

## Datenblatt

# PDFlib TET 5.2

## Text and Image Extraction Toolkit

### Was ist PDFlib TET?

PDFlib TET (Text and Image Extraction Toolkit) extrahiert zuverlässig Text, Bilder und Metadaten aus PDF-Dokumenten. TET stellt den Text eines PDF-Dokuments als Unicode-Strings zur Verfügung und liefert detaillierte Informationen zu Farbe, Glyphen und Fonts sowie die Position auf der Seite. Rasterbilder werden in gebräuchliche Bildformate extrahiert. Optional kann TET die PDF-Dokumente in ein XML-basiertes Format namens TETML konvertieren, das Text und Metadaten sowie Ressource-Informationen enthält.

TET verfügt über einen ausgefeilten Algorithmus zur Inhaltsanalyse und kann damit Wortgrenzen erkennen, Text zu Spalten zusammenfassen, Tabellenstrukturen erkennen und redundanten Text entfernen. Mit der pCOS-Schnittstelle können Sie zudem beliebige Objekte aus einem PDF-Dokument abfragen, zum Beispiel Metadaten oder interaktive Elemente.

Mit PDFlib TET können Sie:

- ▶ PDF-Indexer für eine Suchmaschine implementieren
- ▶ Texte und Bilder aus PDFs weiterverarbeiten
- ▶ PDF-Inhalte in andere Formate konvertieren
- ▶ PDFs abhängig vom Inhalt verarbeiten, z.B. anhand von Überschriften aufteilen (erfordert zusätzlich PDFlib+PDI)
- ▶ Prüfen, ob eine bestimmte Stelle auf der Seite leer ist, z.B. um einen Barcode oder Stempel zu platzieren

### Funktionalität von PDFlib TET

#### Akzeptierte PDF-Eingabe

TET verarbeitet alle gängigen Varianten von PDF-Dokumenten:

- ▶ Alle PDF-Versionen bis Acrobat DC einschließlich ISO 32000-1 und ISO 32000-2 (PDF 2.0)
- ▶ Geschützte PDFs, die zum Öffnen kein Kennwort benötigen
- ▶ Beschädigte PDF-Dokumente werden repariert

#### Alle Schriftsysteme der Welt

TET verarbeitet PDF-Dokumente in allen Schriftsystemen der Welt und implementiert eine für einige Schriftsysteme benötigte spezielle Verarbeitung:

- ▶ Lateinische, griechische und kyrillische Schriftsysteme einschließlich Entfernung von Trennstellen

- ▶ Arabisch und Hebräisch einschließlich logischer Neuordnung von linksläufigem und bidirektionalem Text; Normalisierung der arabischen Präsentationsformen
- ▶ Vereinfachtes und traditionelles Chinesisch, Japanisch und Koreanisch; horizontaler und vertikaler Text
- ▶ Indische Schriftsysteme (ohne Umsortierung von Glyphen)
- ▶ Unicode-Unterstützung für alle anderen Sprachen und Schriftsysteme

#### Unicode

Da Text in PDF normalerweise nicht in Unicode kodiert ist, wandelt PDFlib TET den Text in einem PDF-Dokument in Unicode um:

- ▶ TET konvertiert alle Textinhalte nach Unicode, unabhängig von der im PDF-Dokument verwendeten Encoding-Methode.
- ▶ Ligaturen und andere zusammengesetzte Glyphen werden als Sequenz der entsprechenden Unicode-Zeichen ausgegeben.
- ▶ Glyphen ohne nutzbare Unicode-Information werden erkannt und auf ein konfigurierbares Ersatzzeichen abgebildet, um Fehlinterpretationen zu verhindern.
- ▶ In TET sind verschiedene Workarounds für bestimmte Erzeugerprogramme implementiert, etwa für InDesign- und TeX-Dokumente sowie PDFs, die auf Mainframe-Systemen generiert wurden.

#### Inhaltsanalyse und Worterkennung

TET enthält patentierte Algorithmen zur Inhaltsanalyse:

- ▶ Erkennen von Wortgrenzen zur korrekten Ausgabe von Wörtern
- ▶ Getrennte Silben wieder zu Wörtern zusammensetzen (Enttrennung)
- ▶ Entfernen von redundantem Text, zum Beispiel bei Schatteneffekten oder simulierter Fettschrift
- ▶ Kombinieren von Absätzen gemäß der Lesereihenfolge
- ▶ Korrektes Umsortieren von Text, der über die Seite verstreut ist

#### Seitenlayout, Tabellen- und Listenerkennung

TET analysiert die Seiteninhalte, um Textspalten zu erkennen. Tabellen werden erkannt – einschließlich Tabellenzellen, die sich über mehrere Zeilen oder Spalten erstrecken. Damit lässt sich die Reihenfolge des extrahierten Textes verbessern. Tabellenzeilen und der Inhalt einzelner Tabellenzellen werden als solche erkannt, ebenso Aufzählungen.

## Geometrie

TET liefert präzise geometrische Angaben zum Text, z.B. die Position auf der Seite, Glyphenbreiten und Textrichtung. Für die Textextraktion können Seitenbereiche gezielt ein- oder ausgeschlossen werden, z.B. um Kopf- und Fußzeilen oder Seitenränder zu übergehen.

## Textfarbe

TET analysiert Farbangaben in der PDF-Seitenbeschreibung und liefert für jede Glyphe genaue Farbangabe. Damit lassen sich z.B. Überschriften oder hervorgehobener Text identifizieren. Komplexe Farbräume wie Separation und DeviceN können optional als einfacher Alternativfarbraum extrahiert werden.

## Extraktion von Rasterbildern

Bilder auf den Seiten einer PDF-Datei lassen sich als TIFF-, JPEG-, JBIG2- oder JPEG-2000-Dateien extrahieren. Zusätzlich werden genaue geometrische Informationen (Position, Größe und Winkel) für jedes Bild geliefert. Fragmentierte Bilder werden zu einem größeren Bild zusammengesetzt, um die Wiederverwertung zu vereinfachen. Die originalgetreue Darstellung ist dadurch garantiert, dass beim Extrahieren keine Neuberechnung der Auflösung (Downsampling) und keine Farbkonvertierung stattfindet. Damit erhält TET die höchstmögliche Bildqualität.

## Ignorieren von Artefakten in Tagged PDF

In Tagged PDF und damit insbesondere PDF/UA können irrelevante Inhalte als Artefakt markiert sein (z.B. Kopf- und Fußzeilen). TET ignoriert optional Text- und Bildartefakte.

## PDF-Analyse mit der pCOS-Schnittstelle

Die TET-Bibliothek enthält die pCOS-Schnittstelle zur Abfrage von Detailinformationen über ein PDF-Dokument, wie Dokument-Infofelder, Metadaten, Fontlisten, Seitengröße und vieles mehr (siehe separates pCOS-Datenblatt).

## Konfiguration für problematisches PDF

TET verfügt über spezielle Verfahren zur Behandlung verschiedener Klassen von PDF, deren Text sich mit anderen Produkten nicht extrahieren lässt. Außerdem bietet TET Konfigurationsmöglichkeiten, mit denen sich die Verarbeitung problematischer Dokumente erheblich verbessern lässt:

- ▶ Das Unicode-Mapping lässt sich mit Tabellen anpassen.
- ▶ PDFlib FontReporter ist ein kostenloses Plugin für Adobe Acrobat zur Analyse von Fonts, Encodings und Glyphen in PDF.
- ▶ Eingebettete Fonts werden nach zusätzlichen Daten für das Unicode-Mapping durchsucht. Ist ein Font nicht eingebettet, lässt sich die Extraktion durch externe Fontdateien oder Systemfonts weiter verbessern.

## Unicode-Nachbearbeitung

TET unterstützt Unicode-Nachbearbeitung:

- ▶ Foldings erhalten, entfernen oder ersetzen Zeichen, um z.B. Interpunktionszeichen oder Zeichen aus einem irrelevanten Schriftsystem zu filtern.
- ▶ Dekompositionen ersetzen ein Zeichen mit einer äquivalenten Folge von einem oder mehreren anderen Zeichen, z.B. Ersetzen von schmalen, breiten oder vertikalen japanischen Zeichen oder lateinischen hochgestellten Zeichen (z.B. <sup>a</sup>) mit dem entsprechenden normalen Zeichen.
- ▶ Text kann in alle vier Unicode-Normalformen konvertiert werden, z.B. um Texte in NFC-Form auszugeben, damit sie den Anforderungen einer Website oder Datenbank genügen.

## Dokumentdomänen

PDF-Dokumente können außer den Seiteninhalten noch an anderen Stellen Text enthalten, die oft relevant sind. TET extrahiert Text aus allen Domänen:

- ▶ Seiteninhalte
- ▶ Standard- und benutzerdefinierte Dokument-Infofelder
- ▶ XMP-Metadaten auf Dokument- und Bildebene
- ▶ Lesezeichen und Kommentare (Anmerkungen)
- ▶ Dateianhänge und PDF-Portfolios werden rekursiv verarbeitet
- ▶ Formularfelder
- ▶ Allgemeine PDF-Eigenschaften lassen sich abfragen, z.B. Seitenanzahl, Übereinstimmung mit Standards wie PDF/A, PDF/X usw.

## XMP-Metadaten

TET unterstützt XMP-Metadaten auf mehrere Arten:

- ▶ Mit der integrierten pCOS-Schnittstelle lassen sich XMP-Metadaten für das Dokument, einzelne Seiten, Bilder oder andere Teile des Dokuments programmatisch abfragen.
- ▶ Die TETML-Ausgabe enthält die XMP-Daten des Dokuments und der Bilder, sofern im PDF vorhanden.
- ▶ Werden Bilder in den Formaten TIFF oder JPEG exportiert, bettet TET auch XMP-Bildmetadaten ein, sofern vorhanden.

## TETML stellt PDF-Inhalte als XML dar

Optional kann TET die Inhalte einer PDF-Datei in einer XML-Variante namens TETML darstellen, mit der sich eine Vielzahl der im PDF enthaltenen Informationen in einer Form darstellen lässt, die für gebräuchliche XML-Werkzeuge zugänglich ist. TETML enthält den eigentlichen Text, optional ergänzt durch Fontinformationen und Positionsangaben, Details zu den Ressourcen (Fonts, Bilder, Farbräume) und Metadaten.

TETML umfasst auch interaktive Elemente wie Formularfelder, Anmerkungen, Lesezeichen usw. und kann sogar zur Analyse von JavaScript, Farbraumangaben, ICC-Profilen oder Druckausgabebedingungen verwendet werden.

TETML lässt sich mit XSLT-Stylesheets verarbeiten, um z.B. Filter anzuwenden oder TETML in andere Formate zu konvertieren. Die TET-Distribution enthält Beispiele für XSLT-Stylesheets zur Verarbeitung von TETML.

Der Ausschnitt zeigt TETML-Ausgabe mit Details zu den Glyphen:

```
<Word>
<Text>PDFlib</Text>
<Box llx="111.48" lly="636.33" urx="161.14" ury="654.33">
<Glyph font="F1" size="18" x="111.48" y="636.33" width="9.65">P</Glyph>
<Glyph font="F1" size="18" x="121.12" y="636.33" width="11.88">D</Glyph>
<Glyph font="F1" size="18" x="133.00" y="636.33" width="8.33">F</Glyph>
<Glyph font="F1" size="18" x="141.33" y="636.33" width="4.88">l</Glyph>
<Glyph font="F1" size="18" x="146.21" y="636.33" width="4.88">i</Glyph>
<Glyph font="F1" size="18" x="151.08" y="636.33" width="10.06">b</Glyph>
</Box>
</Word>
```

## TET-Konnektoren

TET-Konnektoren verbinden TET mit anderer Software. Die folgenden TET-Konnektoren ermöglichen die Textextraktion aus PDF in den jeweiligen Software-Umgebungen:

- ▶ TET-Konnektor für die Suchmaschine Lucene
- ▶ TET-Konnektor für den Solr Search Server
- ▶ TET-Konnektor für das Apache Tika-Toolkit
- ▶ TET-Konnektor für Oracle Text
- ▶ TET-Konnektor für MediaWiki
- ▶ TET PDF IFilter für Windows ist als separates Produkt erhältlich. Es extrahiert Text und Metadaten aus PDF und macht sie Such- und Retrieval-Software zugänglich (siehe separates Datenblatt).

## Probleme bei der PDF-Textextraktion

### Enttrennung (Dehyphenation)

TET entdeckt getrennte Wörter, die sich über mehrere Zeilen erstrecken, entfernt den Trennstrich und verbindet die Einzelteile wieder zu einem Wort. Das ist insbesondere für die Volltextsuche wichtig, damit auch Wörter, die nur getrennt vorkommen, gefunden werden. Bindestriche (im Unterschied zu Trennstrichen) werden gesondert behandelt, da sie nicht entfernt werden dürfen.

strategische Grundsätze – der  
 der Nutzung von Synergie-  
 in Branchen sowie in Unter-  
 lusterstellung. So verringert  
 bei der Produkterstellung –  
 g – seit längerem nicht nur

Trennstriche werden korrekt entfernt, Gedankenstriche bleiben jedoch erhalten.

### Schatten und künstliche Faltungen

Digitale Dokumente enthalten oft Text mit Schatten, bei denen der Schatteneffekt durch mehrfache versetzte Platzierung des Textes erzielt wird. Ähnlich werden auch Textfaltungen häufig durch mehrfaches Übereinanderlegen des Textes simuliert. Solche Dokumente enthalten die einzelnen Buchstaben des Wortes mehrfach. Der patentierte Algorithmus von TET zur Schattenerkennung erkennt die Situation und beseitigt die doppelten Zeichen, um die Extraktion irrelevanter Buchstaben zu vermeiden. Während andere Produkte den schattierten oder künstlich gefalteten Text mehrfach extrahieren, entfernt TET die redundanten Zeichen korrekt. Während ein doppeltes Wort bei der Volltextsuche noch gefunden werden könnte, würde die Verdoppelung einzelner Buchstaben wie im Beispiel Treffer im Text vereiteln.

Introduction

Andere Produkte extrahieren »Inttroduccttiion«.  
 TET extrahiert korrekt »Introduction«.

### Akzentzeichen

In vielen Sprachen werden Akzente und andere diakritische Zeichen nah an anderen Zeichen platziert, um kombinierte Zeichen zu bilden. Einige Satzprogramme, vor allem TeX, geben zwei getrennte Zeichen aus (Basiszeichen und Akzent), um ein kombiniertes Zeichen zu erzeugen. Um zum Beispiel das Zeichen *ä* zu erzeugen, wird zuerst der Buchstabe *a* auf der Seite platziert und dann die Umlautpunkte *¨* darüber gesetzt. TET erkennt dies und kombiniert die beiden Zeichen zu dem entsprechenden Buchstaben mit Akzent.

Observatoire de Paris, LERMA, 61 avenue  
 Observatoire **Midi-Pyrénées**, UMR 5572,  
 Department of Astronomy, University of  
 Observatorio Astronomico di Bologna, vi

Andere Produkte extrahieren »Midi-Pyr'en'ees«.  
 TET extrahiert korrekt »Midi-Pyrénées«.

### Ligaturen

Ligaturen kombinieren zwei oder mehr Zeichen zu einer einzigen Glyphe. Die gebräuchlichsten Ligaturen sind für die Kombinationen *fi*, *fl* und *ffi* im Gebrauch. Weniger verbreitet sind zum Beispiel Ligaturen für die Kombinationen *Th*, *sp*, *ct*, *st*. Wenn der Text aus digitalen Dokumenten extrahiert werden soll, müssen Ligaturen analysiert und zerlegt werden, damit sich der Text korrekt verwenden lässt. TET erkennt Ligaturen nach verschiedenen Kriterien und extrahiert korrekt zwei oder mehr Zeichen.

is permanently hidden from Earth.  
**The first photographs** of the hic  
 cial satellite; modern satellites prov

Andere Produkte extrahieren »e rst photographs«.  
 TET extrahiert korrekt »The first photographs«.

### Initialen

Initialen (*Drop Caps*) sind große Anfangsbuchstaben zu Beginn eines Absatzes, wobei der Scheitel der Initiale mit der Höhe der Zeile bündig sind und der Rest des Zeichens sich über mehrere Zeilen nach unten erstreckt. Initialen werden benutzt, um den Start eines neuen Absatzes zu betonen. Wenn sie nicht korrekt behandelt werden, wird das erste Wort in zwei Teilen extrahiert: dem einzelnen ersten Buchstaben und dem verbleibenden Rest des Wortes.

**S**tellen Sie sich vor, Sie stehen an einem  
 Kinder ins Wasser springen und schwim  
 vor, Sie graben am Sandstrand zwei klei  
 Schritte landeinwärts, jeder eine Hand breit, so  
 Kanäle fließen kann. Stellen Sie sich jetzt noch  
 mittels eines Streichholzes und kleiner weißer

Andere Produkte extrahieren zwei Wörter: »S« und »tellen«.  
 TET extrahiert korrekt das ganze Wort »Stellen«.

## Probleme bei der PDF-Textextraktion

### Unicode-Mapping

Unicode-Mapping bildet die Grundlage der Text-Extraktion aus PDF: Jeder Glyphen auf einer Seite muss der korrespondierende Unicode-Wert zugeordnet werden. PDF erschwert diese Aufgabe dadurch, dass es eine Vielzahl von Font- und Encoding-Varianten unterstützt. Obwohl die meisten PDFs genügend Informationen zur Bestimmung der richtigen Unicode-Werte enthalten, ist das nicht immer der Fall. Im schlimmsten Fall enthält das Dokument nicht genügend Informationen, um nützlichen Text aus dem Dokument extrahieren zu können.

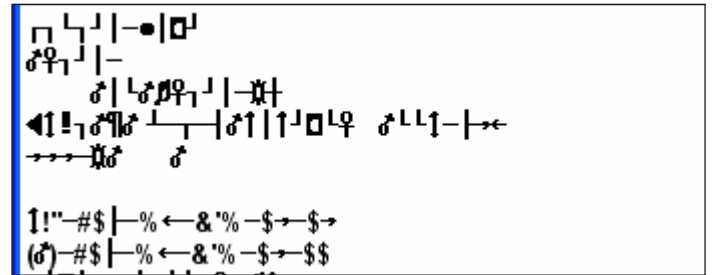
Der patentierte Algorithmus von TET zum Unicode-Mapping implementiert ein kaskadiertes Verfahren, das alle verfügbaren Informationen zur Bestimmung der Unicode-Werte nutzt. Für viele problematische Dokumente kann TET so gültigen Unicode-Text extrahieren, während andere Produkte nur nutzlosen Zeichensalat erkennen.

### Bidirektionaler Text mit Arabisch und Hebräisch

PDF kodiert keinen logischen Text, sondern ist einfach ein Container für die Glyphen auf einer Seite. Arabischer und hebräischer Text läuft von rechts nach links. Diese Texte enthalten aber oft Einschübe von Zahlen oder Namen in westlichen Sprachen, die von links nach rechts zu lesen sind. Aus diesem Grund müssen solche Texte in beiden Leserichtungen interpretiert werden – daher auch der Begriff »bidirektional«. Arabisch bietet noch eine zusätzliche Herausforderung, weil die Zeichen je nach Kontext in bis zu vier verschiedenen Formen benutzt werden können. Diese Kontextformen müssen zu der entsprechenden isolierten Form normalisiert werden.

### Beschädigte PDF-Dokumente

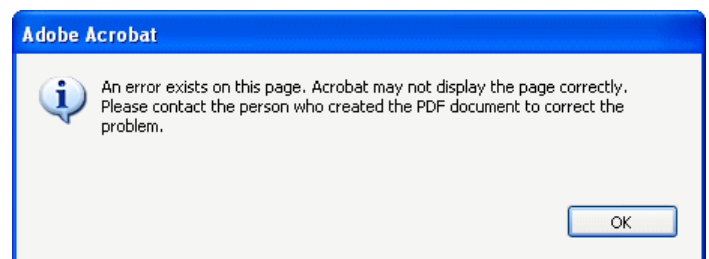
PDF-Dokumente können bei der Übertragung oder aus anderen Gründen beschädigt werden. Der Reparaturmodus von TET stellt viele Arten von beschädigten PDF-Dokumenten wieder her. Manchmal sind die PDF-Daten so stark beschädigt, dass nicht einmal Acrobat sie anzeigen kann. Selbst in solch extremen Fällen extrahiert TET oft noch den Inhalt des Dokuments.



Andere Produkte extrahieren unbrauchbaren Zeichensalat, TET liefert Text.



TET ordnet die Mischung aus links- und rechtsläufigem Text für eine korrekte Textausgabe in die logische Lesereihenfolge um.



Die Seiteninhalte werden nicht einmal in Acrobat angezeigt, TET extrahiert den Text aber trotzdem korrekt.

## Probleme bei der PDF-Bildextraktion

### Farbräume und Kompression

Rasterbilddaten können in PDF mit einer Kombination aus elf Farbräumen und neun Kompressionsfiltern kodiert sein. Übliche Bildformate wie JPEG und TIFF unterstützen jedoch nur einen Teil dieser Möglichkeiten. Die Bild-Engine von TET gleicht die Eigenschaften des PDF-Bildes mit den Möglichkeiten des Bildausgabeformats ab. Unabhängig von der internen Struktur des PDF-Bildes wird das Pixelbild in einem der gängigen Bildformate extrahiert.

### Schmuckfarben

Zusätzlich zu den CMYK-Prozessfarben können Bilder in PDF auch benutzerdefinierte Schmuckfarben verwenden. Diese Farbräume werden auch Separation- (Einkanal-) oder DeviceN- (Mehrkanal-) Farbräume genannt.

TET erzeugt TIFF-Ausgabe mit zusätzlichen Schmuckfarbkanälen. Dies ist für Anwendungen gedacht, die höchste Farbtreue benötigen und keine Farbumwandlung akzeptieren können. Falls ein Bild mit DeviceN-Farbe nur einige der gängigen CMYK-Prozessfarben enthält (z.B. nur Cyan und Magenta), werden die fehlenden Prozesskanäle so hinzugefügt, dass vollständige CMYK-Ausgabe erzeugt werden kann.

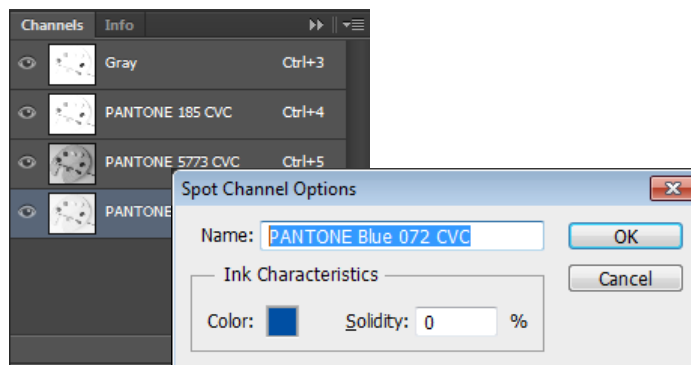
Jedoch können einige Anwendungen Schmuckfarbkanäle eventuell nicht verarbeiten, sondern sind auf einfache TIFF-Ausgabe beschränkt. Um die Verarbeitung zu erleichtern, kann TET in diesem Fall angewiesen werden, einen einzigen Schmuckfarbkanal als Graustufen-TIFF auszugeben.

### Zusammensetzen und Filtern von Bildern

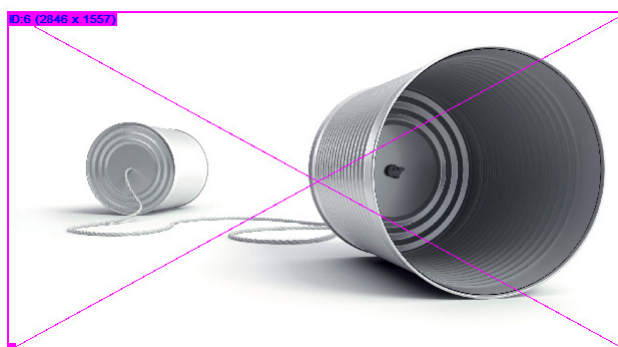
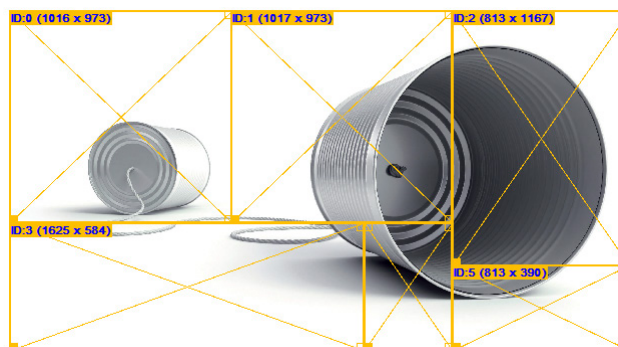
In vielen PDF-Dokumenten werden die Bilder von der Software, die das PDF erzeugt, in kleine Fragmente zerlegt. Was wie ein einziges Bild aussieht, kann durchaus aus zahlreichen kleinen Teilen bestehen. Unter anderem Microsoft Office und TeX produzieren stark fragmentierte Bilder, die aus hunderten oder tausenden kleiner Fragmente bestehen. Adobe InDesign zerlegt Bilder oft in Fragmente von unterschiedlicher Größe. TET erkennt fragmentierte Bilder und setzt die Teile wieder zu einem brauchbaren, größeren Bild zusammen. Denn nur durch das Zusammensetzen und Filtern von TET lassen sich fragmentierte Bilder sinnvoll weiterverwerten.

geräteabhängig	CIE-basiert	spezial
DeviceGray	CalGray	Indexed
DeviceRGB	CalRGB	Pattern
DeviceCMYK	Lab	Separation
	ICCBased	DeviceN

TET verarbeitet alle in PDF verwendeten Farbräume.



Photoshop zeigt die Farbkanäle extrahierter TIFF-Bilder im Optionsdialog für Farbkanäle an (links). Mit einem Doppelklick auf eines der Symbole wird die zugehörige Alternativfarbe angezeigt (rechts).



Obwohl das Bild in kleinere Teile unterteilt ist (oben), extrahiert TET es als ganzes wiederverwendbares Bild (unten).



## Anwendungsmöglichkeiten von TET

TET wird als Programmier-Bibliothek für diverse Entwicklungsumgebungen sowie als Kommandozeilen-Tool für Batch-Prozesse ausgeliefert. Beide Ausführungen bieten den gleichen Funktionsumfang, eignen sich aber für unterschiedliche Einsatzbereiche. Sowohl die TET-Bibliothek als auch das TET-Kommandozeilen-Tool können TETML erzeugen, das XML-basierte Ausgabeformat von TET.

TET unterstützt verschiedene Einsatzvarianten:

- ▶ Die TET-Programmierbibliothek (Komponente) eignet sich zur Integration in Desktop- oder Serveranwendungen. Beispiele für alle unterstützten Sprachbindungen sind im TET-Paket enthalten.
- ▶ Das TET-Kommandozeilen-Tool eignet sich zur Stapelverarbeitung von PDF-Dokumenten. Es erfordert keine Programmierung, sondern kann über Kommandozeilen-Optionen gesteuert und damit in komplexe Abläufe integriert werden.
- ▶ TETML-Ausgabe ist für XML-basierte Workflows und für Entwickler geeignet, die mit dem breiten Spektrum von XML-Werkzeugen und -sprachen wie XSLT vertraut sind.
- ▶ Mit den TET-Konnektoren lässt sich TET in verschiedene gängige Software-Pakete integrieren, wie zum Beispiel Datenbanken und Suchmaschinen.

## Die TET-Produktfamilie

Die TET-Produktfamilie besteht aus den folgenden Produkten:

- ▶ Dem TET-Kernprodukt, das im vorliegenden Datenblatt beschrieben wird.
- ▶ TET PDF IFilter ist als separates Produkt erhältlich. Es eignet sich zur Anwendung mit Microsofts Retrieval-Produkten, wie Windows Search, SharePoint und SQL Server (weitere Informationen finden Sie in einem separaten Datenblatt).
- ▶ Das TET Plugin ist ein kostenloses Plugin für Adobe Acrobat; damit können Sie die hervorragende Text- und Bildextraktion von TET interaktiv testen.

## Unterstützte Entwicklungsumgebungen

PDFlib TET läuft überall – auf praktisch allen Computersystemen. Wir bieten 32- und 64-Bit-Pakete an und unterstützen alle gängigen Varianten von Windows, macOS, Linux und Unix sowie IBM i5/iSeries und zSeries. TET steht außerdem für mobile Plattformen zur Verfügung, darunter iOS und Android.

Der Kern von TET ist in C und C++ programmiert und auf Schnelligkeit und geringen Overhead optimiert. Über ein einfaches API (Application Programming Interface) lässt sich die TET-Funktionalität in zahlreichen Programmiersprachen nutzen:

- ▶ COM für VB, ASP usw.
- ▶ C und C++
- ▶ Java
- ▶ .NET und .NET Core
- ▶ Objective-C (macOS und iOS)
- ▶ Perl
- ▶ PHP
- ▶ Python
- ▶ RPG (IBM i5/iSeries)
- ▶ Ruby
- ▶ Swift

## Vorteile von PDFlib-Software

### Zuverlässig

Weltweit arbeiten viele Tausend Programmierer mit unserer Software. PDFlib erfüllt alle Qualitäts- und Geschwindigkeitskriterien für den Einsatz auf großen Servern. Alle PDFlib-Produkte sind für den zuverlässigen, unbeaufsichtigten 24-Stunden-Betrieb ausgelegt.

### Schnell und einfach

PDFlib-Produkte sind unglaublich schnell – bis zu Tausenden von Seiten pro Sekunde. Die Programmierschnittstelle ist übersichtlich und einfach zu erlernen.

### PDFlib-Produkte sind überall

Unsere Produkte unterstützen alle internationalen Sprachen sowie Unicode. Sie werden von Kunden in der ganzen Welt eingesetzt.

### Professioneller Support

Bei Problemen bietet Ihnen unser Support-Team professionelle Unterstützung. Um den reibungslosen Ablauf unternehmenskritischer Anwendungen zu gewährleisten, können Sie Ihre Software-Lizenz durch einen Supportvertrag ergänzen. Ein Supportvertrag garantiert Ihnen kurze Antwortzeiten und Zugang zu den jeweils neuesten Versionen.

### Lizenzierung

Bei der Lizenzierung können Sie zwischen verschiedenen Modellen für Server-, Integrations-, Firmen- sowie Quellcodelizenzen wählen. Ergänzend bieten wir Supportverträge für umfangreichen technischen Support mit kurzen Reaktionszeiten und kostenlosen Software-Aktualisierungen an.

## Über PDFlib GmbH

PDFlib GmbH ist auf die Entwicklung von PDF-Technologie spezialisiert. PDFlib-Produkte sind seit 1997 weltweit im Einsatz. Das Unternehmen berücksichtigt wichtige technologische Trends, etwa ISO-Standards für PDF. PDFlib GmbH vertreibt alle Produkte weltweit, wobei Nordamerika, Europa und Japan die wichtigsten Märkte darstellen.

## Kontakt

Evaluierungsversionen mit vollem Funktionsumfang und Dokumentation sowie Beispielen sind auf unserer Webseite verfügbar. Weitere Informationen erhalten Sie unter:



### PDFlib GmbH

Franziska-Bilek-Weg 9, 80339 München  
Tel. +49 • 89 • 452 33 84-0  
sales@pdflib.com  
www.pdflib.com