

Optical Character Recognition

Nuovi tool di editing per FineReader Professional

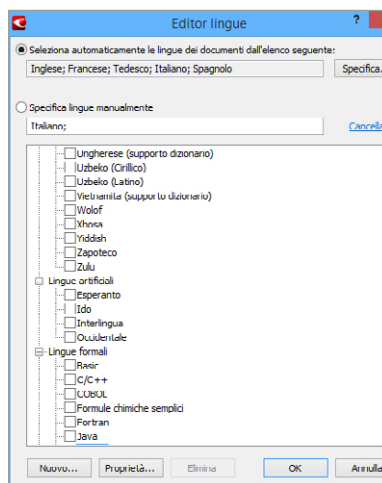
L'interfaccia migliorata e i nuovi strumenti di correzione rendono ancora più efficace l'ottimo software di Ocr prodotto da Abbyy.

■ Anteprima di **Maurizio Bergami**

Ci sono voluti due anni e mezzo di attesa – FineReader 11 è stato lanciato a settembre del 2011 – ma a marzo Abbyy finalmente ha aggiornato il suo noto software di Ocr (Optical Character Recognition, riconoscimento ottico dei caratteri). Anche la versione 11, recensita sul numero 249 di PC Professionale (potete trovare il Pdf dell'articolo sul Dvd virtuale di questo numero) aveva avuto un periodo di sviluppo assai lungo, e le novità erano state relativamente limitate: ricordiamo, in particolare, una modalità di elaborazione “bianco e nero” che riduceva il tempo di riconoscimento dei documenti per i quali il colore non era necessario, e la possibilità di salvataggio diretto in formato ePub. Il motore di scansione era risultato leggermente più preciso di quello di FineReader 10, ma aveva anche fatto un clamoroso passo indietro nel riconoscimento del testo verticale nelle tabelle. Da molti anni il nostro set di documenti di prova per gli Ocr comprende una tabella complessa, creata ad hoc, le cui celle di testo verticale avevano mandato in crisi FineReader 11. La release precedente, al contrario, era stata in grado di riconoscere l'intera tabella alla perfezione. Le difficoltà

sembravano derivare dall'incapacità di gestire in modo corretto la struttura della tabella: usando lo strumento Unione Celle per definire a mano le aree con testo verticale eravamo riusciti ad aggirare l'ostacolo, ma questo approccio evidentemente è impraticabile per chi elabora abitualmente molti documenti di questo tipo. Abbyy aveva riconosciuto il problema, assieme ad alcuni altri bug che avevamo segnalato al termine della prova (poi risolti con un update rilasciato pochi mesi dopo), limitandosi però a segnalarci che “[il comportamento con questa tabella] ... sembra un'eccezione, ma i nostri sviluppatori sono al lavoro per migliorare questi risultati e qualunque miglioramento verrà incluso il prima possibile nelle nuove versioni”. Inutile dire che appena ricevuto FineReader 12 abbiamo immediatamente fatto una verifica, ma le nostre speranze sono andate deluse: il comportamento non è cambiato, e per ottenere un riconoscimento corretto abbiamo dovuto ancora unire a mano le celle che il software aveva identificato erroneamente come composte da elementi separati, e indicare sempre a mano l'orientamento del testo. Abbiamo effettuato varie

prove con altre tabelle per approfondire l'argomento, riscontrando che in effetti il nostro documento è particolarmente ostico per FineReader ma non è il solo a creargli difficoltà: quando le tabelle contengono etichette disposte in verticale spesso il software fornisce un risultato perfetto, ma a volte – e non siamo riusciti ad identificarne il motivo – entra in crisi. In compenso abbiamo notato con soddisfazione che FineReader 12 risolve un problema non gravissimo ma fastidioso e di lunga data: un eccesso di zelo nel riconoscimento delle lettere maiuscole con l'accento simulato tramite l'apostrofo. Uno dei nostri campioni di prova – un articolo di quotidiano – contiene il testo PIU': in passato il software riconosceva l'accento e usava il carattere corretto (Û) ma senza rimuovere l'apostrofo, inserendo quindi nel documento finale PIÛ'. Ora invece FineReader si limita a mantenere l'apostrofo, senza effettuare alcuna sostituzione: i puristi probabilmente preferirebbero un'altra soluzione, ma almeno non è più necessaria la correzione manuale. Con gli altri documenti del nostro set di prova FineReader 12 ha evidenziato progressi in assoluto lievi ma senza dubbio



FineReader riconosce 190 idiomi; l'elenco comprende anche lingue artificiali – ad esempio l'esperanto – e formali (tra cui vari linguaggi di programmazione).

FineReader 12 Professional

VOTO 8,0

Euro **129** Iva inclusa

L'aggiornamento dalla versione precedente costa **89** euro.

PRO

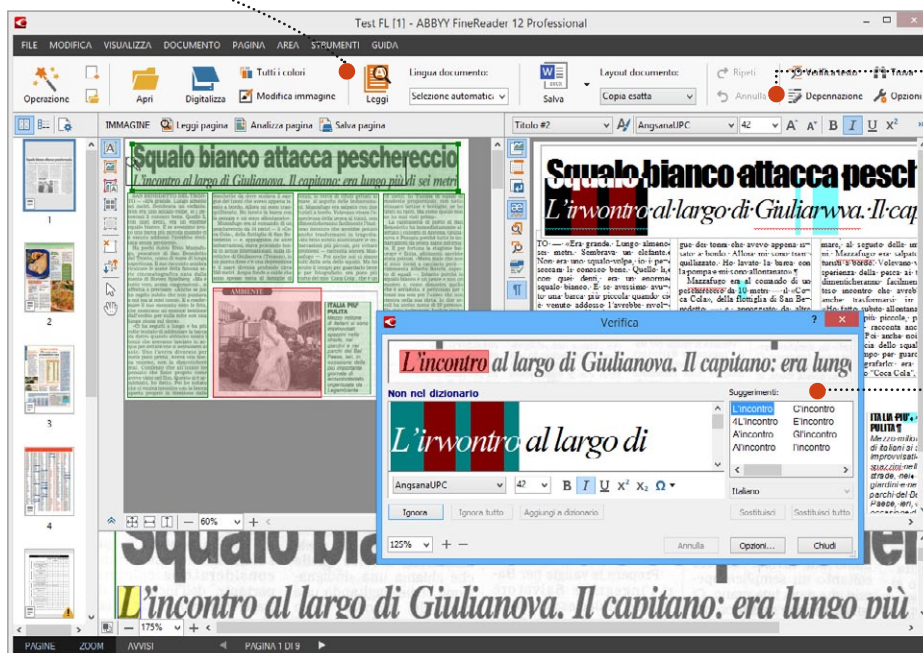
- Ottima precisione del motore di riconoscimento
- Strumenti di editing migliorati
- Elaborazione in background

CONTRO

- Interventi sul testo delle pagine già riconosciute possibili solo a fine elaborazione
- Difficoltà con il testo verticale nelle tabelle

● Produttore: Abbyy; <http://italia.abbyy.com>

NELLO STILE DI WINDOWS 8
L'ispirazione è chiara, ma le modifiche non sono puramente estetiche: il restyling dell'interfaccia ha reso FineReader più efficiente.



RIMOZIONI PERMANENTI
Il pulsante Depennazione consente di marcare il testo che si desidera far rimuovere al momento del salvataggio.

CORREZIONI PIÙ FACILI
La finestra Verifica testo ora permette di intervenire sulla formattazione e di inserire simboli e caratteri speciali.

significativi in considerazione del livello di precisione già raggiunto dalla versione precedente. Ci sono sembrate migliori anche le capacità di conservazione della struttura del documento; in particolare, l'identificazione e la formattazione delle note a piè di pagina nel test effettuato su una scansione del libro Barney's Version sono risultate nettamente più precise. Di contro abbiamo riscontrato la permanenza di alcune imprecisioni nei termini in lingua francese presenti nello stesso libro, imprecisioni che FineReader 10 invece non commetteva. Abbyy dichiara di aver raffinato gli algoritmi di preelaborazione delle immagini, in particolare quelli per il trattamento delle foto digitali, e il nostro test lo conferma: il software ha riconosciuto meglio della versione precedente sia il testo sia il layout del campione che usiamo abitualmente nelle prove, una doppia pagina (fotografata in condizioni per nulla ottimali) di un libro tascabile. Vale la pena di ricordare che il software di Abbyy è in grado di riconoscere una quantità di lingue impressionante: sono 190 in questa versione (tra cui il vietnamita, il coreano, il cinese, il giapponese, il thailandese e l'ebraico), e per 48 di esse è presente un dizionario che non solo migliora il riconoscimento ma facilita anche le operazioni di correzione. All'avvio del programma si nota subito il restyling dell'interfaccia, ora chiaramente ispirata allo stile Modern

di Windows 8. Gli interventi non sono stati però di natura puramente estetica: i ritocchi hanno reso l'interfaccia più razionale, anche se lo spostamento di alcuni pulsanti in nuove toolbar verticali rischia di mettere inizialmente in difficoltà chi conosce molto bene la release 11. Nella barra degli strumenti principali è presente una novità gradita: il pulsante Depennazione, di cui curiosamente né il manuale online né quello scaricabile in Pdf dal sito di Abbyy fanno menzione. La sua funzione è analoga a quella del tool Applica Redazione di Adobe Acrobat: permette di marcare porzioni del testo che verranno annerite e rimosse permanentemente dal documento al momento del salvataggio finale. In fase di correzione abbiamo apprezzato molto le modifiche apportate alla finestra di verifica del testo, che ora consente non solo di modificare il font e di applicare le formattazioni di base (grassetto, corsivo, sottolineato, apice e pedice), ma anche di inserire trattini facoltativi, interruzioni di linea e simboli speciali. Un'altra novità molto promettente sulla carta ma che nella pratica ci ha convinto di meno è l'elaborazione in background: con FineReader 12 non è più necessario aspettare che il programma abbia riconosciuto l'intero documento per poter lavorare sulle parti già trattate. Una possibilità allettante con i documenti molto lunghi, ma tra gli interventi consentiti purtroppo non

vi è la correzione del testo (è invece possibile, ad esempio, modificare le aree immagine o riordinare le pagine); inoltre il carico sulla Cpu dell'elaborazione in background spesso è eccessivo per consentire di lavorare in modo fluido. Inedita è anche la funzione Citazioni Rapide: sostanzialmente si tratta della possibilità di selezionare ed elaborare al volo un'area immagine di una pagina non ancora riconosciuta: il testo viene copiato negli appunti in modo da poter essere incollato dove si vuole. In alcuni casi lo stesso risultato è ottenibile con maggior rapidità con il tool Screenshot Reader fornito a corredo. Questa pratica utility permette di selezionare un'area dello schermo che viene poi riconosciuta automaticamente; il risultato dell'elaborazione viene salvato, a scelta, in un file, in un nuovo messaggio di posta o negli Appunti di Windows.

Da tempo FineReader è un programma maturo: questo aggiornamento non introduce novità rivoluzionarie, ma i ritocchi apportati da Abbyy sono tutti apprezzabili e lo rendono ancora più efficiente. Le difficoltà riscontrate con il testo verticale delle tabelle sono però un problema serio, tale da far calare di un punto la nostra valutazione complessiva, ma a chi non prevede di dover trattare molte pagine contenenti etichette verticali FineReader 12 è un software consigliabile senza esitazione. •