



OmniPage Pro 10 OmniWeb

"Io ho visto cose che voi umani nemmeno immaginate", dice Roy, il replicante, prima di morire, (veramente, la stessa cosa mi disse mio suocero, Dio lo abbia in gloria, guardandomi dritto negli occhi, quando andai a chiedere la mano dell'ultima sua rampolla, variando leggermente il soggetto - Figlio mio, vedrai... -). E noi, di cose, in questi ultimi dieci anni, ne abbiamo viste tante, ma tante. Forse non ce ne rendiamo conto, ma quindici anni fa il DOS ci costringeva a tenere a memoria una serie di comandi da battere alla tastiera, la

grafica era, più o meno, una serie di punti grossi come mattonelle che, con un po' d'immaginazione, lasciavano intravedere le sembianze di un oggetto, e pensare a una macchina che scriveva sotto dettatura era come immaginare che gli asini volassero, che il governo togliesse le tasse, o che Naomi si fosse trasferita ad Atripalda.

E invece, dopo tre lustri, eccoci a considerare quante cose sono accadute, quante meraviglie abbiamo visto senza vederle, susseguentisi l'una dopo l'altra a ritmo sempre più incalzante. Te-

lefonni cellulari, modem, scanner, videoconferenza, monitor grandi come televisori e sottili come un libro, input vocale, WWW. E OCR, tanto per concludere in bellezza.

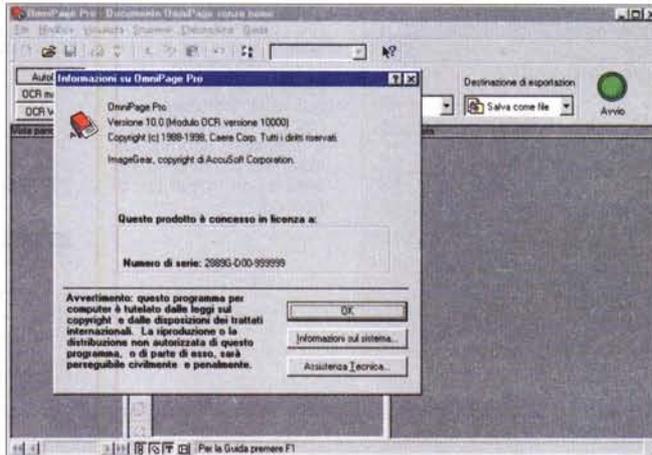
Già, l'OCR, che oggi sta cadendo un po' di interesse; i motivi sono molteplici ma si riducono, poi, a uno singolo, la sempre minore presenza dei documenti in formato cartaceo. Venticinque anni fa scrivevo gli articoli (allora per un'altra rivista, di armi) a mano, e passavo il palinsesto a un vecchio professore in pensione che mi batteva a macchina il pez-

OmniPage Pro 10

Produttore:
 Caere Corporation
 100 Cooper Court
 Los Gatos CA 95032 USA
 http://www.caere.com

Distribuito in Italia da:
 CDC Tel. 0587 2882
 Jensoft Tel. 0532 786911
 Ingram Micro Tel. 02 55351
 Opendate Tel. 0332 803111

Prezzi (IVA esclusa):
 Versione Full L. 1.200.000
 Upgrade da altro OCR L. 299.000
 Aggiornamento L. 189.000



Lo splashscreen di OmniPage Pro.

nale se ne va a farsi benedire!

L'unica è di affidarsi a uno scanner e a un OCR; e se questo secondo è un prodotto dell'esperienza di OmniPage, i risultati sono assicurati.

zo e me lo ritornava. Ovviamente, fiorendo i termini tecnici (scrivevo di ricarica di armi rigate), gli errori erano tanti (anche a causa della mia grafia infernale); quindi il dattiloscritto, di nuovo corretto, ritornava dal buon professore che riscriveva il tutto. A questo punto il materiale poteva essere spedito. Il tutto, se andava ogni cosa per in verso giusto, in una settimana. Oggi adottando un ricognitore vocale e spendendo via Internet si riesce, se si conosce bene l'argomento, a chiudere in un'ora di lavoro.

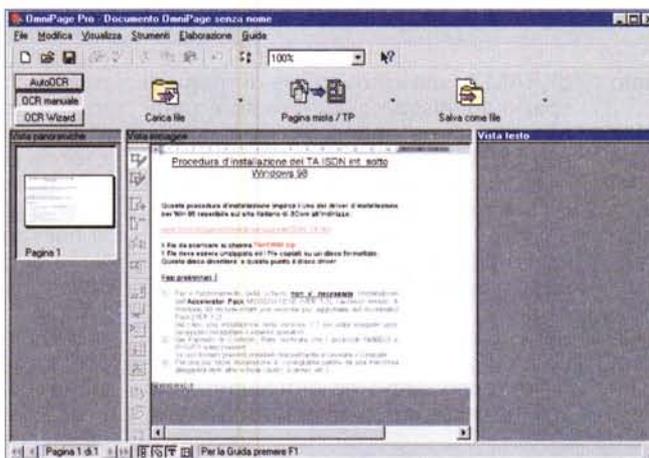
La maggior parte dei documenti e degli scritti, oggi, è in forma elettronica; oggi non si dice più "Ti mando il fascicolo", ma "Ti mando il file", e la fatica risparmiata è tanta, da tutte e due le parti. Ma "la maggior parte", in lingua italiana, non significa tutto. E, quando si è di fronte a un

fascicolo bello voluminoso, o a un pacco di fogli dattiloscritti, iniziano i problemi; certo, c'è la possibilità di leggere a un ricognitore vocale ma ve l'immaginate, che pa...zienza, dettare, come un baccalà, a un microfono per qualche ora. Senza poi tenere da conto che tutta la formattazione del documento origi-

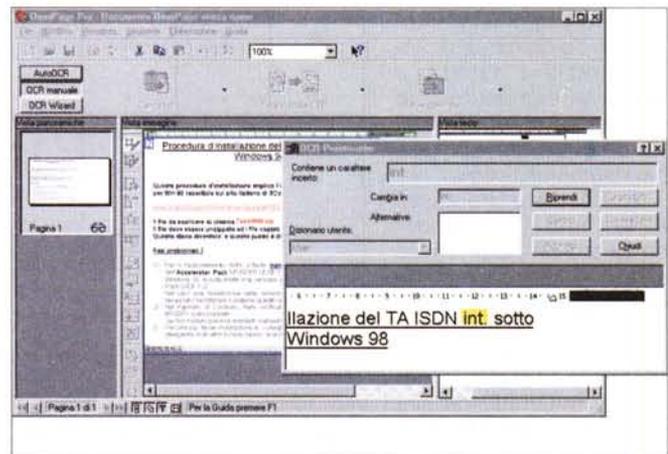
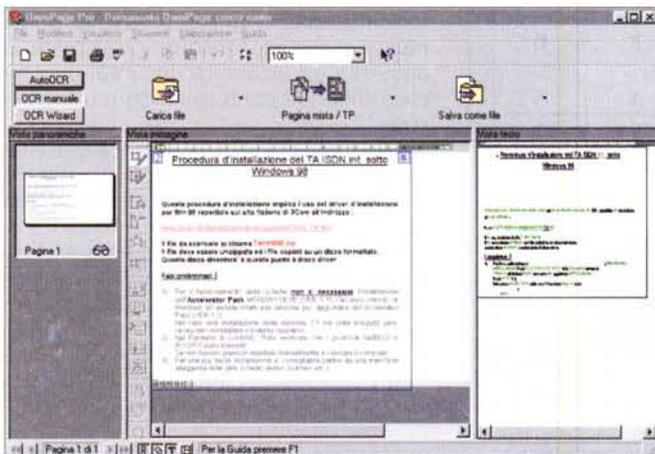
OmniPage e OmniWeb, non solo per trascrivere

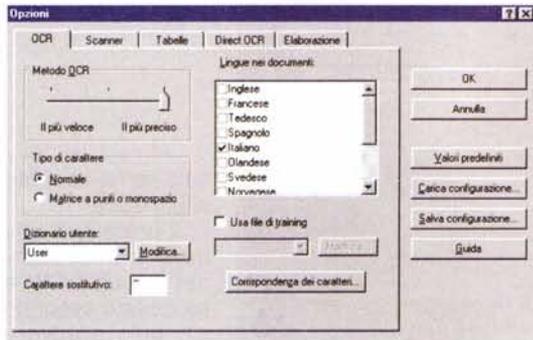
Dei ricognitori di carattere si cominciò a parlare una dozzina di anni or sono, in coincidenza con la comparsa, sul mercato, di scanner a prezzo più basso o, ancora, dei cosiddetti handheld, quella specie di spatole da stuccatori che si strusciavano sulle pagine, e che ci davano una serie di pezzi da assemblare alla men peggio (al modico prezzo di diversi biglietti da centomila, contro il centone che costa oggi uno scanner piano di accettabile qualità).

Ricordo allora di aver parlato, nell'allora esistente rubrica di Intelligenza Artificiale, proprio di questo argomento e delle difficoltà connesse con la tecnica di riconoscimento dei caratteri.

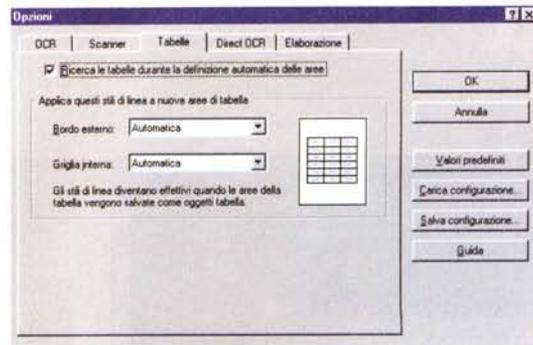


Le diverse fasi del riconoscimento.





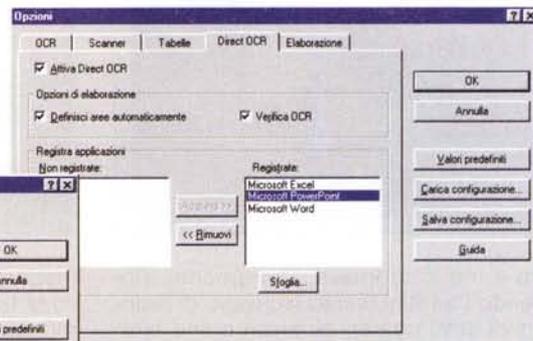
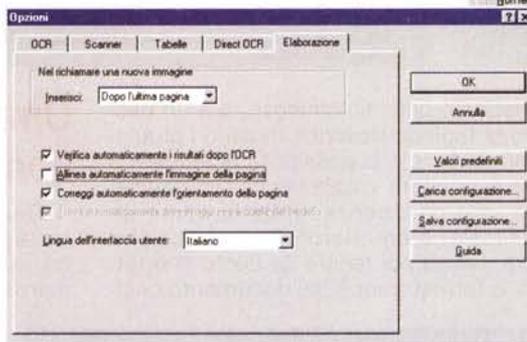
Alcune delle opzioni disponibili; notare le possibilità di registrare programmi come preferenziali, e di riconoscere automaticamente tabelle. E' possibile, inoltre, selezionare più linguaggi e creare un modello di training.



Nello stesso periodo acquistai, per il Mac, un OCR spagnolo (non ne ricordo più neppure il nome) tanto difficile da maneggiare quanto incerto ed esitante nei risultati (figurarsi che occorreva indicare, prima di iniziare, anche il tipo di font adottato nella redazione dell'originale). Ma Caere entrò alla grande sul mercato, fin dall'inizio della comparsa di questa tecnologia, divenendo quasi subito, grazie alla preziosità dei risultati fin da allora forniti, il punto di riferimento di questo genere di prodotto.

Da quella preistorica versione 1 è passato quasi un secolo informatico, e la versione 10, coincide con la fine del millennio, ha aggiunto un altro tassello al gran progetto, non ancora completato, del riconoscimento dei caratteri. Grazie a un'oculata politica di conservazione dell'ambiente di base (alcune versioni sono state pressoché simili nelle funzionalità affacciate verso l'utente, anche se l'avanzamento tecnologico interno non ha mai sofferto di crisi di progresso) che non ha mai subito salti tali da lasciare disorientato l'utente non professionista, OmniPage ha continuato, anche grazie a questo, a mantenere la leadership di quest'ambiente di lavoro, offrendo sempre più risultati di pregio.

Nonostante la potenza e la complessità del software (in parte dimostrata anche dalla grandezza del pacchetto principale) le richieste, in termini di macchina e di memoria, non sono penalizzanti; OmniPage gira su un qualsiasi Pentium, con un minimo di 32 MB



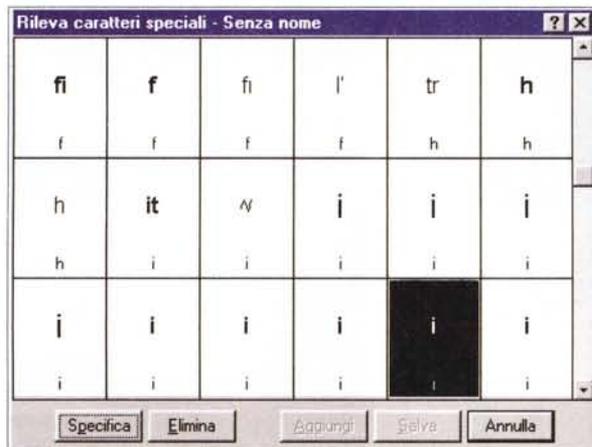
di RAM e una cinquantina di mega di spazio sul disco rigido. Il resto è routine, da una semplice SVGA a, ovviamente, un CD-ROM.

E, ovviamente, uno scanner; sono praticamente supportati tutti quelli presenti sul mercato, TWAIN compliant, vale a dire Agfa, Adara, Artec, Brother, Caere, Epson Fujitsu, LaCie, Microtek, Mustek, Plustek, Ricoh, ScanPlag, Sony, Storm, Umax, Visioneer e, manco a dirlo, Hewlett-Packard. Le funzionalità d'uso, abbiamo detto, non sono variate rispetto a quanto avveniva nelle

versioni precedenti (di cui MC si è occupato in varie riprese negli anni passati), per cui non spenderemo più di poche parole per illustrarle: importazione di un'immagine di un documento in OmniPage Pro, definizione (automatica, manuale o mista) delle parti del documento che si desidera riconoscere come testo, interpretazione del testo stesso, ed esportazione verso un'applicazione preferita (che oggi non significa solo un wp, ma anche, ad esempio, un messaggio di posta elettronica). Anche la finestra di base è la stessa, con una barra di gestione strumenti, e tre finestre di visualizzazione dei documenti, di amministrazione delle aree e di interpretazione vera e propria. Come sempre, è possibile affidare l'intera operazione a un wizard o intervenire personalmente nelle varie fasi. In ambedue i casi, comunque, è lasciata all'utente la scelta di certi parametri, come formato di uscita (livello di mantenimento della formattazione originale presente, destinazione di esportazione, possibilità di acquisire ed elaborare documenti in applicazioni non conformi al formato originale).

Sempre utile è la possibilità di ruotare l'immagine di passi di 90° (la cosa è indifferente ai fini dell'elaborazione, ma questo vale solo per la macchina) o di raddrizzare la pagina se questa è stata posta nello scanner in maniera non precisa (si può lasciare al programma l'incarico di raddrizzare l'immagine in maniera automatica; in questo caso OmniPage può anche decidere di lasciarla così come è se la cosa è ininfluente sul riconoscimento).

OmniPage tenta, per quel che è



L'area di training, forse un tantino noiosa da mettere a punto, ma certamente utile per migliorare efficacia e rapidità di riconoscimento.



possibile, di mantenere corpo e tipo dei caratteri utilizzati nell'originale, e la finestra di verifica e di correzione ortografica è divenuta più agile, e paragonabile a quella di un buon wp. Buona e utile è la possibilità di riordinare o cancellare pagine nel caso di una scansione multipla, e numerosi sono i formati di possibile salvataggio del documento interpretato. Interessante è anche il comando "Invia come posta" che trasforma il prodotto finale in un file attached che può essere unico per tutte le pagine, o frazionato in tanti documenti agganganciati poi a un messaggio di posta elettronica (viene comunque gestita da OmniPage tutta la procedura di creazione del messaggio e, se lo si desidera, di spedizione).

OmniPage Web, un occhio su WWW

Finora, per quanto sofisticato e potente (in condizioni ideali viene assicurato il 99% di resa finale) OmniPage Pro non è una novità.

Lo è invece OmniPage Web, giunto alla versione 1.5, che potremmo definire la versione verticale del Pro, dedicata al Web. Cosa fa è presto detto; la sezione OCR del prodotto trasforma accuratamente un documento stampato (con le sue immagini) in un testo editabile; una volta conclusa quest'operazione, OmniWeb analizza la struttura del documento e lo trasforma in un completo sito Web, dinamico, con pagine separate per ogni capitolo o sezione originaria.

E, per buona giunta, esso crea link di

ipertesto, tool di navigazione, e tabelle "iperlinkate" di contenuti.

La cosa sembra facile e, caso strano, lo è; per creare una pagina Web, usando un documento scritto e OmniPage Web, è sufficiente seguire sei passi fondamentali:

- ✓ trasformazione di un documento cartaceo in un file OP Web;
- ✓ creazione di "zone" da riconoscere, alternativamente e separatamente, come testo e come grafica;
- ✓ applicazione delle tecniche di OCR (fin qui lo sviluppo è lo stesso di OP Pro);
- ✓ outlining della struttura del documento; OP Web esamina il file e crea un layout della struttura del documento originale, basandosi sulla suddivisione in zone effettuata al momento della ricognizione;
- ✓ selezione di parti od oggetti da utilizzare come componentistica HTML. In questa fase, oltre a creare gli opportuni link e ancore, a determinare stringhe "sensibili" e altro ancora, si potrà inserire, nel testo, codice HTML previsto dall'utente per scopi particolari;
- ✓ la fase finale è, infine, rappresentata

OmniPage Web in azione; facile e rapido il wizard di analisi e composizione, soddisfacenti i risultati.

dal salvataggio del prodotto in codice HTML, e dal successivo test (il pacchetto prevede anche una scorciatoia diretta del tipo "Save and Launch...").

Come dicevamo poco prima, quindi, la cosa è abbastanza facile, anche se non è che si possano sviluppare, in questo modo, prodotti di grande sofisticazione. Ma il principio adottato è valido; chi ha polso fermo nella costruzione di pagine WWW non utilizza certo prodotti come OP Web.

Chi invece ha poca dimestichezza con gli editor HTML troverà molto pratico adottare questo sistema, che comunque garantisce pur sempre risultati soddisfacenti, specie se si considera il minimo sforzo applicato.

Conclusioni

Beh, da un pacchetto che è giunto alla versione 10, non ci si poteva aspettare di più e di meno. La spiegazione del paradosso sta nel fatto che, dopo tante versioni, tutto quello che c'era da scoprire, per dirla in parole banali, è stato fatto; ma c'è ancora tutto il tempo e il modo per raffinare sempre di più il buon prodotto già esistente. Salvo poi trovare, nell'evoluzione delle tecnologie, una nuova palestra di sviluppo (parliamo di OmniPage Web) che, per essere all'inizio, avrà tempo e modo di far valere la maestria dello staff che ha portato a far diventare indiscusso leader il prodotto fratello. *MG*