

Datasheet

PDFlib TET PDF IFilter 5

Enterprise PDF Search Per Windows

Che cos'è PDFlib TET PDF IFilter?

TET PDF IFilter estrae il testo e i metadati dai documenti PDF rendendoli disponibili ai programmi di ricerca su Windows. È così possibile ricercare i documenti in locale, attraverso un server aziendale o via web. TET PDF IFilter è uno strumento basato sul Text and Image Extraction Toolkit (TET), la soluzione brevettata di PDFlib per sviluppatori, per estrarre in modo sicuro il testo dai documenti PDF.

TET PDF IFilter è una robusta implementazione dell'interfaccia di indicizzazione IFilter di Microsoft. Supporta qualsiasi applicazione compatibile con IFilter, come SharePoint e SQL Server. Questi prodotti utilizzano filtri specifici – chiamati IFilters – per particolari formati di file, ad es. HTML. TET PDF IFilter è un programma di questo tipo, destinato ai documenti PDF. L'interfaccia grafica per la ricerca può essere Windows Explorer così come un Web o database frontend, uno script o un'applicazione personalizzata. Altrimenti è possibile eseguire query di ricerca direttamente nel programma senza alcuna interfaccia grafica.

Tecnologia TET brevettata

La prima versione di PDFlib TET, il motore di TET PDF IFilter, è stata rilasciata nel 2002 ed è utilizzata in tutto il mondo in ambienti server e desktop. In alternativa all'estrazione dalle pagine PDF dei testi e dei metadati come testo puro, TET estrae anche in formato XML. TET inoltre è disponibile come plugin gratuito per Acrobat. L'utilizzo di questo plugin permette di conoscere e testare le potenzialità di TET.

Vantaggi

TET PDF IFilter offre i seguenti vantaggi:

- ▶ Supporta i testi occidentali, il cinese, il giapponese e il coreano (CJK) nonché le lingue da destra a sinistra come l'arabo e l'ebraico
- ▶ Indicizza i documenti protetti ed estrae il testo anche dai PDF quando Adobe non è in grado
- ▶ Supporta il folding Unicode, la decomposizione e la normalizzazione
- ▶ Utilizzo: thread-safe, veloce e robusto, versioni 32 e 64 bit
- ▶ Rilevamento automatico degli script e della lingua per una ricerca migliorata

Enterprise PDF Search

TET PDF IFilter è disponibile in versione a 32 e 64 bit completamente thread-safe. È possibile realizzare soluzioni aziendali interfacciando TET PDF IFilter con:

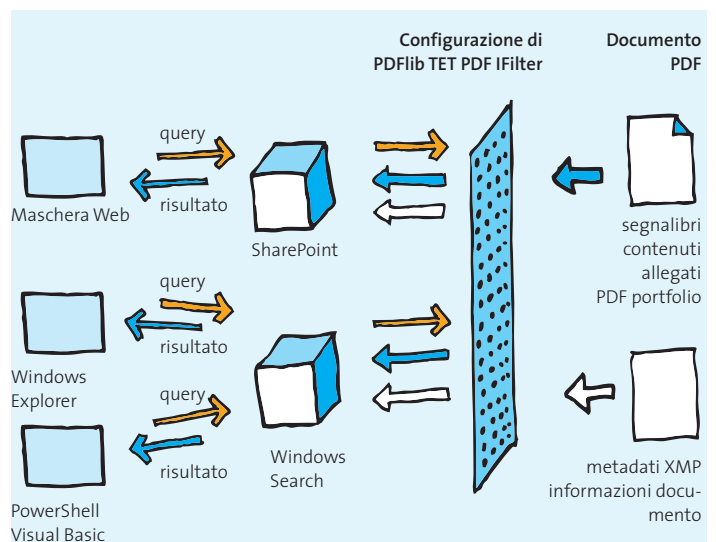
- ▶ Microsoft SharePoint Server 2013 e precedenti
- ▶ Microsoft Search Server
- ▶ Microsoft SQL Server
- ▶ Microsoft Exchange Server
- ▶ Microsoft Site Server

TET PDF IFilter può essere usato con tutti i prodotti di Microsoft e di altre aziende che supportano l'interfaccia IFilter.

Ricerca desktop

TET PDF IFilter può essere utilizzato per ricerche desktop, ad esempio con Windows Search, soluzione già integrata in Windows.

TET PDF IFilter è gratuito per uso non commerciale sui sistemi operativi desktop e rappresenta un modo pratico ed economico per conoscerne le sue potenzialità.



Funzionalità

PDF supportati

TET PDF IFilter supporta i seguenti formati:

- ▶ Tutte le versioni PDF fino ad Acrobat DC, compreso ISO 32000-1 e 2
- ▶ PDF protetti che non richiedono una password per la consultazione
- ▶ I PDF corrotti vengono riparati

Unicode Postprocessing

TET PDF IFilter supporta una serie di fasi di postprocessing in formato Unicode che possono essere richiamate per migliorare la qualità del testo estratto:

- ▶ Mantenimento del folding, rimozione o sostituzione dei caratteri, ad esempio rimozione di punteggiatura o caratteri da script non rilevanti.
- ▶ Decomposizioni per sostituire un carattere con una sequenza equivalente di uno o più caratteri, ad es. sostituzione di caratteri giapponesi stretti, larghi o verticali oppure delle varianti di superscript Latin (ad es. *) con le rispettive controparti standard.
- ▶ Il testo può essere convertito in tutte le quattro forme normalizzate Unicode, ad es. emit NFC per soddisfare i requisiti per testo Web o per le basi di dati.

Internazionalizzazione

Oltre al testo occidentale, TET PDF IFilter supporta il cinese, il giapponese e il coreano (CJK). Tutti gli encoding CJK sono riconosciuti così come la scrittura orizzontale e verticale. Il rilevamento automatico delle impostazioni di localizzazione (lingua e regione) sono inoltre funzionalità molto utilizzate in Estremo Oriente poiché migliorano gli algoritmi di separazione delle parole e di stemming di Microsoft.

Anche le lingue con scrittura da destra a sinistra, come l'arabo e l'ebraico, sono supportate. La forma dei caratteri contestuali viene normalizzata e il testo viene reso in ordine logico.

PDF è molto più che testo impaginato

TET PDF IFilter tratta i documenti PDF come se fossero contenitori al cui interno sono salvate numerose informazioni. TET PDF IFilter indicizza tutte le informazioni nei documenti PDF:

- ▶ Contenuti
- ▶ Segnalibri
- ▶ Metadata (vedi sotto)
- ▶ I PDF embedded e i pacchetti/portofogli PDF vengono elaborati in modo ricorsivo così da permettere di eseguire ricerche nei testi di tutti i documenti PDF.

Metadata XMP e informazioni documento

L'implementazione avanzata di TET PDF IFilter supporta le proprietà di sistema di Windows per i metadata. Indicizza i metadata XMP nonché le informazioni standard e personalizzate dei documenti. L'indicizzazione dei metadata può essere configurata in molti livelli:

- ▶ Le informazioni sul documento, i campi Dublin Core e le altre proprietà XMP vengono convertite nell'equivalente proprietà Windows, come *Title*, *Subject*, *Author*.
- ▶ TET PDF IFilter aggiunge utili pseudo-proprietà, come dimensione della pagina, livello di conformità PDF/A, nomi dei font.
- ▶ Possibilità di cercare tutte le principali proprietà XMP predefinite.
- ▶ Possono essere ricercate anche proprietà XMP personalizzate, come classificazioni aziendali ed estensioni PDF/A.

TET PDF IFilter può, come opzione, salvare i metadata in formato testo, in modo da poter essere utilizzati da motori che non supportano i metadata, come SQL Server.

Benefici di PDFlib

Affidabile

In tutto il mondo sono migliaia i programmatori che utilizzano il nostro software. PDFlib soddisfa tutti i requisiti in termini di qualità e performance per l'utilizzo sui server. Tutti i prodotti PDFlib sono adatti per l'utilizzo continuo su server e per le operazioni di batch senza supervisione.

Semplice e veloce

I prodotti PDFlib sono estremamente veloci e sono in grado di elaborare fino a centinaia di pagine al secondo. L'interfaccia di programmazione è intuitiva e di facile apprendimento.

I prodotti PDFlib sono utilizzati in tutto il mondo

I nostri prodotti supportano tutte le lingue internazionali e Unicode. Vengono utilizzati da utenti in ogni angolo del mondo.

Assistenza professionale

In caso di problemi potete contare sul nostro servizio di assistenza professionale. Per garantire che il workflow a livello aziendale non subisca mai interruzioni, offriamo la possibilità di sottoscrivere un contratto di servizio. In questo modo potrete accedere alle versioni più recenti e potrete contare su tempi di risposta celeri qualora si dovesse verificare un problema.

Licenza

Sono a disposizione diverse tipologie di licenza: server, integrativa, aziendale e codice sorgente. È altresì possibile sottoscrivere un contratto di servizio per usufruire di un'assistenza professionale con tempi brevi di risposta e aggiornamenti gratuiti.

L'azienda PDFlib GmbH

PDFlib GmbH è un'azienda specializzata nel settore delle tecnologie PDF. I prodotti PDFlib vengono utilizzati in tutto il mondo dal 1997. L'azienda segue attentamente gli sviluppi e i trend del mercato, come gli standard ISO per il formato PDF. PDFlib GmbH distribuisce i suoi prodotti in tutti i continenti e detiene una posizione leader nei mercati in Nordamerica, Europa e Giappone.

Contatti

Sul nostro sito Internet sono disponibili le versioni demo complete di documentazione ed esempi. Per maggiori informazioni, contatta:



PDFlib GmbH
 Franziska-Bilek-Weg 9, 80339 München, Germany
 phone +49 • 89 • 452 33 84-0, fax +49 • 89 • 452 33 84-99
 sales@pdflib.com
 www.pdflib.com