

# Datasheet

## PDFlib

### TET PDF IFilter

*Enterprise PDF Search per Windows*

### Che cos'è PDFlib TET PDF IFilter?

PDFlib 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 Extraction Toolkit (TET), la soluzione brevettata di PDFlib, 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. Queste applicazioni usano particolari programmi, chiamati IFilters, per ricercare svariati formati di file, come HTML. TET PDF IFilter è dedicato ai documenti PDF. L'interfaccia grafica per la ricerca può essere Windows Explorer così come un 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 dei testi e 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 le potenzialità di TET.

### Vantaggi

TET PDF IFilter offre i seguenti vantaggi:

- ▶ Estrazione di testo dove Acrobat fallisce
- ▶ Indicizzazione non solo di contenuti, ma anche di metadati, bookmark, allegati e pacchetti
- ▶ Performance: thread-safe, veloce e robusto, versione 32 e 64 bit
- ▶ Prodotto standalone senza dipendenze particolari
- ▶ Rilevamento automatico di linguaggi/script
- ▶ Supporto da parte di un team dedicato

### 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:

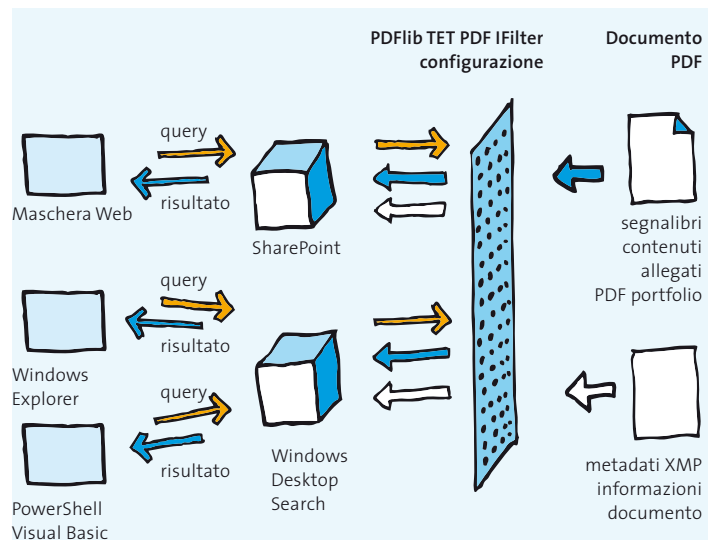
- ▶ Microsoft Office SharePoint Server (MOSS)
- ▶ Microsoft Search Server 2008 e Search Server 2008 Express
- ▶ Microsoft SQL Server Microsoft Exchange Server
- ▶ TET PDF IFilter può essere usato con tutti i prodotti di Microsoft e di altre aziende che supportano l'interfaccia IFilter.

### Ricerca desktop

TET PDF IFilter può essere utilizzato per ricerche desktop, ad esempio con:

- ▶ Windows Desktop Search (WDS): integrato in Windows Vista; disponibile anche come add-on gratuito per Windows XP
- ▶ Windows Indexing Service

TET PDF IFilter è gratuito per uso desktop non commerciale 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 9
- ▶ PDF crittografati che non richiedono una password per la consultazione
- ▶ PDF danneggiati e possibilmente riparabili

### Metadati XMP ed altre informazioni

L'implementazione avanzata di TET PDF IFilter supporta le proprietà di sistema di Windows per i metadati. Indicizza i metadati XMP (linguaggio Adobe per metadati basato su XML) così come altre informazioni sul documento. L'indicizzazione dei metadati 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 *Titolo, Oggetto, Autore*.
- ▶ TET PDF IFilter aggiunge utili pseudo-proprietà, come dimensione della pagina, livello di conformità PDF/A, nomi dei font.
- ▶ Tutte le più importanti proprietà XMP possono essere ricercate, come *dc:rights, xmpRights:UsageTerms, xmp:CreatorTool*.
- ▶ Possono essere ricercate anche proprietà XMP personalizzate, come classificazioni aziendali ed estensioni PDF/A.

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

### Metadati XMP delle immagini

Oltre ai metadati del documento, TET PDF IFilter supporta anche i metadati XMP relativi alle immagini. Nei sistemi moderni, le immagini sono arricchite di vari metadati, dal modello di fotocamera digitale utilizzata, alle modifiche eseguite con Photoshop, al layout utilizzato. TET PDF IFilter estrae questi dati e li rende disponibili alla ricerca. Per esempio è possibile ricercare immagini appartenenti ad una determinata categoria o realizzate da un determinato fotografo.

### Internazionalizzazione

TET PDF IFilter supporta l'estrazione di testo in Cinese, Giapponese e 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.

### 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 le indicizza:

- ▶ Contenuti
- ▶ Segnalibri
- ▶ PDF inclusi ed indicizzati ricorsivamente così da permettere la ricerca in documenti PDF allegati.
- ▶ Indicizzazione di tutti i documenti in un pacchetto PDF. I pacchetti sono una funzionalità introdotta in Acrobat 8 per raggruppare i PDF, la funzionalità è chiamata portfolio in Acrobat 9).

## Benefici di PDFlib

### Affidabile

In tutto il mondo sono migliaia i programmatori che utilizzano il nostro software. I prodotti PDFlib soddisfano 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 da leader di mercato in Nordamerica, Europa e Giappone.

### Contatti

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



#### PDFlib GmbH

Franziska-Bilek-Weg 9, 80339 Monaco, Germania  
 Te. +49 • 89 • 452 33 84-0, Fax +49 • 89 • 452 33 84-99  
 sales@pdflib.com  
 www.pdflib.com