

OMNIPAGE 10

Cartella \Shareware\OmniPage

Il software di riconoscimento ottico, negli ultimi anni ha fatto passi da gigante: dalle primissime versioni (mi ricordo quando ancora possedevo uno scanner manuale), la Caere ha tentato di fornire un prodotto competitivo e il più possibile affidabile, anche perché in un'ottica lavorativa, è molto più semplice e comodo riuscire a catalogare documenti senza spendere troppo tempo in noiose trascrizioni.

L'obiettivo principale di un software OCR è di riportare nel modo più preciso pagine scansionate in documenti di videoscrittura o fogli elettronici mantenendo tutti gli elementi del documento originale incluso il testo, grafici, tabelle e layout.

L'accuratezza è uno dei principali fattori che devono essere considerati quando si vuole misurare la precisione dell'OCR. Altri fattori come il mantenimento del formato e la capacità di mantenere integro il formato originale giocano un ruolo importante nella precisione dell'OCR.

Qui sotto sono riportate le principali caratteristiche scoperte in OmniPage Pro inclusa l'accuratezza e il mantenimento dell'aspetto del documento originale.

Anche se il miglior OCR non è preciso al 100% su tutti i tipi di pagine scansionate, bisogna pur sempre tener conto dei miglioramenti tecnologici raggiunti dall'OCR, permettendogli di essere un metodo migliore, più affidabile, più veloce rispetto alla ritrascrizione completa del documento.

OmniPage Pro 10 include un nuovo OCR di Caere e introduce una nuova tecnologia Predictive Optical Word Recognition Plus+ (POWR++).

POWR++ ha una struttura matematica e probabilistica che dinamicamente gli permette "d'investigare" su tutte le differenti possibilità di carattere che una parola può avere.

POWR++ inizia il processo di riconoscimento analizzando algoritmi e generando una serie iniziale di ipotesi. POWR++ continua aggiornando ipotesi e probabilità. L'interrogazione continua fin tanto che non emerge una chiara e completa risposta.

L'ultimo lettore OCR di Caere incorpora centinaia di esperti con aree di esperienza specifiche e forze complementari, debolezze. Il vantaggio della nuova tecnologia POWR++ è che questo numero esteso di esperti può essere intelligentemente portato per supportare la massima precisione. Come risultato, POWR++ offre grandi miglioramenti nella precisione rispetto ad un'approccio tradizionale.

POWR++ permette a OmniPage Pro di riconoscere caratteri standard, senza doverlo addestrare, da 4 a 72 punti di dimensione. POWR++ riconosce 13 lingue (Brasiliano, Portoghese, Inglese Britannico, Danese, Olandese, Finlandese, Francese, Tedesco, Italiano, Norvegese, Spagnolo, Svedese e Inglese Americano) e per ognuna di queste lingue ha un dizionario completo. Inoltre POWR++ identifica e riconosce più lingue sulla stessa pagina.

Language analyst incorpora una varietà di tecniche che impiegano informazioni linguistiche come la frequenza delle lettere nel linguaggio, n-gram e dizionari che migliorano sensibilmente l'accuratezza della scansione.

De-skew permette automaticamente a OmniPage Pro di raddrizzare l'immagine di documenti storti prima che il riconoscimento avvenga, ovviamente la precisione non può che migliorare.

Questo miglioramento della tecnologia De-skew riduce al minimo la distorsione dei documenti durante il processo di correzione. Inoltre, per gli utenti che vogliono raddrizzare pagine particolari e vogliono farlo manualmente, OmniPage prevede la possibilità di disabilitare questa funzione grazie e adoperare Straighten Page. Questa è una funzione molto importante per gli utenti che utilizzano scanner e per quelli che usano (ADF's).

Il miglioramento della tecnologia Anypage aumenta la precisione nel riconoscimento di documenti a colori e in scala di grigio

Riconoscimento di testi invertiti permette a OmniPage di gestire testi capovolti, dove il testo è bianco su sfondo nero. Questa caratteristica è molto utile per utenti che lavorano molto con le pagine dei giornali, che spesso hanno pagine di questo tipo.

OmniPage Pro include una specifica modalità per riconoscere i fogli elettronici. Questa è un'opzione standard che viene data per mantenere il layout originale del documento che è facilmente disponibile in tutte le modalità OCR.

Massima Precisione – Mantenimento della Formattazione e della Integrità della Pagina Originale.

Nessun'altra applicazione OCR può eguagliare l'abilità di OmniPage Pro nel mantenere l'integrità e la formattazione della pagina originale: la Versione 10 utilizza diversi strumenti e tecnologie per assicurare che il layout della pagina sia sempre più uguale al layout del documento originale. OmniPage Pro inoltre offre più capacità rispetto ad altre applicazioni OCR per il mantenimento dell'integrità delle pagine originali incluso tabelle e colori.

True Page, strettamente parlando, mantiene la formattazione originale della pagina scansata riconoscendo i singoli paragrafi e oggetti all'interno del file. True Page riconosce e mantiene parole e pagine e conserva le dimensioni del font, lo stile del font (corsivo, grassetto...), liste e tabelle, grafici e dimensioni originali della pagina. Gli utenti traggono grossi benefici da questa precisa riproduzione della pagina, infatti oltre ad essere quasi una fotocopia dell'originale, l'utente ha la possibilità di modificarla a loro piacimento.

Nessun'altra applicazione offre questo preciso modo di mantenere il formato originale.

Retain Flowing Columns mantiene molto della formattazione originale della pagina scansata ma riduce il numero di frame usati. Gli utenti in questo caso possono beneficiare di una maggiore editabilità del testo.

Mantenimento del Font e del Paragrafo viene utilizzato quando gli utenti non vogliono prestare attenzione al layout della pagina ma al mantenimento della sua formattazione come parole o frasi in corsivo in modo tale da non dover riformattare il documento. Questo è sicuramente il modo più semplice per editare un documento.

Rimozione della formattazione fornisce il modo più veloce di lavorare quando formattazione e layout non sono richiesti.

Supporto dei colori. OmniPage Pro supporta la scansione di documenti a colori. Mantiene grafici a colori, e li visualizza in riquadri, immagini, testo come in una applicazione di Word. I grafici a colori sono salvati a 150 punti per pollice in un applicativo di Word o in qualunque altra applicazione di testo base.

OmniPage Pro può automaticamente individuare tabelle a griglia e ricrearle come una tavola in Word o Word Perfect in modo tale che gli utenti possano modificare la tavola o il testo al loro interno. Quando una tabella non è a griglia, gli utenti possono indicare che essa possa essere trattata con una tabella oggetto durante il processo di zoning. Se gli utenti preferiscono, possono anche creare una tabella con tabulazioni tra le colonne. Table Tools permette agli utenti di manipolare le tabelle all'interno di OmniPage Pro per un miglior output. Gli utenti possono muovere e rimuovere le righe o le colonne selezionate, inserire singole colonne, o righe divise, o rimuovere tutte le righe o singole colonne. Questi strumenti appaiono sulla nuova barra degli strumenti per facilitare l'accesso.

Anche il Nuovo Table Editing Window all'interno di OmniPage Pro, che può essere accessibile con un doppio click su una cella della tabella nel testo visualizzato, permette agli utenti di editare i contenuti delle tabelle (parole e numeri) prima di esportare il documento in un file di testo.

Questa caratteristica aiuta gli utenti a raggiungere il risultato finale desiderato più facilmente e velocemente.

La nuova tecnologia Font Matching migliora l'interpretazione dei font da parte dell'OCR. A differenza del Font Mapping, che era usato nella precedente versione di OmniPage Pro,

font matching tiene conto di font metrici come l'aspetto dei caratteri, tanto quanto dell'aspetto dei caratteri e lo spazio delle linee di testo.

Ho attinto specificatamente dalla press release inviataci insieme alla autorizzazione per la pubblicazione di questa versione di evaluation, ma è bene specificare un paio di punti.

- Gli utenti della versione demo non hanno diritto al supporto tecnico.
- Gli utenti di altri software OCR (qualunque sia il produttore) si possono aggiornare comprando il competitive upgrade
- Il competitive upgrade può essere comprato (attraverso mail order o attraverso rivenditori di software, ma per tutte queste informazioni vi rimando al sito <http://www.caere.com>).