

ScanTailor è un programma open source ideale per ottimizzare le immagini ottenute attraverso la scansione di un documento cartaceo.

Di **Marco Schiaffino**

Scansioni su misura con ScanTailor

Digitalizzare un volume cartaceo può sembrare facile. In fondo bastano uno scanner, un computer, un programma di grafica e un po' di pazienza. All'atto pratico, però, ci si rende conto che le cose non sono poi così semplici, soprattutto quando si affronta la fase di elaborazione dei file. Se l'obiettivo è quello di ottenere testi leggibili, magari per darli in pasto a un software Ocr, i normali programmi di grafica mostrano rapidamente tutti i loro limiti. Nella maggior parte dei casi ci si trova a dover utilizzare strumenti che sono pensati per altri scopi e che finiscono per rallentare il lavoro. Con ScanTailor la musica cambia: pensato specificatamente per ottimizzare le scansioni di documenti di testo, questo programma offre una serie di strumenti che consentono di snellire le procedure e ottenere subito risultati eccellenti.

ScanTailor è distribuito sotto licenza GNU ed è opera dello sviluppatore indipendente Joseph Artsimovich, che poco più di un anno fa ha deciso di abbandonare il progetto. Il testimone è stato raccolto nell'aprile dello scorso anno da Nate Craun, autore di "A Linux Guide to Book Scanning" (<https://natecraun.net/articles/linux-guide-to-book-scanning.html>). La versione più recente di ScanTailor, in ogni caso, risale

al 2012; è compatibile con Windows 8.1 ed è localizzata anche in italiano. La traduzione è però incompleta e spesso compaiono termini e comandi in inglese o qualche errore grammaticale. In definitiva è consigliabile installare la versione in inglese, che garantisce anche una maggiore aderenza con le indicazioni della guida online.

L'interfaccia del programma è strutturata sul modello di una procedura guidata, ma offre la possibilità di saltare da un passaggio all'altro con un semplice clic del mouse, consentendo di modificare le impostazioni in ogni sezione

senza dover seguire un ordine pre-stabilito. Il flusso di lavoro proposto, però, segue una logica piuttosto coerente e all'atto pratico risulta decisamente efficace. L'area di lavoro, divisa in tre sezioni, è spartana ma estremamente pratica. La sezione centrale mostra un'anteprima del file su cui si sta lavorando, mentre la parte destra permette di selezionare i file inseriti nel progetto. La sezione sinistra è invece dedicata agli strumenti di elaborazione. Questi sono divisi in sei sottosezioni, attivabili con un clic del mouse. La logica del sistema di controllo si adatta perfettamente al tipo di lavoro per cui è pensato il programma.

Software libero
ScanTailor non è semplicemente gratuito: è distribuito con la licenza GPL v.3

L'estrema specializzazione di ScanTailor emerge già nella struttura del sistema di controllo, ispirato alla massima praticità.

Visto che ogni progetto coinvolge normalmente decine (se non centinaia) di scansioni, il flusso di lavoro è pensato per ridurre al minimo gli interventi manuali. Idealmente, si può lasciar fare al programma e limitarsi a correggere eventuali imperfezioni sulle singole scansioni. L'applicazione delle modifiche può essere fatta per ogni singolo file con un clic del mouse, ma la soluzione più pratica è quella di avviare l'elaborazione di tutte le immagini facendo clic sul piccolo pulsante Launch batch processing che compare accanto alla sezione attiva e ha la forma della classica icona "Play". In questo modo è possibile ottimizzare i tempi di lavoro senza pregiudicare la qualità del risultato.

Il primo passo per l'avvio del progetto è la selezione della cartella di origine delle scansioni. Da qui si passa direttamente alla fase di elaborazione, il

**SCANTAILOR
0.9.11.1**

Gratuito

+ PRO

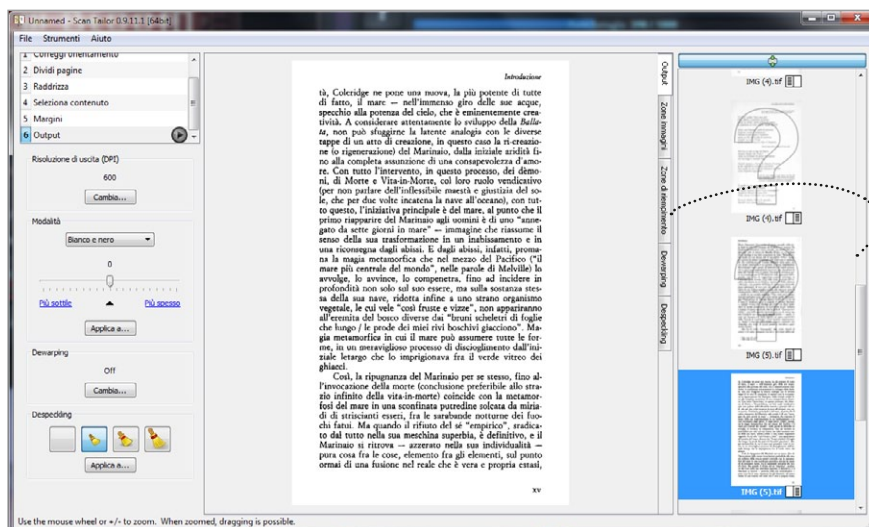
Strumenti specializzati • Eccellente analisi automatica delle immagini

- CONTRO

Traduzione in italiano incompleta e approssimativa • Manca la funzione Annulla

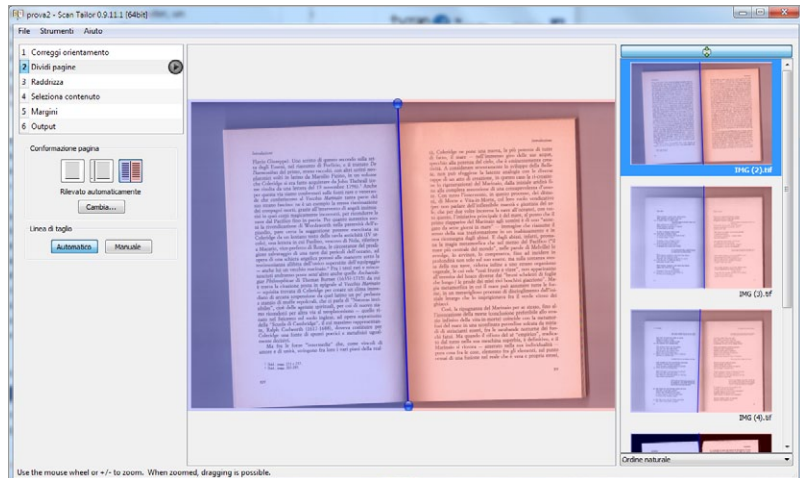
Produttore: www.scantailor.org

**VOTO
8,5**

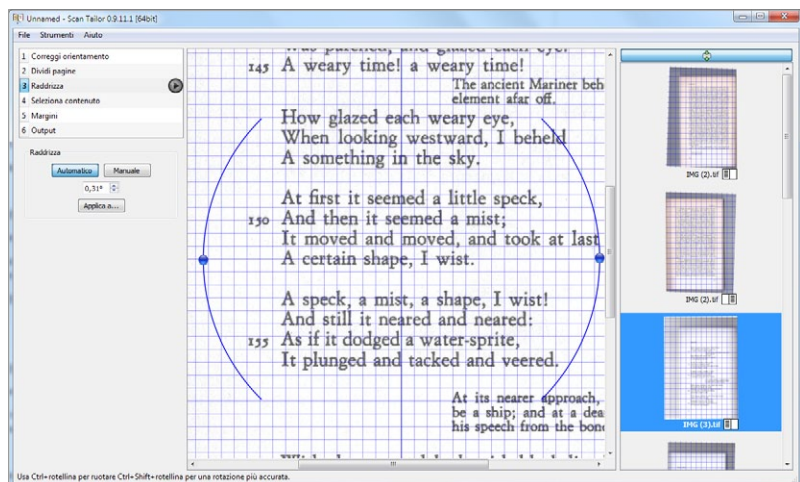


cui primo passaggio è la correzione dell'orientamento delle immagini. Teoricamente si tratta di una funzione superflua, visto che è possibile impostarlo già al momento della scansione. Nel caso in cui ci si dovesse ritrovare con immagini disomogenee, o per una semplice dimenticanza non sia stato impostato l'orientamento corretto, la funzione risulta comunque utile. Anche perché il sistema di controllo consente di applicare la rotazione a tutte le immagini inserite nel progetto, permettendo di portare a termine l'operazione per tutte le immagini con due soli clic del mouse. Le vere potenzialità di ScanTailor emergono già quando si passa alla successiva funzione Dividi pagine. Si tratta di uno strumento pensato per escludere dall'immagine eventuali porzioni di testo delle pagine adiacenti, ma può essere utilizzato anche per ridurre drasticamente i tempi necessari a eseguire la scansione. Il sistema di rilevamento automatico delle pagine è infatti talmente efficace da consentire di digitalizzare due pagine alla volta e lasciare al programma il compito di separarle in fase di elaborazione. In questo modo, se stiamo digitalizzando un volume di piccole o medie dimensioni con un normale scanner piano, diventa possibile dimezzare il numero di scansioni. Nella remota ipotesi in cui l'immagine sia così confusa da mettere in crisi il sistema di rilevamento, è sempre possibile intervenire a mano per fissare la corretta linea di separazione tra le due pagine. Una volta individuate le pagine, il programma le separa automaticamente creando un'immagine per ogni pagina all'interno del progetto. Tutto questo senza che sia necessario muovere un dito.

La stessa accuratezza caratterizza il rilevamento automatico sia della fase Raddrizza sia di quella di selezione del contenuto, per le quali la necessità di un intervento manuale è davvero remota. Nella maggior parte dei casi, quindi, sarà sufficiente selezionare una a una le pagine e lasciare che il software rilevi il contenuto apportando le correzioni necessarie. Le cose si fanno più complicate quando ci si trova a impostare i margini della pagina. La procedura è infatti meno intuitiva, visto che il programma consente di impostare sia i margini tra il contenuto e la pagina, sia l'allineamento rispetto alle altre pagine. Per capire le



L'affidabilità del sistema di rilevamento automatico permette di eseguire la scansione di una doppia pagina lasciando al programma il compito di effettuare la separazione.



La funzione Raddrizza di ScanTailor non sbaglia un colpo. Non si tratta di semplice perfezionismo: una scansione perfettamente allineata facilita non solo la lettura, ma anche il lavoro di un software Ocr.

ripercussioni pratiche delle modifiche serve qualche esperimento, ma qui le cose sono complicate dal fatto che il programma non prevede il comando Annulla. Meglio quindi entrare nell'ottica di salvare spesso il progetto, visto che eventuali modifiche sperimentali non possono essere eliminate con la classica raffica di Ctrl-Z.

La sezione Output rappresenta il passo finale nella produzione delle immagini e contiene una ricca dotazione di strumenti particolarmente utili (tra cui la funzione di *Despeckling*, che permette di ripulire le scansioni da eventuali imperfezioni generate nella procedura di scansione). Il più efficace, però, è quello che consente di modificare lo spessore della linea che compone i caratteri, consentendo

di dare maggiore consistenza al testo o, al contrario, alleggerirlo se dovesse risultare troppo carico. Nel caso in cui le scansioni siano deformate, è possibile anche correggerne la prospettiva attraverso la funzione *Dewarping*. Anche in questo caso, però, pesa l'assenza del comando Annulla.

La modalità di correzione automatica, infatti, è ancora in una fase sperimentale e quella manuale si affida a un sistema intuitivo ma piuttosto impreciso, che prevede il trascinarsi con il mouse di una griglia per applicare la compensazione della distorsione. Piccoli difetti che, come le già citate imperfezioni nella traduzione, non incidono più di tanto sul valore davvero elevato di questo software. •