aperto utilizzando un programma di videoscrittura, il testo appare in questo modo senza *alcun* intervento da parte dell’utente.



Per visionare correttamente l'effetto, è necessario abilitare il modo "WYSIWIG" del programma di videoscrittura, generalmente chiamato modo "layout di pagina". Tuttavia, inviando il documento riconosciuto direttamente a Microsoft Word 97, il "layout di pagina" viene attivato automaticamente.



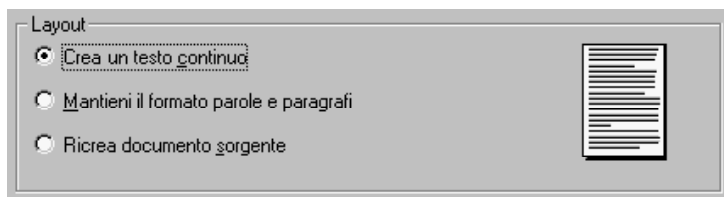


In breve, Readiris non solo riconosce i testi ma è anche in grado di formattarli per l'utente. OCR non è più solo riconoscimento di testo, ma significa anche sempre più **riconoscimento di pagina** o di documento!

FORMATTAZIONE DI TESTO, PARTE 2

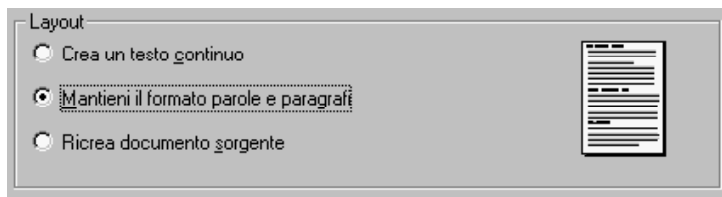
Le altre opzioni di layout sono “Crea un Testo Continuo” e “Mantieni il Formato parole e paragrafi”.

Come illustra l'icona sul lato destro, creare un **corpo del testo** significa produrre un testo non formattato, “continuo”. Il testo verrà catturato, ma ingorando completamente la sua formattazione. Utilizzare questa opzione quando si desidera ricattare solo un testo, ma non il suo layout.



(Generare del testo asiatico implica di creare il corpo di testo – non potete utilizzare “l’autoformattaggio” o mantenere le parole ed il formattaggio di paragrafo. Il numero di formati dei file è ridotto.)

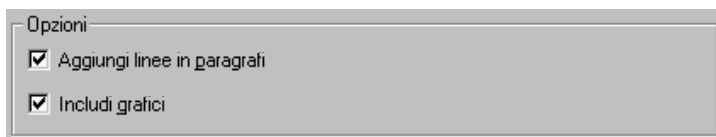
L’opzione “Mantieni il Formato Parole e Paragrafi” costituisce una via intermedia: durante il riconoscimento vengono mantenute la **formattazione della parola** - tipo di font (serif - sans serif, proporzionale - fisso, normale - condensato), corpo e stile del carattere (grassetto, corsivo e sottolineato), e la **formattazione del paragrafo** - tabulazioni e allineamento (a sinistra, centrato, a destra e giustificato).



Non confondere questa opzione di formattazione con l'autoformattazione "completa": questa opzione pone solo un paragrafo di seguito all'altro, non ricrea colonne, non copia la relativa posizione di varie zone di testo, né include gli elementi di grafica come avviene nell'autoformattazione completa.

SALVATAGGIO SEPARATO DEGLI ELEMENTI DI GRAFICA

Nell'esempio riportato, gli elementi di grafica erano inclusi nel testo riconosciuto; questo dipende dall'opzione di formattazione "Includi Grafica". Il salvataggio dei grafici all'interno del testo è possibile solo con l'autoformattazione "completa", e non con formati di testo "poveri" quali ASCII, ANSI, ecc.



Con Readiris è possibile salvare elementi di grafica senza eseguire il riconoscimento testo. Dal momento che il software di OCR Readiris crea immagini in **bianco e nero** e non immagini a colori o con scale di grigio, non ha senso scansionare fotografie, ma è possibile catturare senza problemi elementi di grafica al tratto.

Come? Tracciare una zona di grafica intorno alle illustrazioni, ai disegni, ecc. Selezionare lo strumento "Finestra di Grafica" per creare manualmente finestre di grafica e per disegnare finestre di testo e di tabelle.

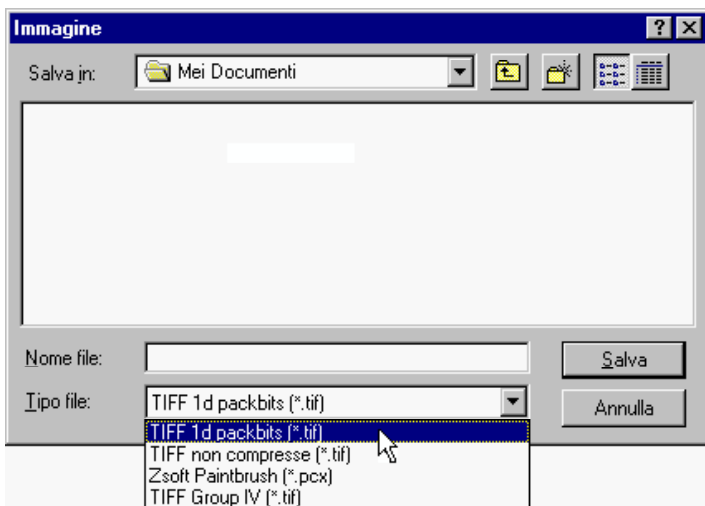


(Analogamente ad altri tipi di finestra, la casella strumenti della finestra di grafica vi dice il numero delle finestre di grafica esistenti. Selezionando una finestra, utilizzare il tasto Alt per visualizzarne il tipo e le coordinate.)

Quindi, scegliere il comando “Salva Grafici” del menu “File”.



Il sistema chiederà di specificare un nome di file.



Stabilire il formato di file di grafica che si intende utilizzare. Selezionare un formato supportato dal software di ritocco immagine o fotografia. Sono disponibili i formati TIFF e Paintbrush (PCX). (È sempre possibile aprire i file Paintbrush con l'accessorio Paint di Windows.)

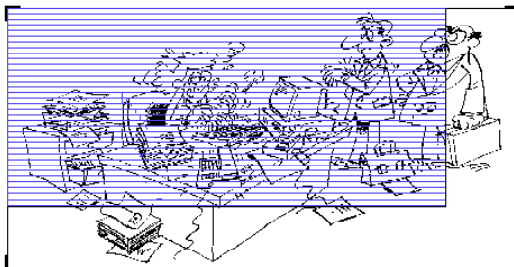


Gli elementi di grafica vengono salvati in un unico file. Non è necessario limitarsi ad un'unica grafica, ma nel caso in cui vengano tracciate numerose finestre di grafica, queste verranno raccolte, "immagazzinate" in un unico file. (Per separarle è possibile utilizzare il comando "Ritaglia" del programma di ritocco immagine o fotografia.)

Non sono consentiti lati inferiori ad 1 mm - bitmap di tali dimensioni difficilmente contengono informazioni. Finestre "irregolari" non rettangolari vengono convertite in rettangoli in modo non visibile: Readiris copre la superficie massima basata sulle coordinate maggiori. Nell'esempio sotto riportato, la finestra di grafica con il bordo nero è completata da un'area oscurata.



ur computer automatically.



nt is read by your scanner. This device acts as the "eye" of yo

LETTURA DI FAX E RICONOSCIMENTO POSTICIPATO

Salvare immagini come file immagine offre un'altra possibilità: è possibile salvare una pagina *intera* e successivamente eseguire l'**OCR posticipato**, come si è fatto con le immagini prescansionate delle esercitazioni.

Scansionare semplicemente il documento e selezionare il comando "Salva Pagina Completa come Immagine" nel menu "File". Vi verrà chiesto di salvare l'intera pagina come file TIFF o Paintbrush (PCX).

Salva pagina completa come immagine...



A questo punto è possibile selezionare il disco come fonte di immagini ed aprire il file di immagini con il tasto "Apri" (o con il comando corrispondente nel menu "Elaborazione"). (Qualora si utilizzi il comando "Apri" del menu "File", non è neppure necessario aggiornare la sorgente di immagini.)

Readiris apre immagini TIFF in bianco e nero (non compresse, pacchetti di bit, compresse di Gruppo 3 e Gruppo 4), immagini Paintbrush (PCX) e bitmap di Windows (BMP). Questa proprietà è particolarmente utile per convertire i **fax** in file di testo editabili!

Se possibile, chiedere ai propri corrispondenti di inviare fax di “buona” qualità - fax con una risoluzione superiore di 200 dpi per ottenere migliori risultati di OCR.

Per aprire immagini, non dimenticare che è possibile utilizzare la funzione **“trascinare selezione”**: dopo aver trascinato un file di immagini da Esplora Risorse (Explorer) di Windows er alla zona di immagine di Readiris, è immediatamente possibile aprire tale file!

RICONOSCIMENTO DI TABELLE

Fino ad ora sono stati spiegati il riconoscimento di testi e di fax e il salvataggio degli elementi di grafica. Si passerà ora all’elaborazione di una tabella. Prendere una tabella numerica ed eseguirne la scansione oppure aprire l’immagine campione TABLE.TIF nella cartella di Readiris.

L’immagine TABLE.TIF contiene in realtà due tabelle e non a caso! L’analisi della pagina divide in zone l’intera tabella e Readiris è in grado di ricostruirla, ricreando le tabelle cella dopo cella nel foglio elettronico o inserendo un oggetto tabella all’interno del file del vostro sistema di videoscrittura.

Analizziamo le diverse soluzioni iniziando dalla tabella “con griglia” o “incorniciata”. È caratterizzata da bordi intorno alle celle.



Readiris - D:\Readiris\Table.tif

File Modifica Impostazioni Visualizzazione Azione Apprendimento Registria Guida

Proc. guidata
Scansione
Ordina
Riconoscimento
Inglese
Sorgente
Analisi pagina
Apprendimento
Formato
Scanner

Reading Tables

Readiris recognizes tabular data and recreates them cell by cell in worksheets or as table objects inside wordprocessor files.

To insert tables as table objects, you must retain the word and paragraph formatting or recreate the source document: see the "Format" button on the main toolbar.

The page analysis detects "gridded" and "ungridded" tables. "Gridded" or "framed" tables have borders around the cells - as does the example below. The borders of the table cells get recreated.

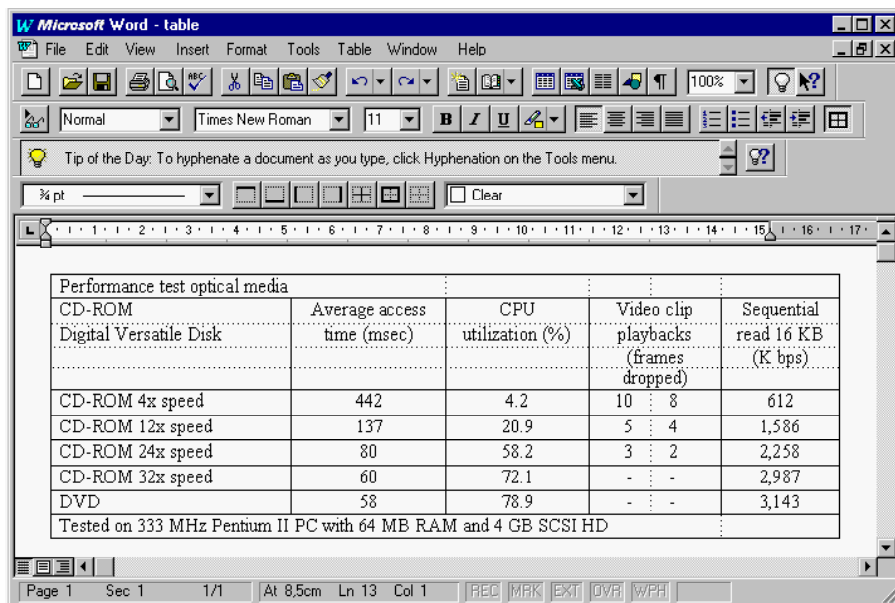
Performance test optical media				
CD-ROM Digital Versatile Disk	Average access time (msec)	CPU utilization (%)	Video clip playbacks (frames dropped)	Sequential read 16 KB (K bps)
CD-ROM 1x speed	442	4.2	10 8	612
CD-ROM 12x speed	197	20.9	5 4	1,586
CD-ROM 24x speed	80	98.2	3 2	2,208
CD-ROM 32x speed	60	72.1	- -	2,867
DVD	58	78.9	- -	3,143

Tested on 333 MHz Pentium II PC with 64 MB RAM and 4 GB SCSI HD

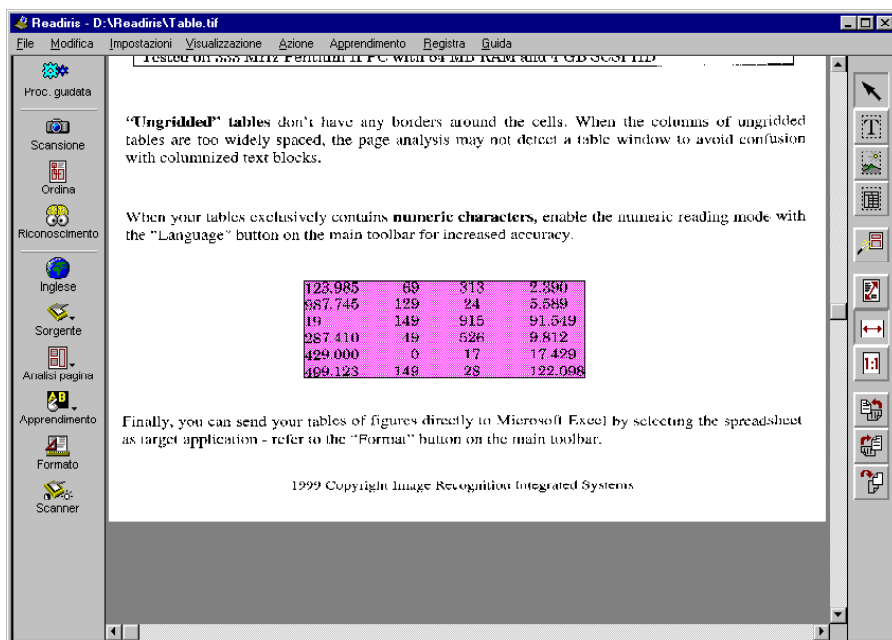
"Ungridded" tables don't have any borders around the cells. When the columns of ungridded tables are too widely spaced, the page analysis may not detect a table window to avoid confusion with columnized text blocks.

When your tables exclusively contains **numeric characters**, enable the numeric reading mode with the "Language" button on the main toolbar for increased accuracy.

Eseguire il riconoscimento dopo aver abilitato l'opzione di layout, "Mantieni il Formato Parole e Paragrafi" o "Ricrea Documento Sorgente" e le tabelle vengono ricreate. Aprire il sistema di videoscrittura (se la tabella riconosciuta non è stata inviata direttamente a Word) per esaminare il risultato. (È anche possibile che siano stati inclusi paragrafi di testo nel file di testo).



Passiamo ora all'esempio "senza griglia". Non vi sono bordi intorno alle celle. L'analisi della pagina individua comunque la tabella.

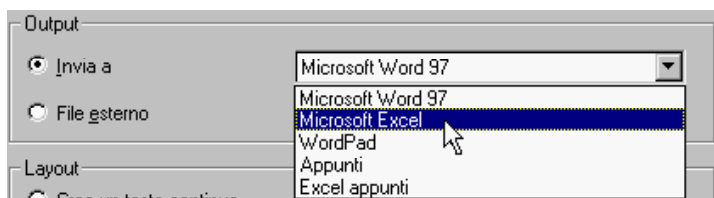


Per una precisione ottimale di OCR, limitare il riconoscimento ai **simboli numerici** mediante il pulsante "Lingua". (La modalità numerica non riguarda strettamente i numeri, ma include i simboli da 0 a 9, +, *, /, %, , (virgola), . (punto), (,), -, =, \$ e £.)

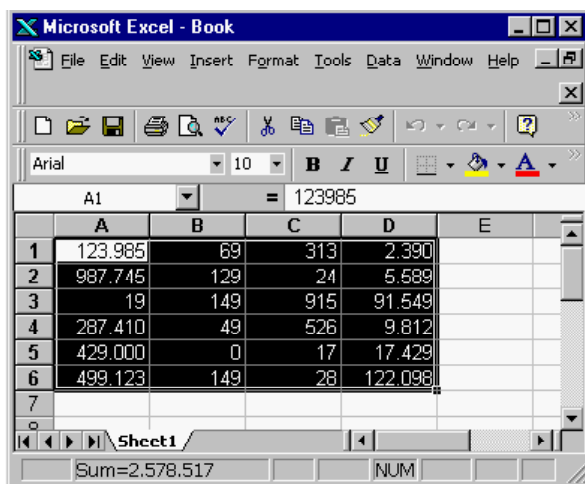


Poiché questo procedimento è possibile solo quando la tabella non contiene simboli alfabetici - altrimenti non sarebbe possibile riconoscere correttamente le porzioni di testo -, ora è possibile attivare la modalità numerica, ma non per la prima tabella.

In questo caso, invieremo il risultato OCR direttamente al foglio elettronico Microsoft Excel. Selezionare quindi Excel come applicazione target con il tasto “Formato”.



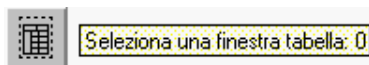
Il foglio elettronico viene avviato automaticamente con questo risultato: viene ricreata la tipica struttura a tabella con righe e colonne. È possibile ora passare all'elaborazione dei dati.



È possibile anche inviare le tabelle agli appunti (opzione “Invia ad Appunti Excel”) e salvarle in un file esterno (formato file Excel - CSV indica dati “separati da virgola”).



Potreste avere tabelle “senza griglia” non individuate dall’analisi della pagina come zona tabella perché le colonne sono troppo spaziate. Readiris cerca di evitare confusione con i blocchi di testo organizzati in colonne. Per creare manualmente una finestra di tabella, cliccare “Finestra di Tabella” nella barra strumenti immagine, e procedere come d’abitudine; la casella strumenti indica di nuovo il numero di finestre di tabelle.



LA FUNZIONE “CONNECT”, UN ALTRO COLLEGAMENTO DIRETTO

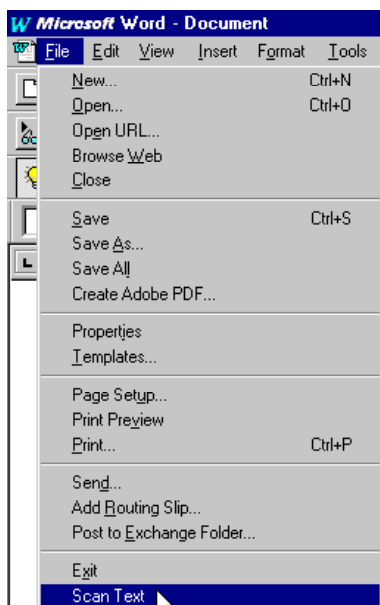
Analogamente alla funzione “Invia a” che esporta automaticamente il risultato di lettura, la funzione “Connect” consente un collegamento OCR diretto fra lo scanner e le applicazioni Windows. La sola differenza consiste nel modo “inverso” di elaborazione: i documenti non vengono riconosciuti all’interno di Readiris per esportare il risultato a un’applicazione target; Readiris viene richiamato dall’interno dell’applicazione.

Per alcune operazioni manuali nessuna procedura di **installazione** può sostituirsi all’utente, pertanto, per informazioni relative all’installazione consultare il sistema di guida in linea o il file CONNREAD.DOC nella cartella di Readiris - per utilizzare la capacità “Connect” da Word, Excel e WordPerfect è necessaria una macro. (Il file “Leggimi” (“Read Me”) viene visualizzato automaticamente al termine del programma di installazione.)



Se installata correttamente, la caratteristica "Connect" aggiunge un nuovo comando al menu delle applicazioni. Negli esempi sotto riportati, il comando "Scansiona Testo" viene aggiunto al menu "File" di Word e il menu "Scansiona Tabella" viene aggiunto alla barra del menu di Excel.





Segue il funzionamento: quando si lancia la macro “Connect”, Readiris *non* può girare sullo sfondo. Per far girare Readiris fare clic sul nuovo comando del menu dell’applicazione. Prima che Readiris inizi a girare, è possibile che l’interfaccia “Connect” ponga alcune domande. Procedere normalmente e terminata l’operazione uscire dall’interfaccia “Connect”. I risultati del riconoscimento vengono inseriti nel punto in cui si trova il cursore.

USO DELLA GUIDA IN LINEA

Qui termina la panoramica di Readiris. Non è possibile includere nel presente manuale alcune informazioni dell’ultimo minuto. Quindi, per ulteriori informazioni su Readiris si consiglia di consultare il sistema di guida in linea.

Per fare ciò, passare al menu “Guida”. Il comando “Argomenti della Guida” e il relativo tasto di scelta rapida F1 consentono di navigare attraverso i numerosi argomenti della guida.

