

# Readiris<sup>TM</sup>PRO9

## BENUTZER- HANDBUCH



Readiris Pro

© 2004 I.R.I.S. Alle Rechte vorbehalten

OCR Technologie von I.R.I.S.

Connectionist, AutoFormat und linguistische Technologie von I.R.I.S.

© 2004 I.R.I.S. Alle Rechte vorbehalten

## **ZEIT SPAREN, NEU SCHREIBEN VERMEIDEN!**

---

Herzlichen Glückwunsch zur Anschaffung von Readiris. Diese Software wird Ihnen ohne Zweifel bei der Erfassung von Texten, Tabellen und Grafiken eine große Hilfe sein.

So effizient Computer auch sind, Sie müssen erst Ihre Daten eingeben. Wenn Sie schon einmal einen 15-seitigen Text oder eine lange Tabelle neu geschrieben haben, wissen Sie, wie langweilig und zeitaufwendig dies sein kann. Mit diesem OCR-System auf dem neuesten Stand der Technik, das Ihren Text automatisch in Ihren Computer eingibt, erreichen Sie noch nie zuvor gesehenen Komfort und Effizienz!

Scannen Sie ein gedrucktes oder maschinengeschriebenes Dokument ein, geben Sie die einzulesenden Bereiche an - oder lassen Sie das System diese Bereiche für Sie ermitteln - und führen Sie die Zeichenerkennung durch. Mehrseitige Dokumente werden von Anfang bis Ende in einem einzigen Durchgang verarbeitet. Ein paar Mausklicks statt langer Arbeitsstunden: Readiris konvertiert Ihre gedruckten Dokumente in ein computerlesbares Dateiformat, und dies bis zu 40 mal schneller, als wenn Sie alles neu schreiben müßten.

Der OCR Assistent führt Sie bequem durch den OCR-Prozeß. Damit Sie schnell und bequem die gewünschten Ergebnisse mit Readiris erzielen, brauchen Sie nur einige einfache Fragen zu beantworten. Sie können den erkannten Text sowohl direkt in Ihre Textverarbeitung und in Ihre Spreadsheet senden. Um Faxe zu erkennen und PDF Dateien zu konvertieren, reicht es, die Bilddateien vom Windows Explorer zum Readiris Fenster zu ziehen. Mit einem Rechtsklick mit der Maus können Sie ein Bild sofort zu Readiris schicken.

Readiris ist in der Lage, Tabellen-Data zu erkennen, und sie als Worksheets oder als Tabellen-Objekte in Ihrer Textverarbeitung wiederherzustellen; die numerische Daten sind sofort für weitere Bearbeitung bereit.

Gestützt auf die Connectionist-Technologie von I.R.I.S. bietet Readiris dem Anwender beste OCR-Technik. Readiris kombiniert Omnifont-Erkennung mit selbstlernenden Techniken basierend auf einem neuraler Eigentums-Netzwerk. Readiris ist in der Lage automatisch neue Zeichen zu lernen, indem es den



linguistischen Zusammenhang analysiert: Das linguistische Know-how von Silben und Wörtern führt zu einem Fortschritt der OCR-Leistung.

Readiris kann bis zu 107 Sprachen erkennen: Alle amerikanischen und europäischen Sprachen werden unterstützt, darunter die zentraleuropäischen, baltischen, griechischen, türkischen und kyrillischen („Russischen“) Sprachen. (Optional können auch vier Asiatischen Sprachen - Japanisch, traditionelles und vereinfachtes Chinesisch sowie Koreanisch - verarbeitet werden.) Readiris bewährt sich ebenfalls sehr gut bei gemischten Zeichensätzen: Die Software findet die „westlichen“ Wörter, die häufig in griechischen, kyrillischen und asiatischen Dokumenten erscheinen - viele nicht transkribierbare Eigennamen, Markenbezeichnungen usw. werden in diesen Sprachen mit westlichen Schriftzeichen geschrieben.

Readiris benutzt die Sprachwissenschaft *während* der Erkennungsphase und nicht hinterher. Hieraus resultiert die Tatsache, dass Readiris Dokumente aller Art, auch Dokumente einer schlechten Qualität sowie Faxe und Dokumente, die mit einem Matrixdrucker gedruckt wurden, mit einer großen Genauigkeit liest. Readiris bewährt sich auch bei schlecht gescannten oder kopierten Dokumenten mit zu hellen oder zu dunklen Zeichen. Zusammenlaufende Zeichen („Ligaturen“) werden segmentiert und getrennte Formen wie Nadeldruckzeichen werden wiederhergestellt.

Die Anwendungskontrolle markiert die Zweifelsfälle und erhöht somit die Genauigkeit des Systems. Alle vom Anwender bestätigten Lösungen werden gespeichert, was die Geschwindigkeit und das Vertrauen in das System erhöht. Durch den ständigen Einsatz von Readiris wird das Programm immer intelligenter. Dieses starke Lernmodul ermöglicht es Ihnen, Readiris auf spezielle Zeichen, wie zum Beispiel mathematische Zeichen, Dingbats und nicht zuletzt auf Zeichen mit einer schlechten Form, wie man sie in reellen Dokumenten finden kann, zu trainieren.

Damit das Ergebnis noch besser wird, erkennt Readiris nicht nur Texte, sondern ist auch in der Lage sie zu *formatieren*. Wenn Sie die „Autoformatierung“ benutzen, kann Readiris eine Faxe kopie des gescannten Dokuments herstellen: Die Wort-, Absatz- und Seitenformatierung des Originals werden beibehalten.

Schriftart, Schriftgröße und Schriftauszeichnung vom Originaldokument werden während der Erkennungsphase beibehalten. Die Spalten, Textblöcke und Grafiken befinden sich an der gleichen Stelle wie in den Originaldokumenten. Und da Readiris Graustufen- und Farbscans unterstützt, können Sie beliebige Grafiken einlesen, ob Lineart, Schwarz-weißfotos oder Farbillustrationen. Wenn ein Dokument Tabellen enthält, reorganisiert Readiris diese in Zellen und stellt die Zellenrahmen der Originaltabellen wieder her.

Mit anderen Worten erlaubt es Readiris Ihnen, eine originalgetreue Kopie Ihrer Dokumente zu speichern. Sie ist jetzt aber eine kompakte Textdatei geworden, und nicht langer eine Bilddatei! Es sind mehrere Format-Ebenen vorhanden, aus denen der Benutzer auswählen kann.

Readiris unterstützt eine Vielzahl beliebter Scanner: Zahlreiche Flachbettscanner, Sheetfedscanner und Kombigeräte („MFPs“ oder „Multifunctional Peripherals“) sowie Digitalkameras können benutzt werden. Readiris unterstützt ebenfalls den Twain Scanning-Standard und einige Scan-Plattformen.

## **INHALTSÜBERSICHT**

---

Zeit sparen, Neu schreiben vermeiden! .....	III
Inhaltsübersicht .....	V
Warenzeichen und Urheberrechte .....	VII

### **Kapitel 1: Installation**

Systemvoraussetzungen .....	1-1
Installation von Readiris .....	1-1
Entfernen der Readiris Software .....	1-4
Readiris Entfernungsprogramm .....	1-5
Windows „Installieren/Entfernen“-Assistent .....	1-5
Installierung der Software-Optionen .....	1-6
Installation von Zusatzprogrammen .....	1-8
Installierte Dateien .....	1-10
„Lese Mich“ Datei und Dokumentation .....	1-10
Scanner Treiber .....	1-10
Vergessen Sie nicht registrieren zu lassen! .....	1-10



Produktunterstützung bekommen .....	1-12
-------------------------------------	------

## **Kapitel 2: Kurze Führung**

Starten der Software .....	2-1
Beim ersten Start .....	2-2
Entdecken Sie das Readiris Interface .....	2-3
Beginnen mit einer ersten Einführung .....	2-6
Auf Bilder zoomen .....	2-10
Eins, Zerlegung eines gescannten Bildes .....	2-13
Anderthalb, Fenster sortieren .....	2-15
Zwei, Manuelle Fensterbestimmung eines Bildes .....	2-17
Drei, Fenstervorlagen speichern .....	2-22
Mit Readiris um die Welt! .....	2-23
Bei Bedarf verändert Readiris die Sprachen .....	2-29
Bestimmen der Dokumenteigenschaften .....	2-32
Readiris wird immer intelligenter! .....	2-34
Lernen.....	2-37
Nicht Lernen .....	2-37
Löschen .....	2-38
Rückgängig .....	2-38
Ende .....	2-39
Abbrechen.....	2-39
Die Rolle der Zeichenwörterbücher .....	2-39
Das Ergebnis direkt in Ihre Anwendung senden .....	2-42
Das Ergebnis in einer Textdatei speichern .....	2-46
Portable Dokumente erstellen... ..	2-50
... oder lesen .....	2-55
Mehrere Seiten erkennen .....	2-56
Mehrseitige Dokumente bearbeiten .....	2-65
Ein neues Dokument starten .....	2-67
Textzonen erkennen .....	2-68
Die Aussicht der Textausgabe .....	2-69
Scannereinstellung .....	2-70
Bringen Sie Farbe in Ihre Text-Scans! .....	2-73
Andere Geräte, andere Auflösung .....	2-76
Standard Einstellungen speichern .....	2-80
Vordefinierte Einstellungen speichern .....	2-81

Dokumente scannen .....	2-82
Justieren der gescannten Bilder .....	2-85
Lassen Sie den OCR Assistenten für Sich arbeiten .....	2-89
Readiris erstellt das Layout .....	2-91
Spalten bitte, keine Rahmen! .....	2-95
Textformatierung, Teil 2 .....	2-97
Text mehrmals exportieren .....	2-98
Grafiken getrennt speichern .....	2-99
Faxe erkennen und zurückgestellte Erkennung .....	2-101
Tabellenerkennung .....	2-103
Online-Hilfe bekommen .....	2-108

## **WARENZEICHEN UND URHEBERRECHTE**

---

Die Readiris Software wurde von I.R.I.S. entworfen und entwickelt. OCR, Connectionist, AutoFormat und linguistische Technologie von I.R.I.S. Die Readiris-Software, die OCR-Technologie, die linguistische Technologie, das Online-Hilfesystem und dieses Handbuch sind urheberrechtlich von I.R.I.S. geschützt.

AutoFormat, Cardiris, Connectionist, die I.R.I.S. linguistische Technologie, das I.R.I.S. Logo und Readiris sind Warenzeichen von I.R.I.S.

Von Apache entwickelter XML-Parser. Dieses Produkt beinhaltet Software, die von der Apache Software Foundation ([www.apache.org](http://www.apache.org)) entwickelt wurde.

Acrobat und Reader sind (eingetragene) Warenzeichen von Adobe. AsianBridge ist ein Warenzeichen von TwinBridge. AsianSuite ist ein Warenzeichen von UnionWay. Excel, Windows und Word sind eingetragene Warenzeichen von Microsoft. Intel ist ein eingetragenes Warenzeichen von Intel.





# Kapitel 1

## INSTALLATION

In diesem Kapitel werden die Systemvoraussetzungen und die Installation von der Readiris Software besprochen.

### SYSTEMVORAUSSETZUNGEN

---

Die folgenden Grundbedingungen sind für die Installation von Readiris erforderlich:

- ☐ ein 486 Intel PC oder kompatibel. Ein Pentium PC wird empfohlen.
- ☐ 32 MB RAM. 64 MB RAM ist zur Verarbeitung von Graustufen- und Farbbildern empfohlen.
- ☐ 110 MB freier Festplattenspeicherplatz. 95 MB reichen aus, wenn die Beispieldateien nicht von CD-ROM kopiert werden.
- ☐ das Betriebssystem Windows XP, Windows 2000, Windows 98 oder Windows NT 4.0.

Bestimmte **Scannertreiber** funktionieren gegebenenfalls nicht unter den neusten Versionen von Windows. Schlagen Sie bitte in der Dokumentation Ihres Scanners nach, welche Plattformen unterstützt werden.

### INSTALLATION VON READIRIS

---

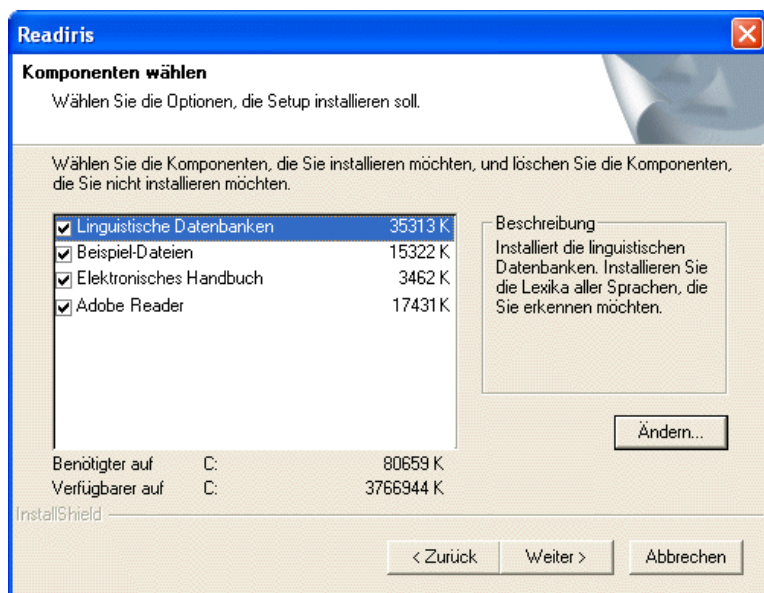
Die Readiris Software ist ausschließlich auf einer **selbststartenden CD-ROM** lieferbar. Um das Programm zu installieren, legen Sie die CD-ROM in das CD-ROM-Laufwerk ein und warten Sie, bis sie aktiviert wird. Folgen Sie den Anweisungen auf dem Bildschirm.



Wenn die CD-ROM nicht automatisch aktiviert wird, nachdem sie in das CD-ROM-Laufwerk gelegt wurde, aktivieren Sie den Befehl MENU.EXE, um die Software zu installieren.

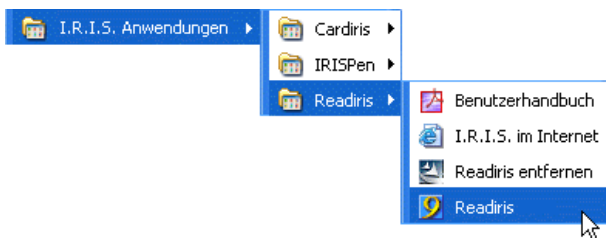
Benutzer von Windows XP, Windows 2000 und Windows NT müssen sicherstellen, dass Sie die erforderlichen **Zugriffsrechte** haben. Wenden Sie sich bei Bedarf an Ihren Systemadministrator.

Es werden mehrere Installationsoptionen angeboten. Sie sollten die **linguistischen Datenbanken** für alle Sprachen, mit denen Sie zu arbeiten beabsichtigen, installieren. Standardmäßig werden *alle* Wörterbücher installiert. Es ist ratsam, die **Beispielbilder** zu installieren, die in den Lernabschnitten dieses Handbuchs verwendet werden.



Wenn erforderlich, sollte ebenfalls Adobe Reader, zum Lesen der Softwaredokumentation benötigt, installiert werden. Das **elektronische Handbuch** wird standardmäßig auf Ihre Festplatte kopiert. Sie können es aber auch auf der CD-ROM lassen.

Das Untermenü „I.R.I.S. Anwendungen - Readiris“ des Menüs „Programme“ wird automatisch vom Installationsprogramm erstellt.



Außerdem legt das Installationsprogramm eine Readiris **Verknüpfung** auf dem Windows Schreibtisch an, mit dem Sie Readiris direkt vom Schreibtisch aus starten können.



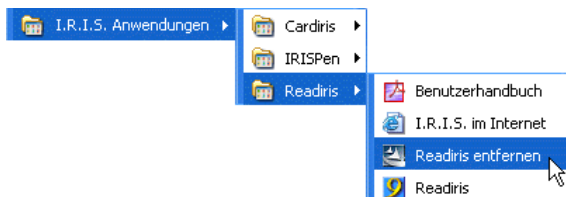
## ENTFERNEN DER READIRIS SOFTWARE

---

Es gibt nur zwei Möglichkeiten, um die Readiris Software zu entfernen: durch das Programm „Readiris Entfernen“ und das Windows Entfernungs-Assistent. Es wird davor gewarnt, das Readiris Programm durch Löschen einzelner Softwarekomponenten zu entfernen.

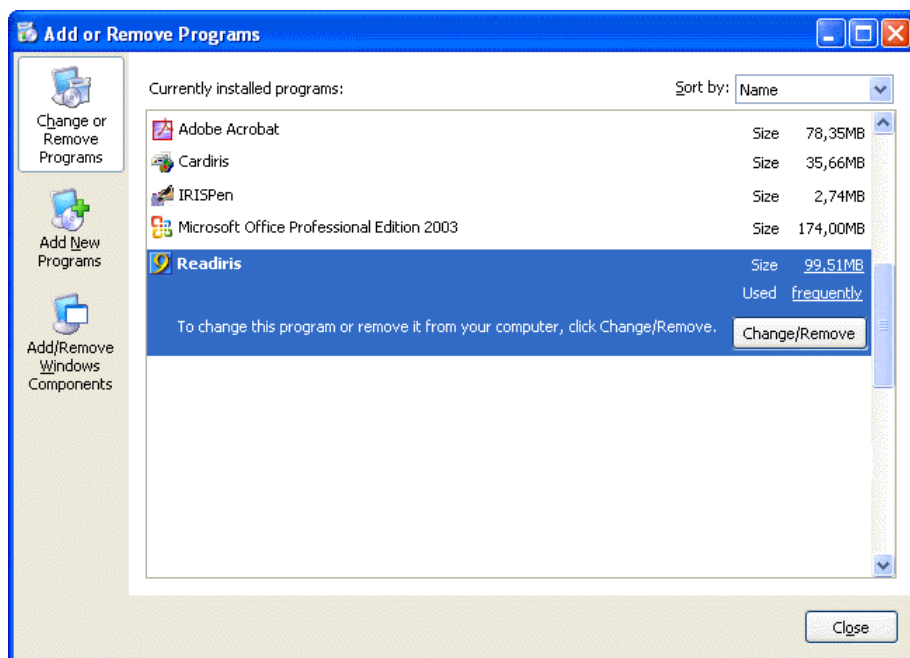
## **Readiris Entfernungsprogramm**

Um das Readiris Entfernungsprogramm zu starten, selektieren Sie „Readiris Entfernen“ in das Untermenü „I.R.I.S. Anwendungen - Readiris“ und folgen Sie den Anweisungen auf Ihrem Bildschirm.



## **Windows „Installieren/Entfernen“-Assistent**

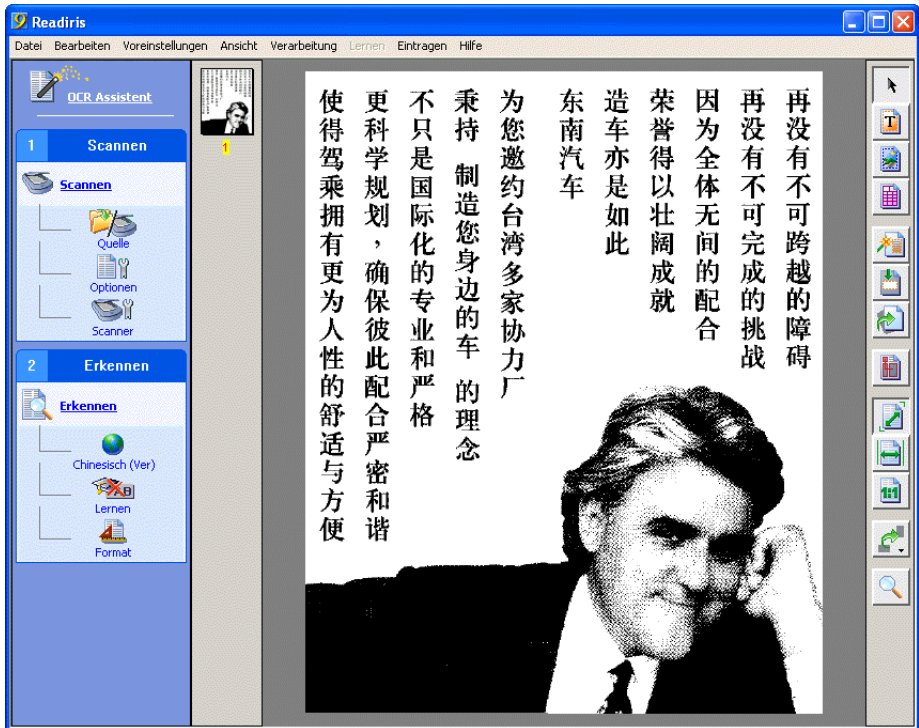
- ☐ Klicken Sie auf „Einstellungen“ unter dem Windows Menü „Start“ und gehen Sie zur „Systemsteuerung“.
- ☐ Klicken Sie auf „Software“.



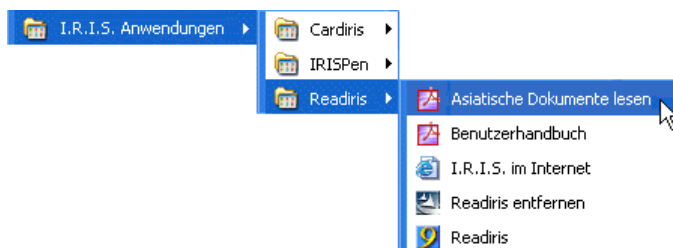
- ☐ Folgen Sie den Anweisungen auf dem Bildschirm um die Readiris Software zu entfernen.

## INSTALLIERUNG DER SOFTWARE-OPTIONEN

Es ist nur eine einzige Software-Option für Readiris verfügbar: Das „Asiatisches OCR-Modul“, mit dem es möglich ist Japanisch, traditionelles und vereinfachtes Chinesisch und Koreanisch einzulesen. Dieses Softwaremodul wird ebenfalls auf einer selbststartenden CD-ROM geliefert.



Diese Option installiert die Dokumentation zur Erkennung von Dokumenten in asiatischen Schriften.



## INSTALLATION VON ZUSATZPROGRAMMEN

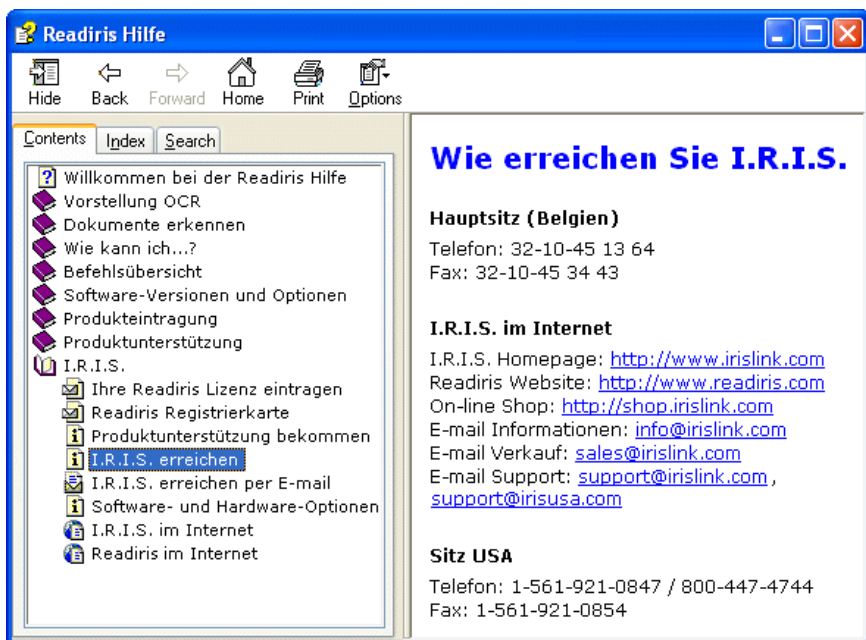
---

Je nach Software-Paket kann Readiris mit einer Version des Zusatzprogrammes Cardiris, einem **Karteiprogramm**, geliefert werden.

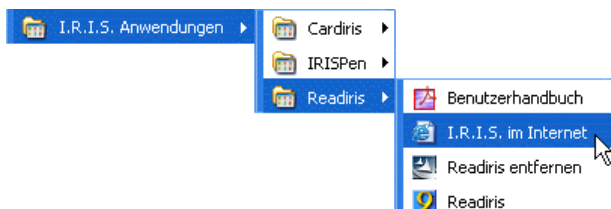
Wenn dieses Zusatzprogramm sich auf Ihrer CD-ROM befindet, wird es auch automatisch beim Aktivieren der CD-ROM installiert. Sie brauchen nur den Anweisungen auf dem Bildschirm zu folgen.

Um mehr über zusätzliche Software zu erfahren, können Sie sich jederzeit mit I.R.I.S. in Verbindung setzen; der Befehl „I.R.I.S. Erreichen“ befindet sich im Readiris „Hilfe“ Menü und erklärt Ihnen, wie Sie sich mit I.R.I.S. in Verbindung setzen können.





Ein Anwendungssymbol im Untermenü „I.R.I.S. Anwendungen - Readiris“ des „Programme“ Menüs bringt Sie sofort zur I.R.I.S. **Homepage**. Das Gleiche gilt für dem Startschirm und das Kommando „I.R.I.S. im Internet“ im „Hilfe“ Menü von Readiris.





## **INSTALLIERTE DATEIEN**

---

Das Installationsprogramm hat eine Mappe erstellt, in dem sich die Readiris Dateien befinden. Es wird davor gewarnt, das Readiris Programm oder einzelne Softwarekomponenten manuell zu entfernen. Verwenden Sie hierzu das Readiris Entfernungsprogramm oder den Windows Entfernungsassistenten. Siehe oben.

### **„Lese Mich“ Datei und Dokumentation**

README.HTM      „Lese Mich“ Datei (im HTML Format)

MANUAL.PDF      Benutzerhandbuch (im Adobe Acrobat Format)

### **Scanner Treiber**

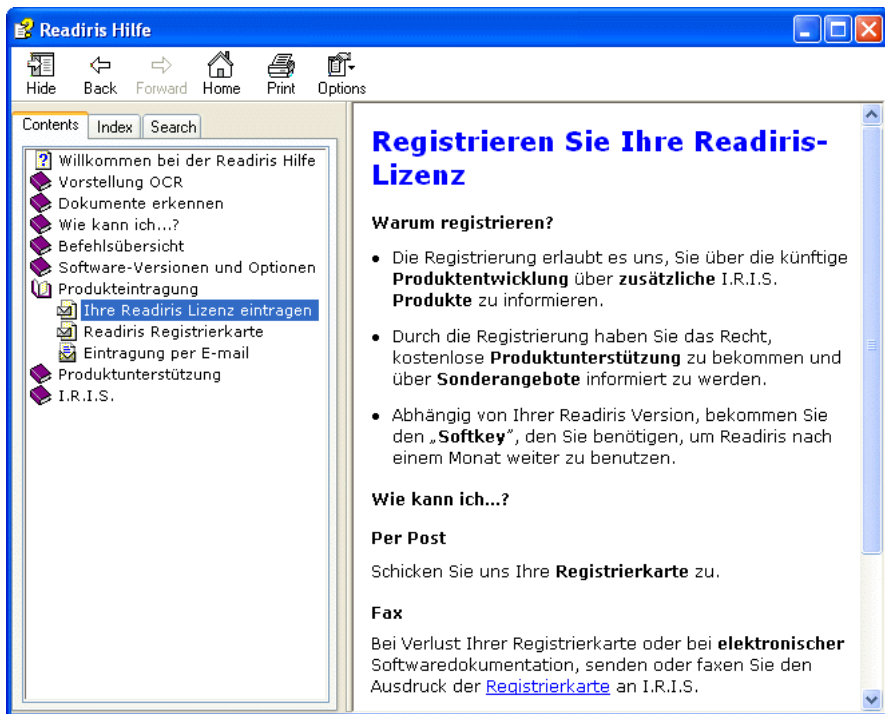
Zögern Sie nicht, Ihren Scanner-Fabrikanten oder seinen Vertreiber zu kontaktieren, falls Probleme mit Scanner-Treibern auftreten. Bei den meisten Fabrikanten können Sie die neuesten Versionen der Scanner-Treiber von deren Website herunterladen.

## **VERGESSEN SIE NICHT REGISTRIEREN ZU LASSEN!**

---

Vergessen Sie nicht Ihre Readiris Lizenz registrieren zu lassen! Die Registrierung erlaubt es uns, Sie über die weitere Produktentwicklung sowie über I.R.I.S. Zusatzprogramme zu informieren. Die Vorteile, sowie die kostenlose **Produktunterstützung** und **Sonderangebote**, erhalten nur registrierte Anwender.

Es gibt mehrere Registriermöglichkeiten: Sie können uns Ihre Registrierkarte zusenden oder zufaxen, sich während der Bürozeiten telefonisch registrieren lassen, oder die Registrierung auf der I.R.I.S. Website ausfüllen!



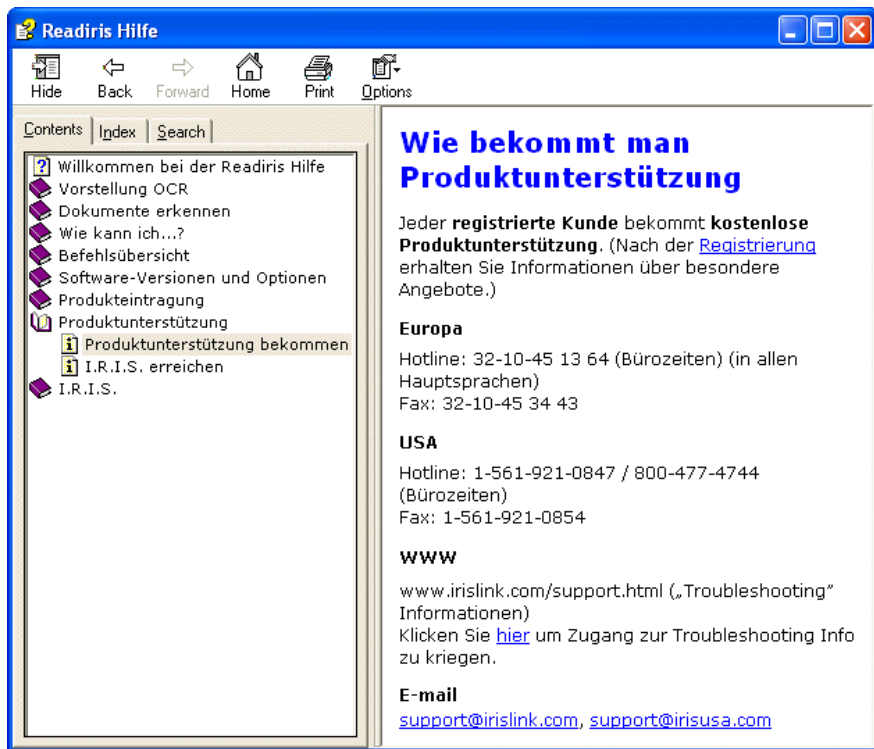
Der Readiris **Eintragungsassistent**, den Sie im Menü „Eintragen“ in der Readiris Software finden, führt Sie auf einfache Art durch den Eintragungsprozeß.



Abhängig von Ihrer Readiris-Version, erhalten Sie den „**Softkey**“, den Sie brauchen können um Readiris nach einem Monat weiter zu benutzen.

## PRODUKTUNTERSTÜTZUNG BEKOMMEN

Über den Befehl „Produktunterstützung“ im Menü „Hilfe“ von Readiris erfahren Sie genau, wie Sie technische Unterstützung bekommen können. Geben Sie eine deutliche Beschreibung des Problems, einschließlich der damit verbundenen Readiris Daten, Ihres Scanners und Ihres Computers.





# Kapitel 2

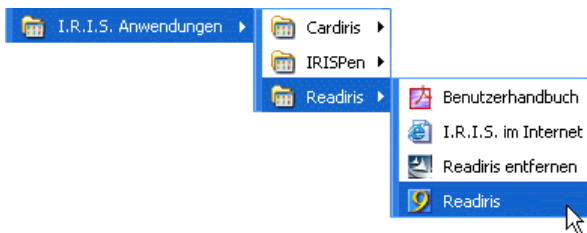
## KURZE FÜHRUNG

Readiris ist ein OCR-System auf dem neuesten Stand der Technik, das mit einer Vielzahl von fortgeschrittenen Eigenschaften ausgerüstet ist. Dieses Kapitel befaßt sich mit den wichtigsten Eigenschaften und gibt Ihnen viele Tips und Anweisungen bezüglich der Anwendung von Readiris.

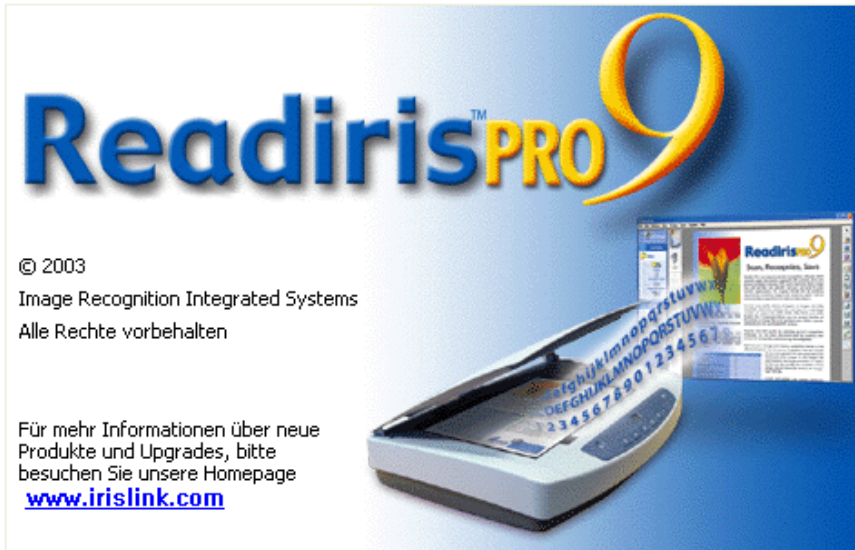
### STARTEN DER SOFTWARE

---

Klicken Sie die Readiris Anwendung im Submenü „I.R.I.S. Anwendungen - Readiris“ oder klicken Sie die Abkürzung in dem Schreibtisch.



Der Readiris Startbildschirm sowie das Anwendungsfenster erscheinen. Im Startbildschirm sind die Versionsnummer und Urheberrechtsvermerke für die Readiris Software aufgeführt. Außerdem können Sie von hier aus direkt zur I.R.I.S. **Homepage** gelangen - klicken Sie auf die URL, um die Website von I.R.I.S. zu öffnen. Mit einem Mausklick außerhalb diese URL verschwindet dieser Startbildschirm.



Das nächste Fenster bezieht sich auf den OCR Assistenten. Klicken Sie im Moment auf „Abbrechen“.

## BEIM ERSTEN START

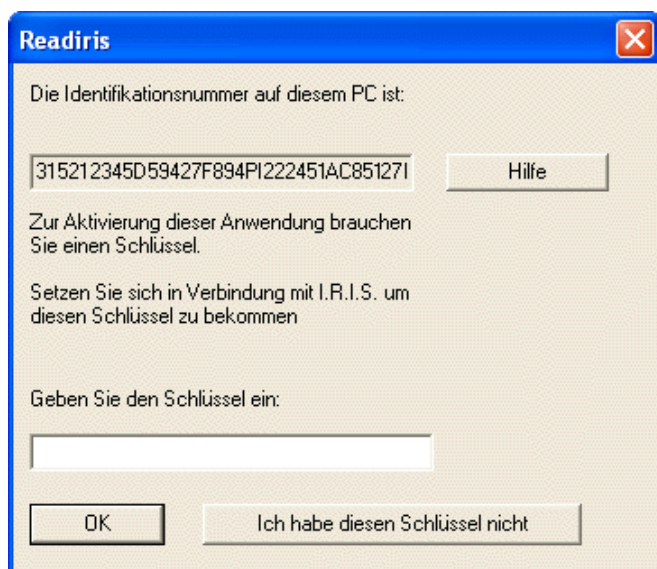
---

Je nach Software-Paket werden Sie beim ersten Start gebeten, Ihre Lizenz zu registrieren.

In diesem Fall ist der Gebrauch von Readiris auf einem Monat beschränkt. Bei Registrierung bekommen Sie einen **Softkey** gratis um die Software nach 30 Tagen weiter zu benutzen.

Um diesen „Softkey“ zu generieren, benötigen Sie unbedingt die **Identifikationsnummer**; bitte vergewissern Sie sich, dass diese Nummer verfügbar oder notiert ist, wenn Sie Ihre Lizenz registrieren.

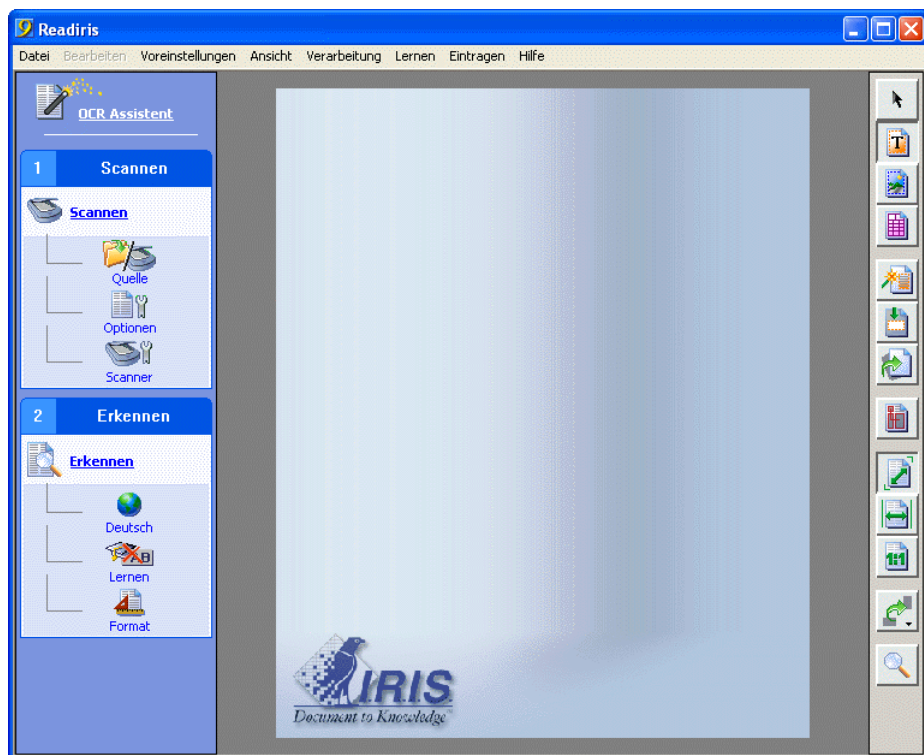




## ENTDECKEN SIE DAS READIRIS INTERFACE

---

Das Readiris Anwendungsfenster enthält nicht nur **Befehlsmenüs**, sondern auch zwei Toolbars, die Ihnen den schnellen Zugriff zu den meisten Befehlen ermöglichen. Anfänglich sind die Menüs in Bezug auf der Vorschau ausgeblendet: Solange noch keine Bilddateien geöffnet wurden, bleiben sie ausgeblendet.



Das gleiche gilt für die **Bild-Toolbar**, die sich rechts in Ihrem Anwendungsfenster befindet. Die Bild-Toolbar enthält alle für die Bildvorschau notwendigen Befehle. Die **Haupttoolbar** links ermöglicht schnellen Zugang zu allen häufig genutzten allgemeinen Befehlen.

Damit Sie wissen, welcher Befehl einer bestimmten Schaltfläche entspricht, brauchen Sie nur den Cursor kurz über die Schaltfläche zu bewegen: Das **QuickInfo** teilt Ihnen die Funktion der Schaltfläche mit.



Die gescannten Bilder werden im Fenster oder **Bildbereich** angezeigt. Zur Erkennung der Bilddateien können diese in den Bildbereich (oder auf das Readiris Anwendungssymbol) gezogen werden („Drag & Drop“).

Sobald Seiten verarbeitet werden, erscheint eine weitere Toolbar, die **Seiten-Toolbar**, an der linken Seite. In dieser Toolbar sind die verschiedenen Seiten des Dokuments dargestellt. Sie können auf die Seiten-Befehle zugreifen, indem Sie mit der rechten Maustaste klicken („Kontext“ Menü).





## BEGINNEN MIT EINER ERSTEN EINFÜHRUNG

---

Die beste und einfachste Art, sich mit Readiris vertraut zu machen, ist es zu benutzen. Zu Ihrer Software gehört eine Zahl von **vorgescannten Bildern**, die Ihnen den Start auch ohne Scanner ermöglichen. Diese Bilder schauen wir uns an.

Die Schaltfläche „Quelle“ auf der Haupttoolbar bestimmt, ob Sie den Scanner oder vorgescannte Bilder als Bildquelle benutzen werden.

Farb-, Graustufen- und Schwarzweißbilder werden gleichermaßen unterstützt. Mit Readiris können Sie Adobe Acrobat PDF Dokumente, JPEG Bilder, Paintbrush Bilder (PCX), DCX Faxbilder (eine mehrseitige Version des Paintbrush Formats), PNG Bilder, TIFF Bilder (nicht komprimiert, LZW, PackBits, Gruppe 3 und Gruppe 4 komprimiert), mehrseitige TIFF Bilder und Windows Bitmaps (BMP) öffnen.

Diese Funktion ist sehr praktisch, wenn Sie Ihre **Faxe** zur weiteren Bearbeitung in Textdateien konvertieren möchten.

Wenn Sie ein vorgescanntes Bild als Bildquelle öffnen möchten, sollten Sie mit der „Quelle“ Schaltfläche die „Bilddateien“ (und nicht den Scanner) als Bildquelle wählen.



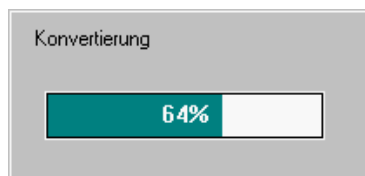
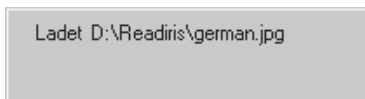
Klicken Sie anschließend auf die Schaltfläche „Öffnen“. (Wenn Sie die Festplatte als Bildquelle wählen, wird „Scannen“ durch „Öffnen“ ersetzt, und der entsprechende Befehl „Scannen“ im Menü „Verarbeitung“ wird wiederum durch den Befehl „Öffnen“ ersetzt.)



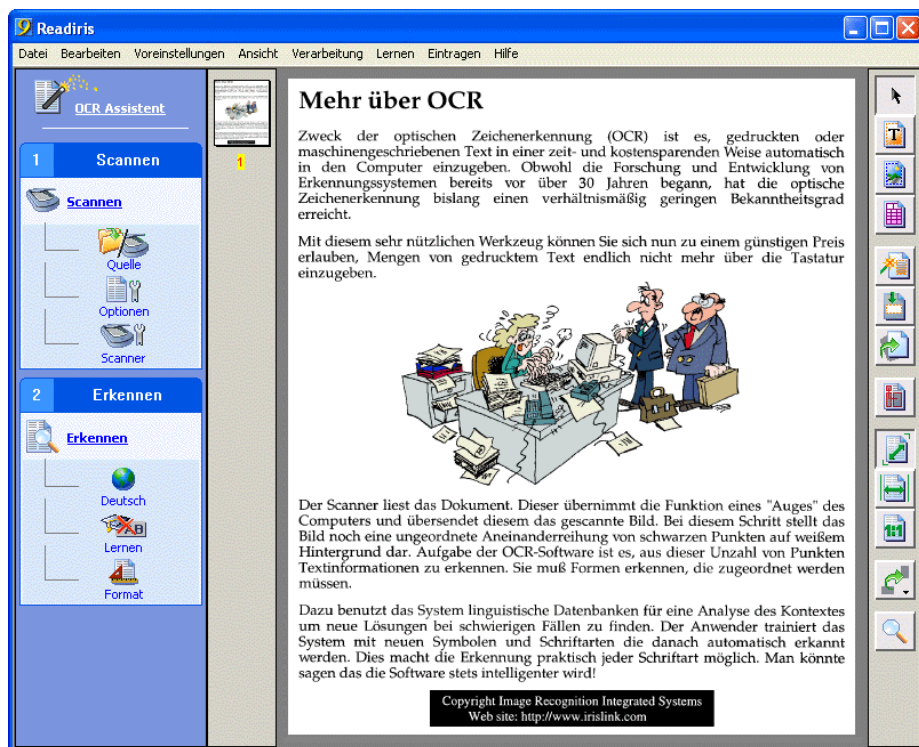
Um das vorgescannte Bild direkt zu öffnen, können Sie auch den Befehl „Öffnen“ im „Datei“ Menü selektieren. Diese Funktion arbeitet sogar wenn Ihr Scanner als Bildquelle läuft.



Sie werden aufgefordert, eine Bilddatei auszuwählen. Wählen Sie die Datei GERMAN.JPG im Readiris Ordner. Da dieses Bild ein Farbbild ist, wird es nicht nur von Festplatte eingelesen, sondern es wird auch eine Schwarzweiß-Fassung für den OCR-Vorgang erstellt.



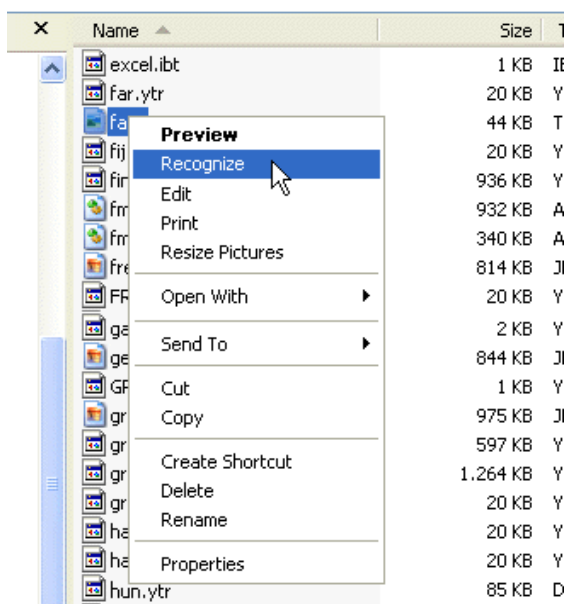
Dieses Bild wird schließlich im Bildbereich angezeigt. In der Seiten-Toolbar sieht man, dass eine einzige Seite in Readiris eingelesen wurde.



Eine dritte Möglichkeit ist, vorgescannte Bilder mit Hilfe der „**Drag & Drop**“-Funktion zu öffnen. Ziehen Sie Bilder vom Windows Explorer in den Readiris Bildbereich oder auf das Readiris Anwendungssymbol, und schon werden die Bilder geöffnet.



Sie können Bilder ebenfalls vom Windows Explorer aus öffnen: **Klicken Sie mit der rechten Maustaste** auf eine Bilddatei und wählen Sie aus dem „Kontext“ Menü den Befehl „Erkennen“. (Dieser Befehl erscheint nur, wenn der Dateityp der ausgewählten Datei unterstützt wird.)



Hierdurch wird jedoch nicht direkt der OCR-Vorgang gestartet. Um dem Benutzer vollkommene Handlungsfreiheit zu lassen, wird lediglich Readiris gestartet und das ausgewählte Bild geöffnet.

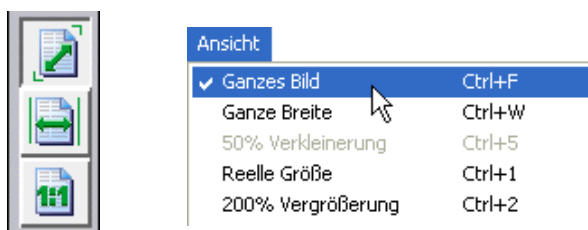
Die Bild-Toolbar rechts im Readiris Anwendungsfenster enthält alle notwendigen Befehle für die Bildvorschau: Funktionen zur Angabe der zu bearbeitenden Blöcke, Funktionen zum Drehen des Bildes, zum Ein- und Auszoomen usw.

## AUF BILDER ZOOMEN

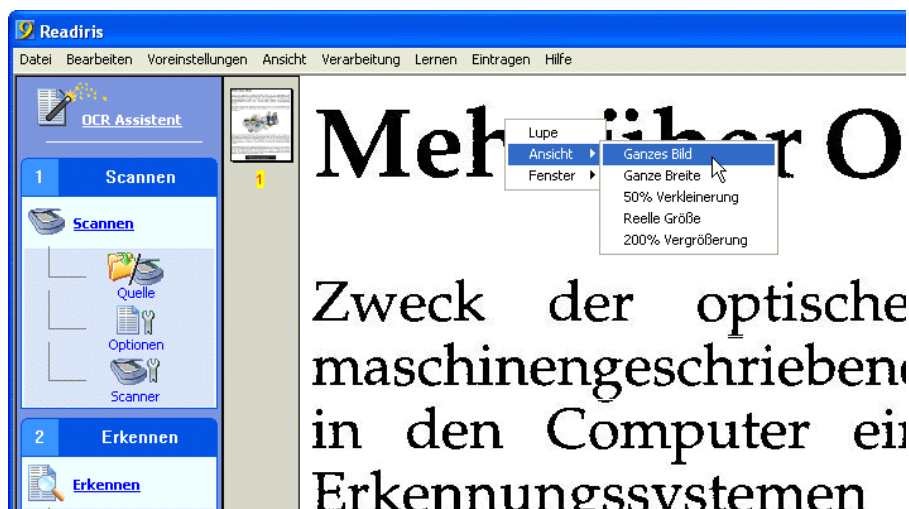
Readiris verfügt über mehrere Befehle um das gescannte Bild zu **zoomen**, zum Beispiel um die Scanqualität zu überprüfen.



Auf der Bild-Toolbar befinden sich Schaltflächen, mit denen Sie das gescannte Bild auf die wahre Größe bzw. auf Seitenbreite oder auf Bildschirmgröße einzoomen können. Das Menü „Ansicht“ enthält noch zwei zusätzliche Zoomstufen: Sie können das Bild auf 50% oder 200% der Originalgröße zoomen. Wenn das Bild auf seine tatsächlichen Größe gezoomt wird, stimmt ein Bildschirmpixel überein mit einem Bildpixel. (Es sind Tastenkombinationen für alle Zoomstufen vorhanden!)



Die Zoom-Stufen können mit einem rechten Mausklick eingestellt werden. Klicken Sie die rechte Maustaste, um das „Kontext“ Menü zu öffnen, und wählen Sie die geeignete Zoomstufe.

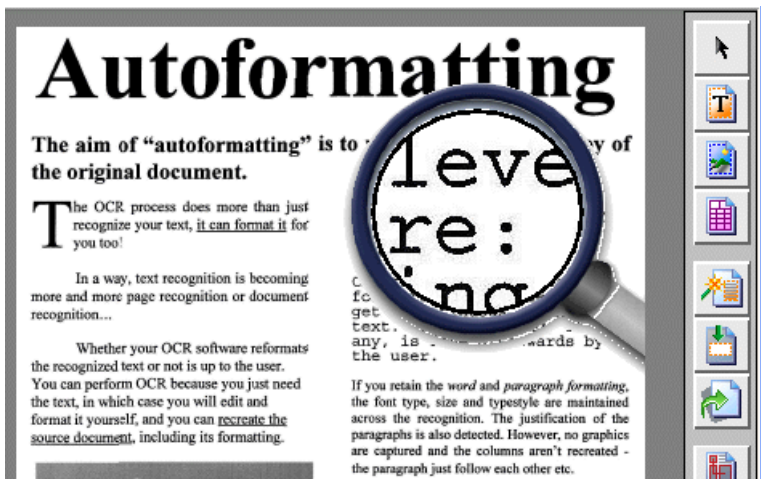


Außerdem können Sie mit der rechten Maustaste auf einen Bereich des gescannten Bildes *doppelklicken*, um automatisch auf 100%-Zoom zu schalten. Wiederholen Sie die Operation, um den Zoom aufzuheben.

Schließlich können Sie die **Lupe** verwenden, um Details des gescannten Dokuments heranzuzoomen. Die Lupe ist ebenfalls im „Kontext“ Menü verfügbar, wenn Sie mit der rechten Maustaste auf das Bild klicken.



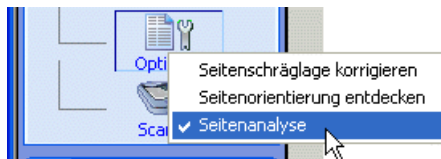
Mehr über O



## EINS, ZERLEGUNG EINES GESCANNTEN BILDES

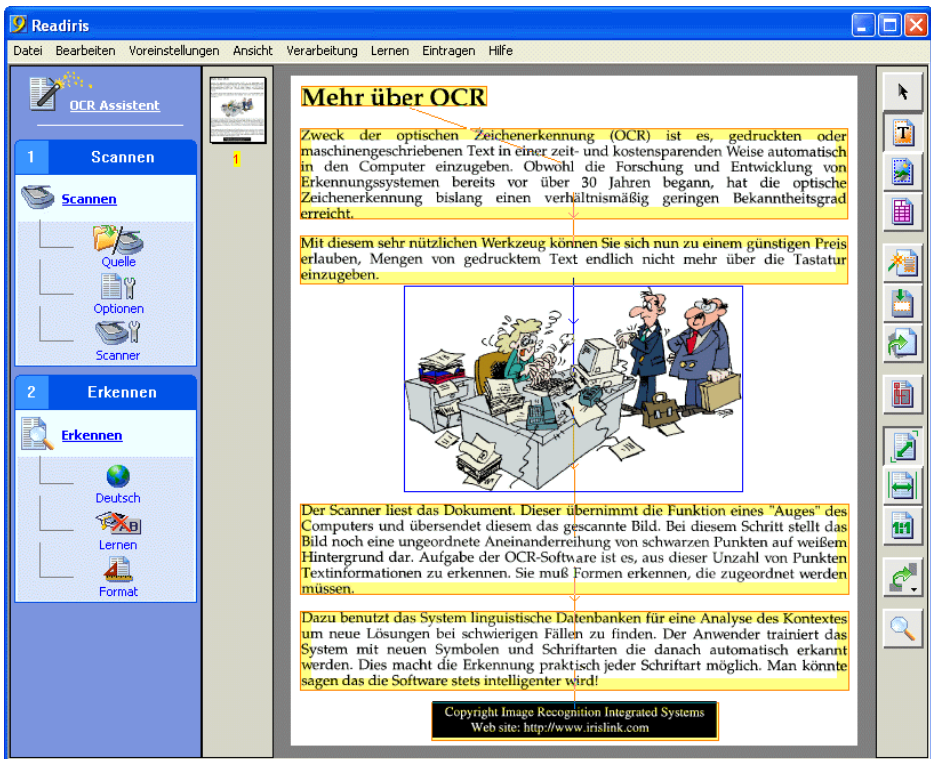
Nachdem das Bild gescannt wurde, muß bestimmt werden, welche Teile als editierbare Text verarbeitet werden müssen. Hierzu werden Rahmen, sogenannte „Fenster“, um die ausgewählten Bereiche gezogen.

Wenn die Option „Seitenanalyse“ mithilfe der Schaltfläche „Optionen“ in der Haupttoolbar (im Menü „Voreinstellungen“) aktiviert wurde, führt Readiris diese Operation automatisch durch.





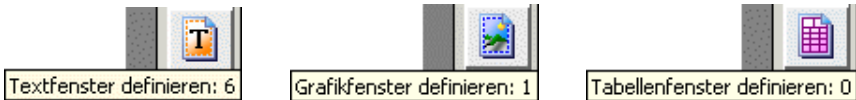
Die automatische Seitenanalyse ist besonders nützlich, wenn **Texte in Spaltenform** und Dokumente mit einem komplizierten Seitenlayout (die vielleicht Grafiken oder Tabellen enthalten) eingelesen werden müssen.



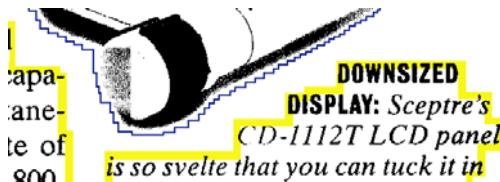
Die Seitenanalyse verwendet drei **Arten von Fenstern**: Text-, Grafik- und Tabellenfenster. Readiris unterscheidet zwischen Textblöcken, Tabellen und Grafikzonen (die Fotos, Illustrationen usw. enthalten) auf einer Seite. (Wie man Grafiken speichert und Tabellen erkennt, wird weiter unten noch detailliert erklärt.)

Ein **Farbcode** gibt den Fenstertyp an: Textzonen sind gelb, Tabellen violett und Grafikfenster blau umrandet.

Die Zahl der Fenster wird jederzeit in dem QuickInfo der „Textfenster“, „Grafikfenster“ und „Tabellenfenster“ Schaltflächen angezeigt.



Die Seitenanalyse ist schnell, toleriert Schrägstellung und arbeitet sehr präzise: „unregelmäßige“ Schatten werden eingegrenzt.



Die Seitenanalyse ermittelt sogar Zonen mit **weißem Text auf schwarzem Hintergrund**. Solche Eingaben sind kein Problem bei der Erkennung: Obwohl die Vorschau das gescannte Dokument originalgetreu auf dem Bildschirm anzeigt, ist Readiris in der Lage das Bild bei Bedarf zu „invertieren“, damit solche Textblöcke erkannt werden. (Ebenfalls können Sie mit Ihrem Scanner *ganze* Negativbilder erzeugen um weißen Text auf einem schwarzen Hintergrund zu verarbeiten. Siehe unten.)

## ANDERTHALB, FENSTER SORTIEREN

Readiris erkennt nicht nur die verschiedenen Blöcke, sondern *sortiert* sie selbst: Die Zonen werden von oben nach unten, von links nach rechts sortiert, damit man auch mit Spalten umgehen kann.

Sie können die **Sortierreihenfolge** natürlich auch ändern. Klicken Sie hierzu auf die Schaltfläche „Sortieren“ in der Bild-Toolbar. Der Cursor wird zum Zeigefinger, wenn der Sortiermodus aktiviert ist.



Danach klicken Sie die Fenster, die Sie einschliessen wollen, an. Fenster die *nicht* angeklickt werden, werden einfach ignoriert, von der Erkennung ausgeschlossen. Es ist weiters sehr einfach zu sehen, welche Fenster angewählt wurden und welche nicht: Die angewählten Fenster haben eine starke Farbe, nicht angewählte Fenster haben eine schwächere.

**Readiris**

Datei Bearbeiten Voreinstellungen Ansicht Verarbeitung Lernen Eintragen Hilfe

**OCR Assistent**

**1 Scannen**

Scannen

Quelle

Optionen

Scanner

**2 Erkennen**

Erkennen

Deutsch


Lernen

Format

### Mehr über OCR

Zweck der optischen Zeichenerkennung (OCR) ist es, gedruckten oder maschinengeschriebenen Text in einer zeit- und kostensparenden Weise automatisch in den Computer einzugeben. Obwohl die Forschung und Entwicklung von Erkennungssystemen bereits vor über 30 Jahren begann, hat die optische Zeichenerkennung bislang einen verhältnismäßig geringen Bekanntheitsgrad erreicht.

Mit diesem sehr nützlichen Werkzeug können Sie sich nun zu einem günstigen Preis erlauben, Mengen von gedrucktem Text endlich nicht mehr über die Tastatur einzugeben.



Der Scanner liest das Dokument. Dieser übernimmt die Funktion eines "Auges" des Computers und übersendet diesem das gescannte Bild. Bei diesem Schritt stellt das Bild noch eine ungeordnete Aneinanderreihung von schwarzen Punkten auf weißem Hintergrund dar. Aufgabe der OCR-Software ist es, aus dieser Unzahl von Punkten Textinformationen zu erkennen. Sie muß Formen erkennen, die zugeordnet werden müssen.

Dazu benutzt das System linguistische Datenbanken für eine Analyse des Kontextes um neue Lösungen bei schwierigen Fällen zu finden. Der Anwender trainiert das System mit neuen Symbolen und Schriftarten die danach automatisch erkannt werden. Dies macht die Erkennung praktisch jeder Schriftart möglich. Man könnte sagen das die Software stets intelligenter wird!

Copyright Image Recognition Integrated Systems  
Web site: <http://www.irislink.com>

Die Seitenanalyse wird als Standardeinstellung eingestellt. Wenn die Seitenanalyse inaktiviert wurde weil Sie zum Beispiel versehentlich einige Fenster gelöscht haben, können Sie wieder aktiviert werden, indem Sie die Schaltfläche „Seite Analysieren“ in der Bild-Toolbar anklicken.



Wenn Sie asiatische Dokumente bearbeiten, wählen Sie die Dokumentensprache, *bevor* Sie die Seitenanalyse ausführen. Für diese Sprachen gelten ganz spezifische Handlungsweisen: Im allgemeinen ist der Zeilenabstand in asiatischen Dokumenten größer als in westlichen Dokumenten, der Text besteht aus kleinen Ikonen („Ideogrammen“), die in westlichen Dokumenten mit Grafiken verwechselt werden können, und der Text kann von oben nach unten und von rechts nach links gelesen werden. Und falls Sie vergessen haben, die korrekte Sprache zu wählen, können Sie dies nachträglich tun. Readiris führt die Analyse der Seite automatisch neu durch!

Manche Dokumente weisen zahlreichen „Streupunkte“ auf der Seite auf, das eigentliche Bild ist ggf. mit einem schwarzen Seitenrand umgeben usw. Um alle kleinen Fenster zu entfernen - wobei davon ausgegangen wird, dass sie keinen wirklichen Text enthalten - und das verbleibende Fenster neu anzuordnen, können Sie im Menü „Bearbeiten“ den Befehl „Kleine Fenster Löschen“ wählen.

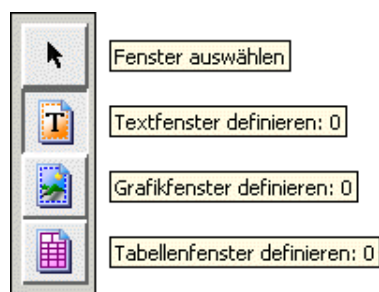
Kleine Fenster löschen

Ctrl+M



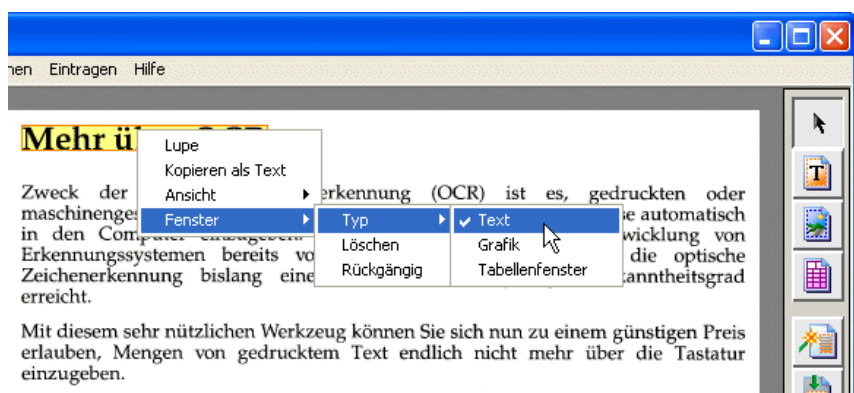
## **ZWEI, MANUELLE FENSTERBESTIMMUNG EINES BILDES**

Durch die Seitenanalyse werden die Fenster automatisch bestimmt. Alternativ dazu kann die Fensterbestimmung auch manuell durchgeführt werden mit Hilfe der **Fenster-Tools** von Readiris.



Um ein rechteckiges Fenster um eine gewünschte Zone zu **zeichnen**, wählen Sie die entsprechende Schaltfläche in der Bild-Toolbar, bewegen Sie den Cursor in die obere linke Ecke und ziehen Sie das Fenster durch bewegen des Cursors in die untere rechte Ecke und klicken Sie nochmals. (Seiten von weniger als 1 mm. sind nicht erlaubt: Sie würden ohnehin kein Zeichen enthalten.)

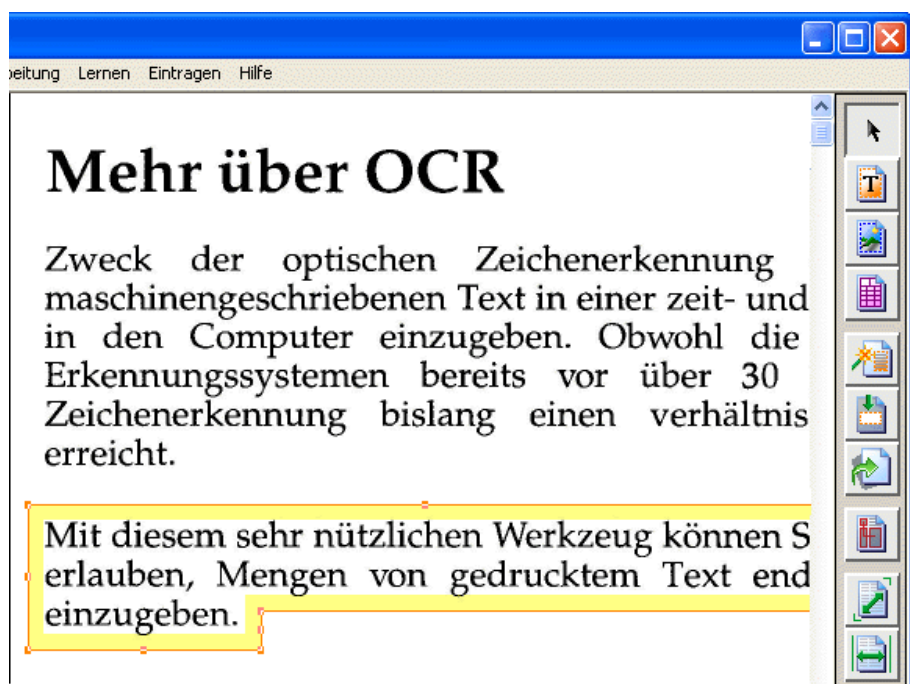
Keine Sorge: Falls Sie den falschen Zonentyp gewählt haben, können Sie den Typ schnell ändern, indem Sie mit der rechten Maustaste auf ein Fenster klicken und den Befehl "Fenster - Typ" im "Kontext" Menü wählen.





Die Fenster werden automatisch sortiert: Pfeile zeigen die Sortierreihenfolge an.

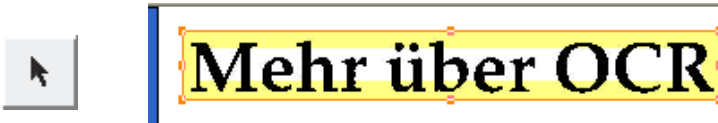
Sie können auch „unregelmäßige“ Textblöcke umrahmen, indem Sie **polygonale Fenster** erstellen. Sie können polygonale Fenster erstellen durch Zusammenzufügen recteckiger Fenster. Wenn zwei Fenster (derselben Art) einander berühren, macht Readiris automatisch ein polygonales Fenster daraus, als ob Sie ein Haus bauen, in dem Sie ein Zimmer nach dem anderen hinzufügen... (Es ergibt keinen Sinn, polygonale Tabellenfenster zu erstellen!)





Weiterhin kann manuelle Fensterbestimmung mit Fenstersortierung kombiniert werden: selbst wenn die Sortierungsfunktion aktiviert ist, können Sie neue Fenster zeichnen. Sie verwenden dann die Sortierungsfunktion, um einige der erkannten Fenster zu behalten und bestimmen dann manuell neue Fenster, dort, wo die Seitenanalyse nicht die gewünschten Ergebnisse hervorgebracht hat. Wenn Sie Fenster in der Sortierungsfunktion bestimmen, werden alle Fenster, die Sie nicht vorher angewählt haben, sofort gelöscht!

Um Fenster zu verändern, zu verschieben und zu löschen müssen sie zuerst **bestimmt** werden. Selektieren Sie hierzu die Fenster-Auswahlfunktion (die „Pfeil“-Schaltfläche) in der Bild-Toolbar und klicken Sie in das Fenster. Es erscheinen jetzt rechteckige Aktivpunkte in jeder Ecke und in der Mitte des Fensters.

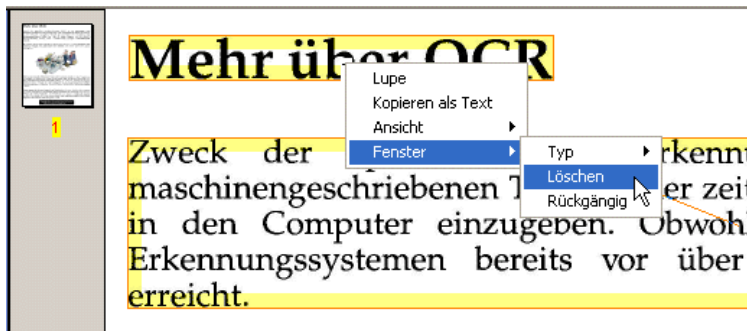


Um die **Selektion** eines Fensters zu **annulieren**, klicken Sie außerhalb des Fensters. Um **zusätzliche Fenster** zu selektieren, halten Sie die Umschalt Taste gedrückt, während Sie die Fenster anklicken. Um ein **Fenster einschließlich** des sich wiederum in diesem Fenster (eines anderen Typs) befindlichen Fenster zu aktivieren, halten Sie die Strg-Taste gedrückt und klicken Sie dabei das Hauptfenster an.

Soweit die Fensterauswahl. Um ein Textfenster zu **verändern**, selektieren Sie zuerst das Fenster, bewegen anschließend den Cursor über einen der Aktivpunkte und bewegen Sie die Maus um die Fenstergröße zu verändern.

Um ein Fenster zu **verschieben**, selektieren Sie einfach das Fenster und ziehen Sie es zur gewünschten Position.

Um ein Fenster zu **löschen**, selektieren Sie das (die) Fenster, klicken Sie mit der rechten Maustaste auf ein Fenster und wählen Sie den Befehl „Fenster - Löschen“ im „Kontext“ Menü. Hierdurch werden alle ausgewählten Fenster sowie das Fenster unter dem Mauszeiger gelöscht.



Alternativ können Sie Zonen auswählen und den Befehl „Ausschneiden“ oder „Löschen“ im Menü „Bearbeiten“ wählen. Der Befehl „Ausschneiden“ schneidet das Fenster aus und fügt es in ein interner Puffer ein. Der Befehl „Löschen“ löscht das Fenster definitiv. Werden Fenster eingefügt, kommen sie an ihre ursprüngliche Position und müssen zum neuen Platz gezogen werden.

Es können also *alle* bekannten Befehle aus dem Menü „Bearbeiten“ bei den Fenstern angewendet werden: Sie können sie löschen, ausschneiden, kopieren und einfügen. Der Befehl „Rückgängig“ kann ebenfalls angewendet werden. Wenn Sie versehentlich einige Fenster gelöscht, bewegt, verändert usw. haben, wird der Befehl rückgängig gemacht mit dem Befehl „Rückgängig“.

Bearbeiten		
Rückgängig		Ctrl+Z
Ausschneiden		Ctrl+X
Kopieren		Ctrl+C
Einfügen		Ctrl+V
Löschen		Delete
Kleine Fenster löschen		Ctrl+M
Alles auswählen		Ctrl+A

Für alle Befehle können Tastenkombinationen benutzt werden! Beispielsweise: Damit alle bestehenden Fenster gelöscht werden, können Sie den Befehl „Alles

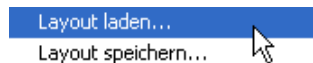


Auswählen“ im Menü „Bearbeiten“ selektieren, oder die Tastenkombination Strg+A und den Befehl „Löschen“ aus dem Menü „Bearbeiten“ oder die Taste Delete benutzen. Jetzt kann das Layout wiederhergestellt werden. Um das vorherige Layout wiederherzustellen, selektieren Sie „Rückgängig“ oder benutzen Sie die Tastenkombination Ctrl+Z.

## **DREI, FENSTERVORLAGEN SPEICHERN**

---

Die Fenstervorlagen können mit vorbestimmten **Fensterstrukturen** mit dem Befehl „Layout Speichern“ im Menü „Datei“ gespeichert werden und in den Speicher geladen werden mit dem Befehl „Layout Laden“.



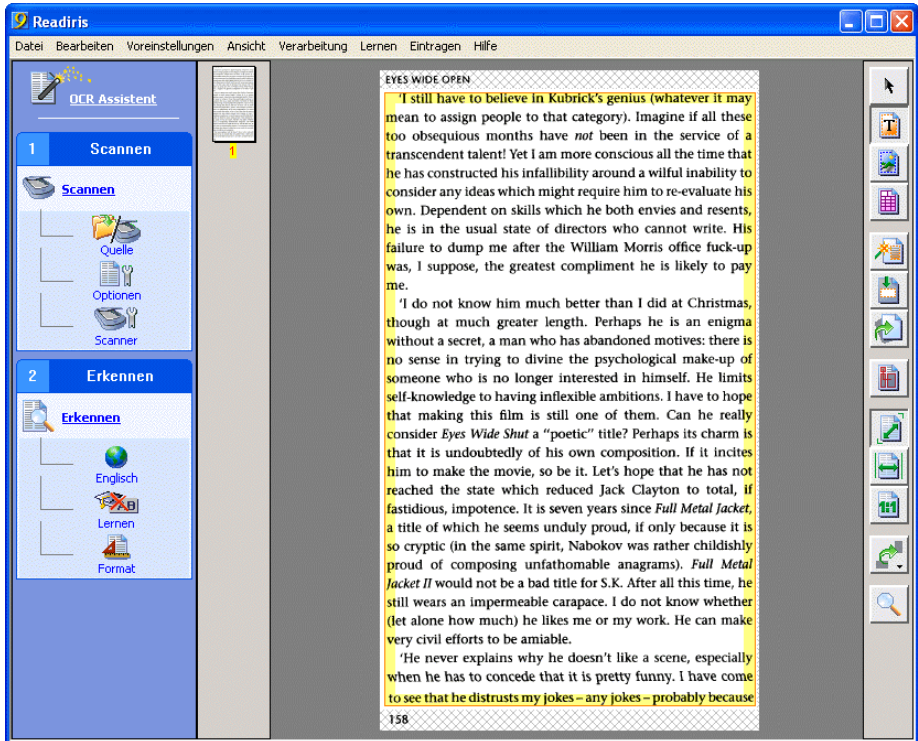
Wenn Sie Dokumente mit einem ähnlichen Layout lesen müssen, zum Beispiel ein 50-seitiges Protokoll, wo Dachzeile und Fußnoten aus bestimmten Gründen unsichtbar bleiben sollen, dann kann eine einzige Vorlage benutzt werden um 50 Seiten in Zonen einzuteilen.

Wenn Sie eine solche Vorlage in den Speicher laden wird die Seitenanalyse automatisch inaktiviert. Die Layoutvorlage bleibt solange aktiviert bis die Seitenanalyse auf der Haupttoolbar wieder aktiviert wird.

Es gibt eine gute Alternative zu Zoneneinteilungs-Vorlagen: Das Vorschau-Tool „Außenseite von Gebiet Ignorieren“ begrenzt die Zerlegung der Seite auf den „zugeschnittenen“ Teil des Bildes.



Aktivieren Sie dieses Tool und ziehen Sie einen Rahmen um den Teil des Bildes, den Sie verarbeiten möchten. Wenn Sie ein mehrseitiges Dokument bearbeiten, können Sie dieselbe Außenzone auf allen Seiten von der Seitenanalyse ausschließen. (Führen Sie die Seitenanalyse erneut durch, um den Bildzuschnitt aufzuheben, oder ändern Sie die Zonen manuell.)



## MIT READIRIS UM DIE WELT!

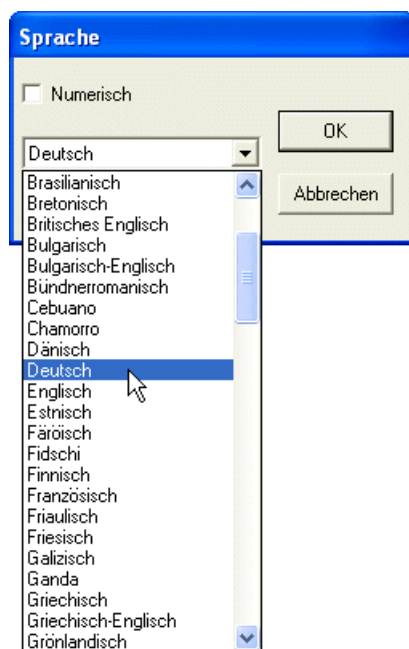
Vorausgesetzt Sie haben die Fensterbestimmung korrekt durchgeführt, dann sind Sie jetzt fast soweit, dass Sie die Zeichenerkennung aktivieren können. Wir sagen „fast“, weil wir die Sprachenauswahl und die Dokumenteinstellungen noch nicht kontrolliert haben!

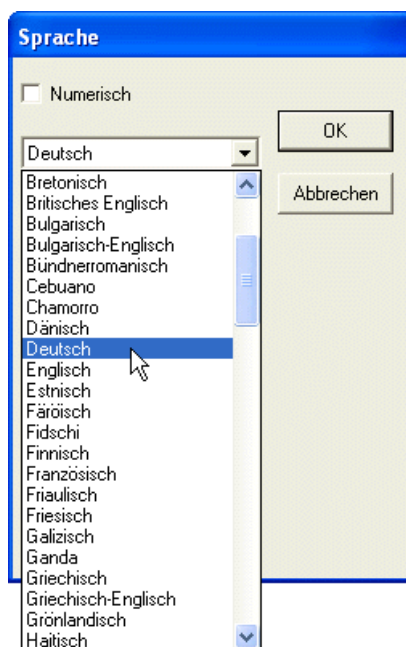


Die Spracheneinstellungen finden Sie auf der Haupttoolbar.



Um die Dokumentensprache zu verändern klicken Sie „Sprache“ an.





Sie können auch direkt eine Taste auf der Tastatur drücken: wenn Deutsch meistens gewählt wird und Sie möchten jetzt Occitanisch selektieren, drücken Sie dann die „O“ Taste auf Ihrer Tastatur um direkt Zugang zur occitanischen Sprache zu bekommen. Für mehrere Sprachen mit dem gleichen Anfangsbuchstaben drücken Sie den Buchstaben mehrmals, bis Sie die gewünschte Option bekommen. Zum Beispiel: Readiris kann Dänisch und Deutsch lesen. Wenn Sie das „D“ einmal drücken, selektieren Sie Dänisch, wenn Sie das „D“ ein zweites Mal drücken, selektieren Sie Deutsch und beim dritten Mal sind Sie wieder bei Dänisch. (Um einen *anderen* Buchstaben, sagen wir ein T, einzugeben, gehen Sie zuerst auf Zurück, bevor Sie das T eingeben.)



Readiris beschränkt sich nicht nur auf Deutsch: Es werden bis zu 107 **Sprachen** unterstützt! Readiris unterstützt alle amerikanischen und europäischen Sprachen sowie die zentraleuropäischen, griechischen, türkischen, kyrillischen („russischen“) und baltischen Sprachen.

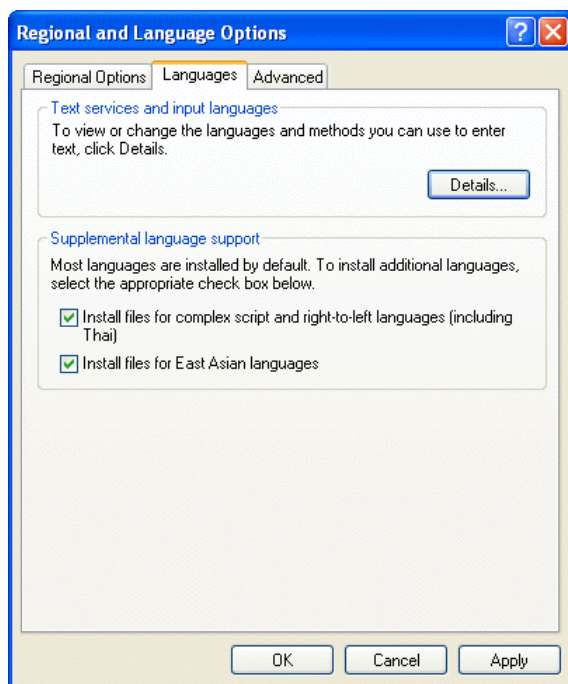
Es ist optional möglich, **asiatische Dokumente** zu lesen: Die Software-Option „Asiatisches OCR Modul“ ermöglicht die Erkennung von Japanisch, traditionellem und vereinfachtem Chinesisch sowie Koreanisch. (Vereinfachtes Chinesisch wird auf dem chinesischen Festland und in Singapore verwendet, während traditionelles Chinesisch in Hong Kong, Taiwan, Macau und den chinesischen Gemeinschaften in Übersee verwendet wird.)

Denken Sie auch daran, dass die britischen und amerikanischen - oder sagen wir „internationalen?“ - Varianten der englischen Sprache ebenfalls differenziert werden.

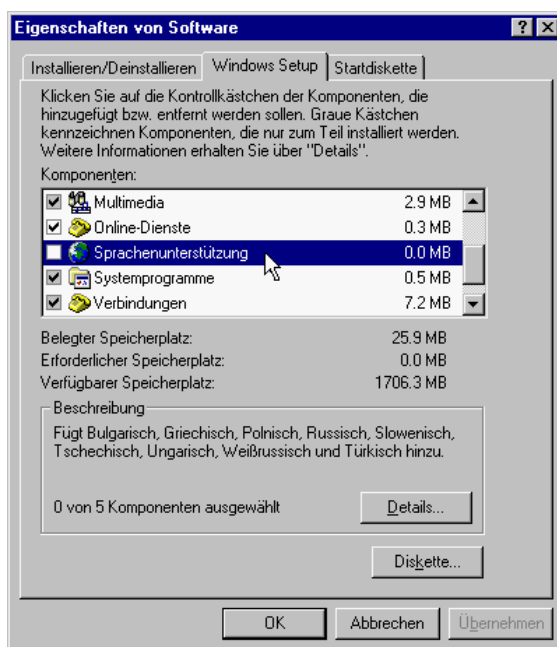
Merken Sie sich auch, dass Sie die angepaßte Windows Konfiguration installieren müssen, damit zentraleuropäische, griechische, türkische, kyrillische und baltische Zeichen angezeigt werden. Vielleicht werden Sie das **Windows Bestandteil „Sprachenunterstützung“** installieren müssen, bevor Ihr Windows System in der Lage ist mit diesen Sprachen zu arbeiten.

Auf einem Rechner mit Windows XP, 2000 oder Windows NT 4.0 wählen Sie das Symbol „Regionale Einstellungen (und Sprachen)“ in der „Systemsteuerung“.





Wählen Sie auf Ihrem Windows ME oder Windows 98 Rechner in der „Systemsteuerung“ das Symbol „Software“, um zu erfahren, ob das Modul „Sprachenunterstützung“ auf Ihrem PC installiert ist.



Um asiatische Dokumente zu betrachten und zu bearbeiten, können Sie eine asiatische Windows Version installieren oder eine spezielle „Emulationssoftware“ (wie UnionWay AsianSuite oder Twinbridge AsianBridge) unter einer westlichen Windows-Version betreiben, um die Ideogramme der asiatischen Sprachen korrekt darzustellen. Schließlich können Sie auch Word 2003, Word 2002 oder 2000 zum Betrachten und Bearbeiten solcher Dokumente verwenden: Office 2003 System, Office XP und 2000 wurden speziell im Hinblick auf die Verarbeitung von Dokumenten in vielen verschiedenen Sprachen entwickelt.

Weitere Informationen zu diesem Thema finden Sie in der Readiris „Liesmich“-Datei.

Es ist sehr wichtig, die dokumenteneigenen Sprache zu selektieren. So weiß die Software genau, welche **Zeichen** sie zu erkennen hat. Dieses ist natürlich abhängig von Ihrer Sprachenauswahl. Multilinguistische Unterstützung ermöglicht die Erkennung von „exotischen“ Zeichen wie ç, ñ, ψ und ø.

Desweiteren benutzt die Software die **linguistischen Datenbanken** sehr ausgiebig, um die Ergebnisse zu kontrollieren. Nehmen wir einmal an, das Wort „Präsident“ muß gelesen werden und auf dem Buchstaben „r“ befindet sich ein Tintenfleck, so dass das „r“ wie ein „f“ aussieht. Bei der Kontrolle im deutschen Wörterbuch wird Readiris automatisch feststellen, dass das Wort „Präsident“ gelesen wurde und dass ein „f“ in diesem Kontext keinen Sinn ergeben würde. Diese „**autodidaktische**“ **Technik** ist natürlich sehr abhängig vom linguistischen Kontext.

Die linguistische Unterstützung kann sehr hilfreich sein bei der Erkennung von „**Zweifelsfällen**“, so wie das „O“, das als „0“ gelesen werden könnte. Ein anderes typisches Beispiel ist „l“ und „I“, die oft eine identische Form haben - denken Sie an Texte, die mit einer alten Schreibmaschine geschrieben wurden! Der linguistische Kontext hilft Ihnen zu erkennen, ob es sich um „l“ handelt oder um „I“.

Folgendes Beispiel illustriert die Schattenvariationen von „l“ und „I“. Die Schatten der ersten Linien sind deutlich abgegrenzt, die der zweiten Linien sind zweifelhaft abgegrenzt. Dieses Problem wird durch die Sprachwissenschaft gelöst. Wenn der Kontext nicht ausreicht, liegt die Unterscheidung beim Benutzer.

193 1950s. 1hr  
Well, Rossellini

## **BEI BEDARF VERÄNDERT READIRIS DIE SPRACHEN**

Aber das ist noch lange nicht alles. Readiris ist in der Lage, ohne Eingriff des Anwenders, in der Mitte eines Textes von einer Sprache zu einer anderen zu wechseln! Wenn „Westliche“ Worte auftauchen in griechischen, kyrillischen oder

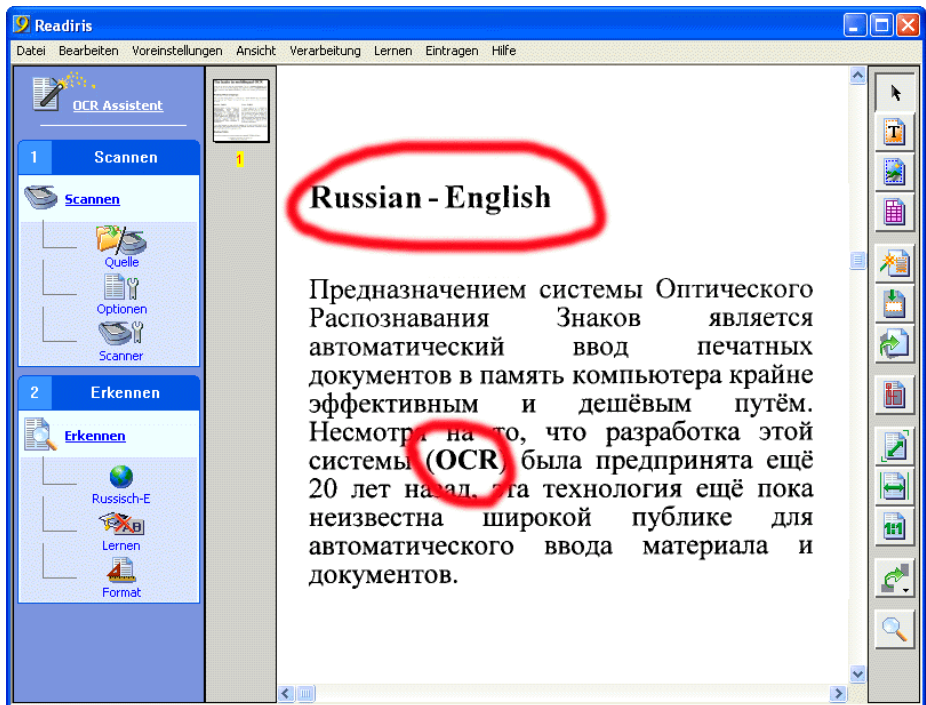


asiatischen Dokumenten - viele Eigennamen, Markennamen usw. werden nicht umgesetzt, sondern mit „normalen“ lateinischen Zeichen geschrieben - kann Readiris kann sofort zum richtigen Alphabet wechseln. Mit anderen Worten, es ist möglich, „**gemischte**“ **Alphabete** von griechischen, kyrillischen oder asiatischen und westlichen Zeichen gleichzeitig zu aktivieren.

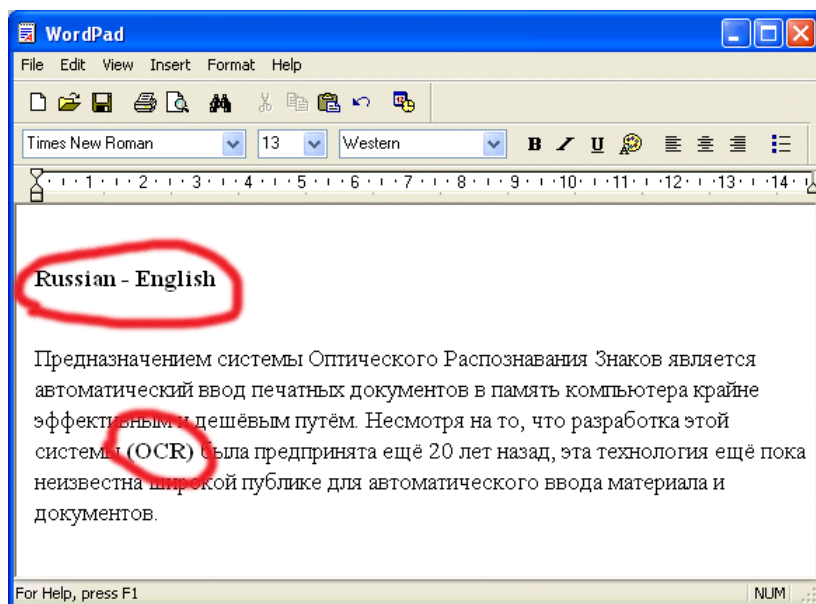
Vergewissern Sie sich, dass die „Griechisch-Englisch“ oder eine übereinstimmende kyrillische Einstellung gewählt wurde - zum Beispiel „Weiß-Russisch-Englisch“. Mit anderen Worten: Selektieren Sie nicht nur „Griechisch“ oder „Weiß-Russisch“ als Dokumentensprache in der Hoffnung, dass die westlichen Symbole perfekt erscheinen werden!



Hier sehen Sie ein Beispiel eines russischen Textes mit einigen englischen Wörtern - öffnen Sie die Bilddatei ALPHABETS.TIF wenn Sie dieses selbst einmal ausprobieren möchten.



Das Endergebnis sieht aus wie folgt: Sie werden vielleicht doch noch einen kyrillischen **Zeichentyp** selektieren müssen, damit der russische Text richtig angezeigt wird.



Um **andere Sprachen** mit einander auf einer Seite zu **mischen**, wählen Sie die Sprache mit dem größten Zeichensatz. Wenn Sie ein englisches Dokument haben, das ebenfalls die französische Übersetzung enthält, selektieren Sie am besten Französisch als Sprache, damit auch die betonten Buchstaben ç, é und ù erkannt werden können.

## BESTIMMEN DER DOKUMENTEIGENSCHAFTEN

Nachdem die Sprache eingestellt wurde, befassen wir uns mit den anderen Eigenschaften des Dokuments. Sie können die Erkennung noch besser gestalten, indem Sie einige Dokumentmerkmale spezifizieren: Typ der Schriftart und

Zeichenlaufweite. (Diese Befehle finden keine Anwendung auf asiatische Dokumente.) Was bedeutet dies nun alles?

Beginnen wir mit dem Befehl „Typ der Schriftart“ im Menü „Voreinstellungen“. Bei den Schrifttypen wird zwischen „normalen“ Dokumenten und **Punktmatrix**-Dokumenten unterschieden. „Entwurf“- oder „9-Nadel“-Punktmatrixsymbole bestehen aus isolierten, getrennten Punkten, für deren Erkennung höchst spezialisierte Erkennungsroutinen erforderlich sind.

**ape-descended life**

Punktmatrixausdrucke mit „Briefqualität“ (auch als „25-Nadel“- oder „NLQ“-Ausdrucke bezeichnet) erfordern die Einstellung „normal“, so wie dies auch für die **Druckqualitte** Druck, Maschinenschrift, Laserdruck und Tintenstrahl-Druck der Fall ist.

Die Einstellung „Automatisch“ besagt, dass Readiris automatisch den Schriftartmodus ermittelt. Lassen Sie Readiris den Schriftmodus in allen Fllen automatisch erkennen, auer wenn Sie sicher sind, dass nur Punktmatrix-Dokumente eingelesen werden! (Selbstverstndlich ist „Automatisch“ der Standardwert.)



Der Schrifttyp ist im Quickinfo der Schaltflche „Erkennen“ angegeben: Wenn dem Quickinfo keine Angabe hinzugefgt ist, ist die automatische Erkennung der Druckqualitt aktiviert. Wenn die Nachricht „Punktmatrix“ im Quickinfo erscheint, ist der Punktmatrix-Lesemodus aktiviert.





Die **Laufweite** kann mit dem Befehl „Zeichenabstand“ im Menü „Voreinstellungen“ eingestellt werden.



Bei *festen* Schriftarten haben alle Buchstaben die gleiche Breite. Ein „i“ nimmt genau so viel Platz ein wie ein „w“, wie in diesem Satz vorgeführt. Wir denken hier an Dokumente, die mit einer Schreibmaschine geschrieben wurden, wo eine feste Entfernung der Buchstaben von einander vorgegeben ist.

Bei der *Proportionalschrift* ist die Laufweite abhängig von der Form des Zeichens. Zeichen wie „m“ und „w“ sind breiter und brauchen mehr Platz als die „feinen“ Zeichen so wie „i“ oder „j“. Fast alle Bücher, Zeitschriften und Zeitungen sind in Proportionalschrift gedruckt.

Hier ist es am einfachsten, wenn Sie diese Option in der Standardeinstellung „Automatisch“ lassen, damit Readiris den Zeichenabstand automatisch erkennen kann.

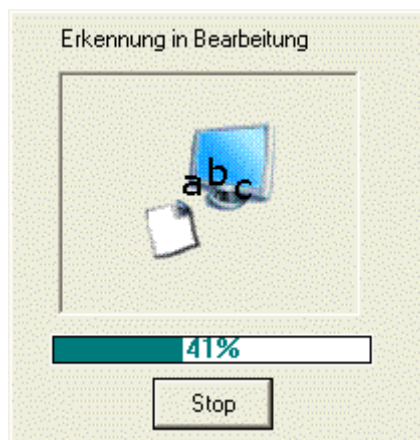
## **READIRIS WIRD IMMER INTELLIGENTER!**

Wenn die Dokumentensprache und die Dokumenteigenschaften eingestellt sind, aktivieren Sie das interaktive Lernen und klicken Sie die Schaltfläche „Erkennen“ an.



Das OCR-Programm wird angezeigt. Sie können die Schaltfläche „Stop“ anklicken, um die Texterkennung abubrechen.





Am Ende jeder Erkennungsphase aktiviert Readiris die interaktive Lernphase. Die Lernfunktion muß aber aktiviert sein mittels der Schaltfläche „Lernen“ auf der Haupttoolbar.

(Das interaktive Lernen findet keine Anwendung auf asiatische Dokumente: Lernen in diesen Sprachen ist zwecklos, weil es sich hier um tausende verschiedener Symbole handelt - und Sie müßten in der Lage sein, die Ideogramme mit einer westlichen Tastatur einzugeben!)

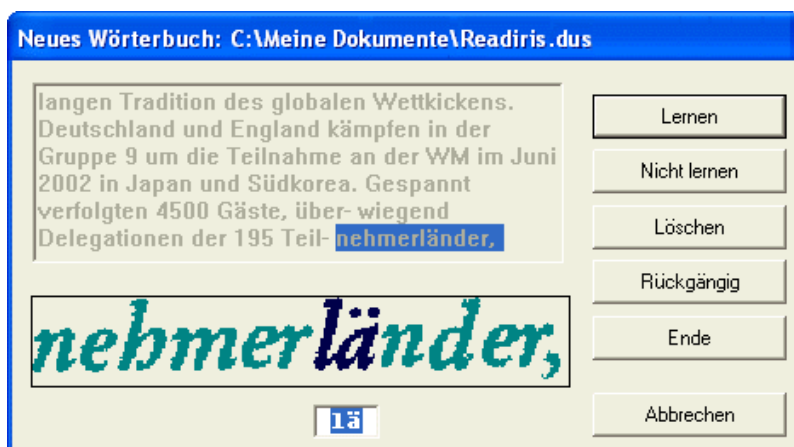
**Zeichentraining** bietet eine Möglichkeit, um die Genauigkeit des Erkennungssystems zu verbessern. Das Training ist eine große Hilfe beim Lesen von Texten mit beschädigten Zeichen wie man in reellen Dokumenten findet, und beim Lesen von stilisierten Schriftarten, welche Readiris nicht optimal erkennen kann.

Das Lernsystem wird auch angewendet um das System mit **Sonderzeichen** vertraut zu machen: so können mathematische und wissenschaftliche Zeichen und Dingbats von Readiris erkannt werden. Einige Beispiele: Readiris kann lernen, das „ $\pi$ “-Symbol als „pi“ oder das Dingbat „☎“ als „Tel“ zu erkennen. (Die Liste



mit den erkannten Zeichen kann aber nicht mit den Symbolen „π“ und „☎“ aufgestreckt werden.)

Der erkannte Text wird schrittweise auf dem Bildschirm wiedergegeben. Das System stoppt, wenn es Zweifel hat, ein Zeichen - oder eine Zeichenreihe wie bei „zusammenlaufenden“ Zeichen („Ligaturen“) - richtig zu erkennen. Sie werden immer in ihrem Kontext angezeigt, die Zweifelsfälle werden eingeblendet. Nicht erkannte Zeichen werden durch eine Tilde (das „~“ Symbol) dargestellt.



Kontrollieren Sie zuerst, ob das richtige Zeichenwörterbuch und den richtigen Wörterbuch-Modus aktiviert wurden. Beide werden immer im Titelfeld des Lernfensters angezeigt. Wenn nicht, klicken Sie die Schaltfläche „Abbrechen“ an; das Dokument wird jetzt wieder angezeigt mit dem Fensterlayout. Aktivieren Sie nun das richtige Zeichenwörterbuch oder den richtigen Wörterbuch-Modus und fangen Sie neu an. (Die Wirksamkeit von „Zeichenwörterbuch“ wird bald besprochen werden.)

Wenn erforderlich, geben Sie über die Tastatur das richtige Zeichen (oder die richtige Zeichengruppe) ein und selektieren Sie einen der nachstehenden Befehle.

## **Lernen**

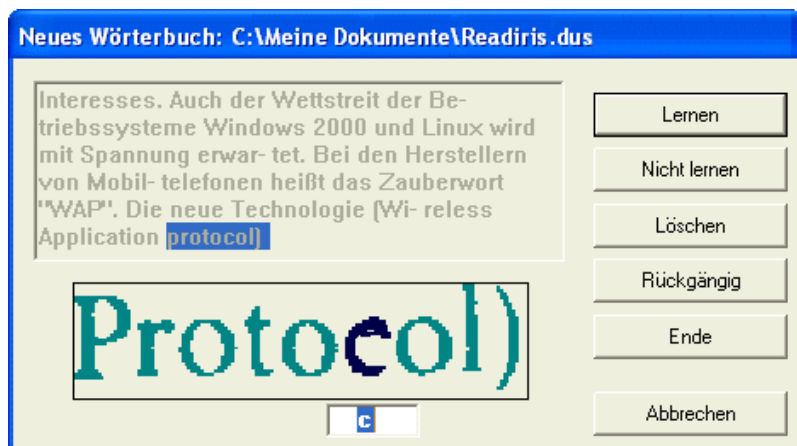
Sie sind mit der vorgeschlagenen Lösung einverstanden oder Sie haben die Lösung korrigiert. Readiris speichert das Zeichen dann im Zeichenwörterbuch als „sicher“ ab. Dieses Zeichen wird dann künftig nie wieder angezeigt werden. Eine Bestätigung Ihrerseits ist nicht mehr erforderlich.

Im oben erwähnten Beispiel hält das System bei einige zusammenlaufenden Zeichen an. Wir klicken auf „Lernen“, um die Zeichen, die nicht mit anderen verwechselt werden können, zu akzeptieren.

## **Nicht Lernen**

Sie sind mit der vorgeschlagenen Lösung einverstanden, oder Sie haben die Lösung korrigiert. Der Unterschied zu der Schaltfläche „Lernen“ liegt darin, dass das erlernte Symbol als „unsicher“ im Zeichen-Wörterbuch abgespeichert wird. Künftig wird das System die „gelernte“ Lösung vorschlagen, sie muß aber immer wieder bestätigt werden.

Diese Schaltfläche wird bei Symbolen, die leicht zu verwechseln sind, benutzt: ein schlecht geformtes „e“ kann als „c“ gelesen werden, ein schlechtes „t“ ähnelt einem „r“ usw.



Das oben gezeigte „c“ ist stark beschädigt - eigentlich sieht es mehr aus wie das „e“-Symbol. Jetzt sollten Sie die Schaltfläche „Nicht Lernen“ anklicken, damit es nicht mit dem „e“-Symbol verwechselt wird.

## **Löschen**

Das angezeigte Zeichen verschwindet von der Ausgabedatei. Diese Schaltfläche wird benutzt, um bestimmte „Verunreinigungen“ so wie zum Beispiel Flecke, Kaffeefflecke usw. zu eliminieren. Diese könnten als Punkte, Kommas usw. erkannt werden. Es läßt auch jedes ungewollte Symbol verschwinden.

## **Rückgängig**

Widerruft die zuvor erteilten Befehle. Sie können die letzten 32 Entscheidungen rückgängig machen.

## **Ende**

Der Lernprozeß wird abgebrochen, aber das OCR-System läuft weiter im Auto-Modus. Alle Entscheidungen werden ohne Bestätigung des Anwenders akzeptiert.

Klicken Sie diese Schaltfläche an wenn Sie feststellen, dass die Erkennung sehr akkurat durchgeführt wurde und Korrekturlesen nicht erforderlich ist.

## **Abbrechen**

Verwechseln Sie „Ende“ nicht mit „Abbrechen“! Bei „Abbrechen“ gibt es keine Textausgabe und Sie müssen neu anfangen. Bei „Ende“ wurde der Text erstellt, aber noch nicht überprüft!

## **DIE ROLLE DER ZEICHENWÖRTERBÜCHER**

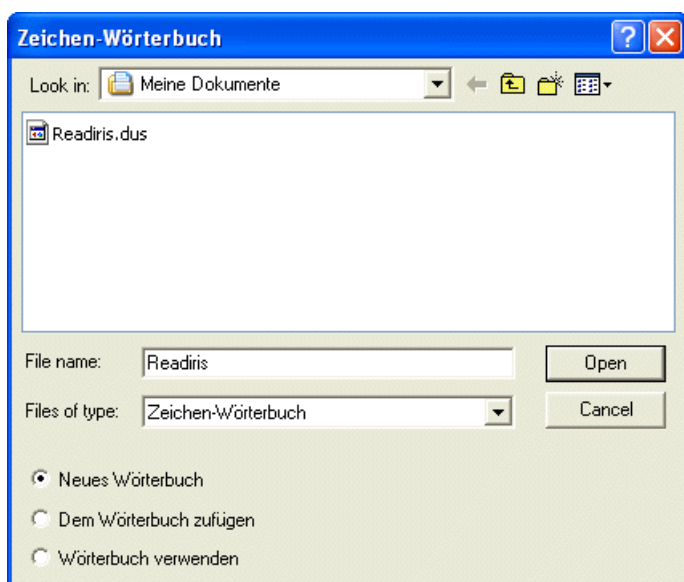
---

Die Ergebnisse jeder Lernphase werden übergangsweise im Speicher festgehalten. Sie sollten jedoch in sogenannten „Wörterbüchern“ gespeichert werden, damit sie auch später noch verwendet werden können.

(Benutzerlexika dürfen nicht mit Fontwörterbüchern verwechselt werden! Fontwörterbücher enthalten Buchstabenformen, die während der interaktiven OCR-Phase erlernt wurden, während es sich bei Benutzer-Lexikon um linguistische Datenbanken handelt, die den Erkennungsprozess unterstützen.)

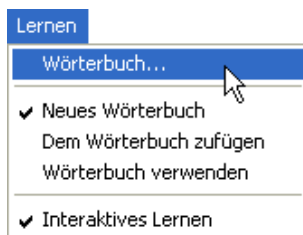
Zeichenwörterbücher sollten in den Speicher geladen werden, um die darin enthaltenen Informationen beim Erkennen ähnlicher Dokumente zu nutzen. So kann Readiris die „Intelligenz“ der Font-Wörterbücher verwenden. Man kann sagen, dass Readiris bei jeder Nutzung intelligenter wird!

Wie funktioniert das? Die Funktion der Zeichenwörterbücher wird vom Menü „Lernen“ kontrolliert: Es muß ein Wörterbuch mit dem Befehl „Wörterbuch“ gewählt werden und der Handlungsmodus muß bestimmt werden.



**Zeichen-Wörterbücher** haben eine Kapazität von 500 Zeichen. Für Sonderanwendungen empfehlen wir Ihnen separate Wörterbücher anzulegen, zum Beispiel nach Typ des Dokuments. Wörterbücher haben die Standarderweiterung \*.DUS. Wenn ein Wörterbuch voll ist, ist es zwecklos, noch weiter zu lesen: Die Ergebnisse können dann weder im Speicher, noch im Wörterbuch festgehalten werden.

Der Wörterbuchmodus kann im Befehl „Wörterbuch“ oder direkt im Menü „Lernen“ eingestellt werden. Es gibt drei Wörterbuchmodi: Neu, zufügen und verwenden.



Wenn Sie „Neues Wörterbuch“ selektieren, geben Sie an, dass die Ergebnisse in einem *neuen* Wörterbuch gespeichert werden müssen. (Wenn Sie ein bereits existierendes Wörterbuch selektieren, wird der Inhalt gelöscht.)

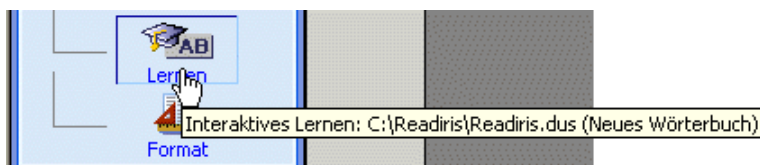
Der Modus „Zufügen“ gibt an, dass die Ergebnisse in einem bereits *existierenden* Wörterbuch gespeichert werden: Das Erkennungssystem nutzt bereits die zusätzliche Intelligenz der Wörterbücher und Sie fügen neue Zeichen hinzu. Mit anderen Worten: Mit dieser Option wird ein Zeichen-Wörterbuch schrittweise aufgebaut.

(Wenn Sie für ein neues Wörterbuch einen neuen Dateinamen eingeben und den Zufügen-Modus aktivieren, wird ein leeres Wörterbuch erstellt, das Sie ergänzen können.)

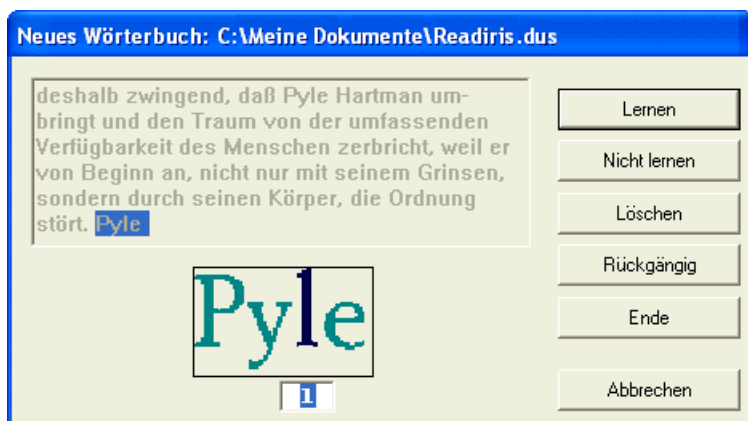
In der letzten Option „Wörterbuch Verwenden“, funktioniert das Wörterbuch nur im Lesen-Modus („Read Only“). Es werden dem Wörterbuch *keine* neuen Zeichen zugefügt. Sie verwenden es aber.

Wenn Sie nur eine Seite lesen müssen, selektieren Sie den Modus „Neu“. Bei mehreren Seiten der gleichen Schriftart selektieren Sie den Modus „Neu“ für die erste Seite, den Modus „Zufügen“ für einige darauffolgende Seiten und den Modus „Verwenden“ für den Rest des Dokumentes.

Das QuickInfo der Schaltfläche „Lernen“ zeigt Ihnen an, welches Wörterbuch aktiv ist und in welchem Modus es arbeitet.



Wenn Sie im interaktiven Lernbereich sind, werden die Wörterbücher und ihre Modi im Fenstertitel angezeigt. Wenn es sich um einen Fehler handelt, sollten Sie „Abbrechen“ klicken.



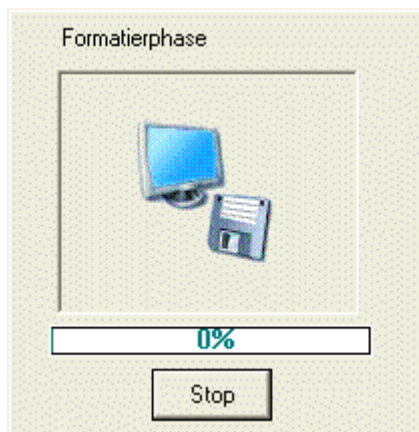
## **DAS ERGEBNIS DIREKT IN IHRE ANWENDUNG SENDEN**

Die interaktive Lernphase enthält auch die Zeichenerkennung. Da Microsoft Word, falls nicht anders eingegeben, automatisch Ausgabeziel ist, wird Ihre Textverarbeitung automatisch, falls notwendig, am Ende der Erkennung gestartet und der erkannte Text hineingesendet.

Am Bildschirm erscheint ggf. ein Fortschrittsbalken, der den Fortschritt der Dokumentformatierung anzeigt. (Ob dieser Balken erscheint oder nicht, ist von



der Größe des Dokuments und der Komplexität der auszuführenden Formatierung abhängig.)

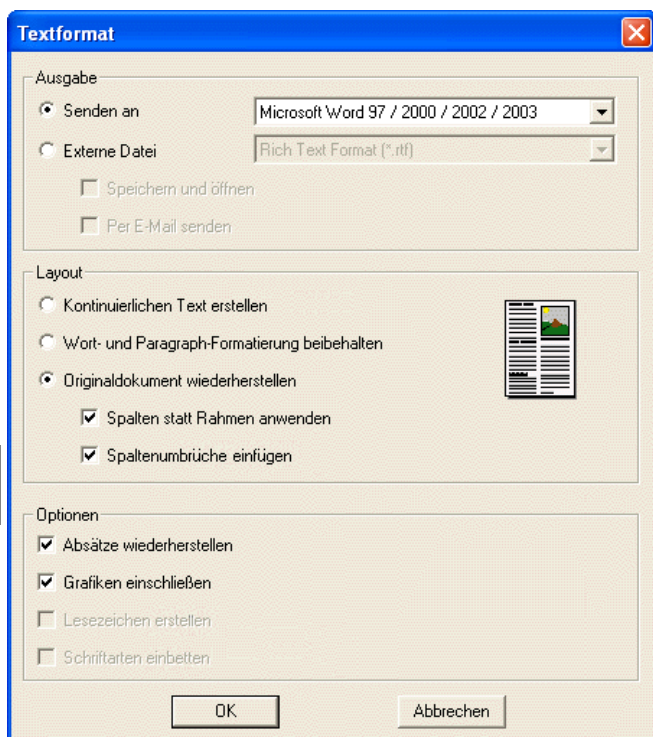


Das gescannte Bild erscheint mit den definierten Bereichen wieder auf dem Bildschirm. Es kann jetzt weiter bearbeitet werden und wird solange auf dem Bildschirm angezeigt, bis Sie eine andere Seite scannen.

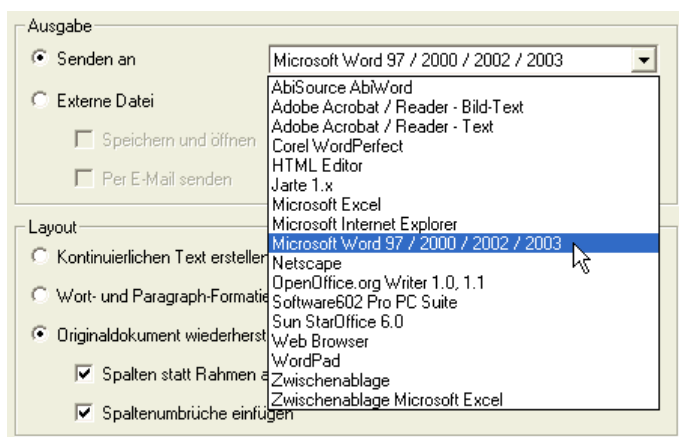
Sie haben jetzt ein Papierdokument in eine editierbare Textdatei verwandelt. Ein Vorgang der 40 Mal schneller ist, als das manuelle Abschreiben! Sie können es jetzt mit dem Bild, das Sie in Ihrem Readiris Fenster haben, vergleichen.

Readiris schlägt Ihnen drei verschiedene Methoden vor, um Ihr OCR-Ergebnis zu speichern: Das erkannte Dokument direkt in eine gewünschte Anwendung senden, es in einer externen Datei speichern oder es in die Windows Zwischenablage kopieren.

Das **Ausgabeziel** wird mit Hilfe des Befehls „Format“ auf der Haupttoolbar (oder des Befehl „Textformat“ unter dem Menü „Voreinstellungen“) gewählt.



Die Option „Senden an“ bietet eine direkte OCR-Verbindung zwischen Ihrem Scanner und Ihren Windows-Anwendungen: Sie **senden** die gescannten Dokumente direkt an Ihre Textverarbeitung, Tabellenkalkulation, Ihren Web Browser, an Adobe Reader usw.!



Am Ende der Erkennung wird die gewünschte Anwendung geöffnet und das erkannte Dokument wird in einer neuen Textdatei oder einem Worksheet geöffnet.

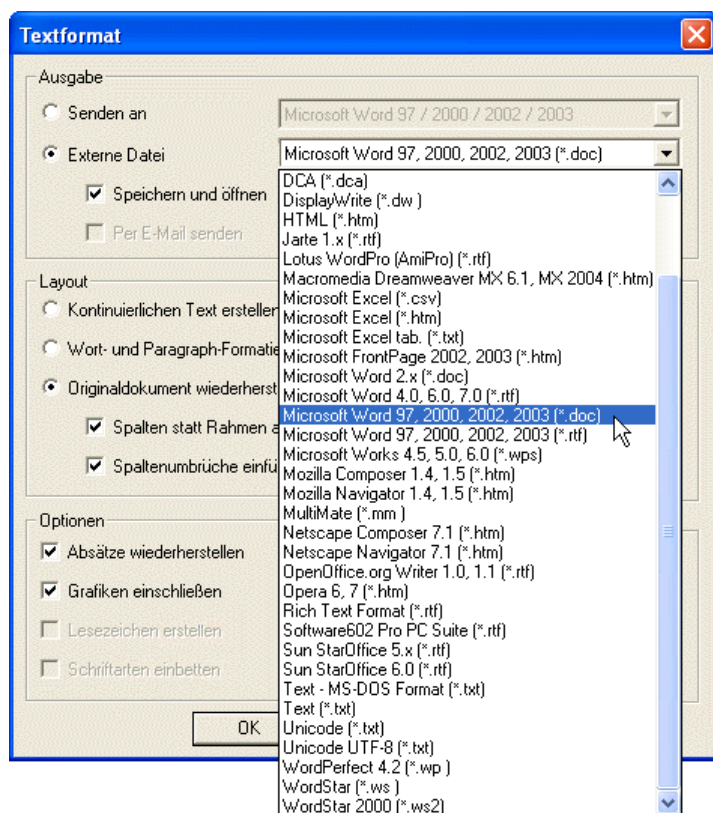




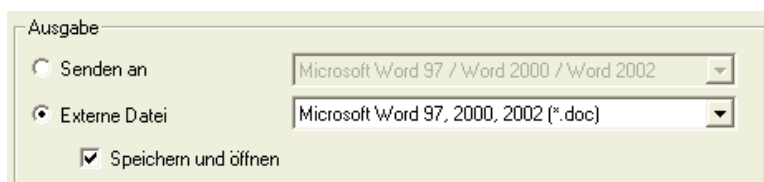
Vergessen Sie nicht, dass Sie mit der Option „Senden an“ den erkannten Text auch in die Windows **Zwischenablage** kopieren können. Es ist also nicht unbedingt notwendig, das Ergebnis zu exportieren... oder es in einer externen Datei zu speichern!

## **DAS ERGEBNIS IN EINER TEXTDATEI SPEICHERN**

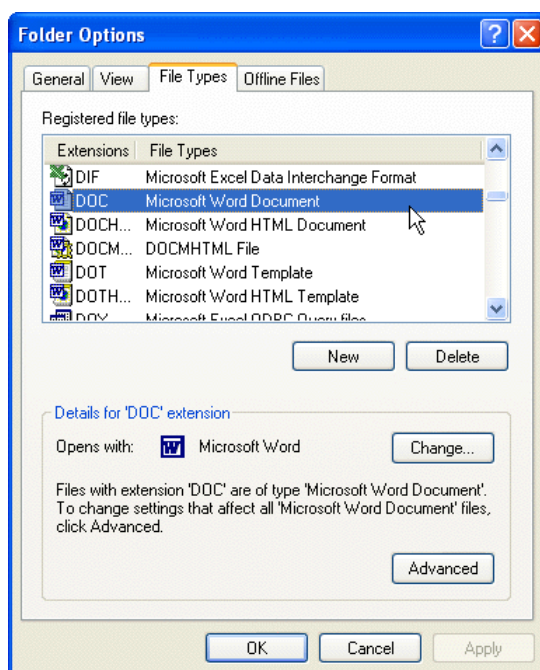
Sie können das OCR-Ergebnis tatsächlich als „externe“ Datei abspeichern. Readiris unterstützt auch hier eine Vielzahl von Dateiformaten, darunter alle bekannten Textverarbeitungen, Tabellenkalkulationen, HTML-Editoren usw.



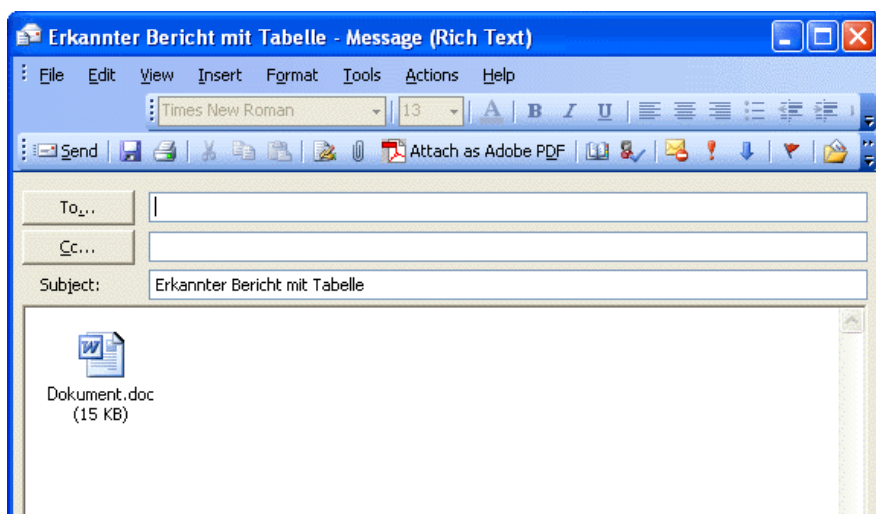
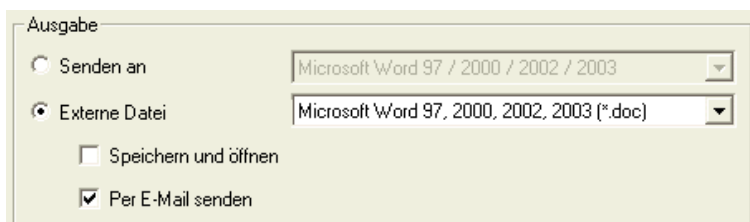
Die Option „Speichern und Öffnen“ ähnelt weitgehend der „Senden“-Funktion: Das erkannte Dokument wird nach dem Speichern gleich geöffnet.



Allerdings wird hierbei ein anderes Verfahren zur Auswahl der Zielanwendung eingesetzt. Welche Anwendung gestartet wird, hängt von den **in Windows festgelegten Dateitypen** ab. Das System verhält sich so, als hätten Sie die Ausgabedatei im Windows Explorer doppelgeklickt... (Bei Verwendung des Befehls „Senden an“ spricht Readiris die Zielanwendungen direkt an.)



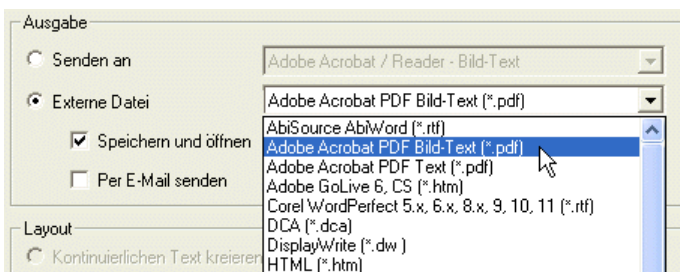
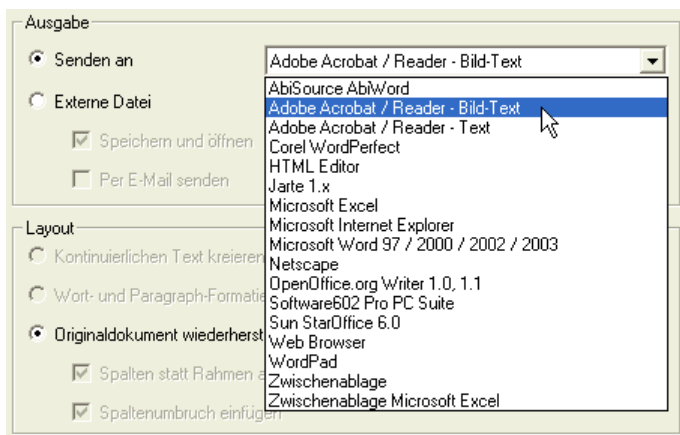
Der Befehl „Per E-Mail Senden“ erzeugt eine neue **E-Mail**-Nachricht, an die das erkannte Dokument angehängt wird. Kennen Sie eine schnellere Methode, ein Papier-Dokument rasch zu verbreiten...?





## PORTABLE DOKUMENTE ERSTELLEN...

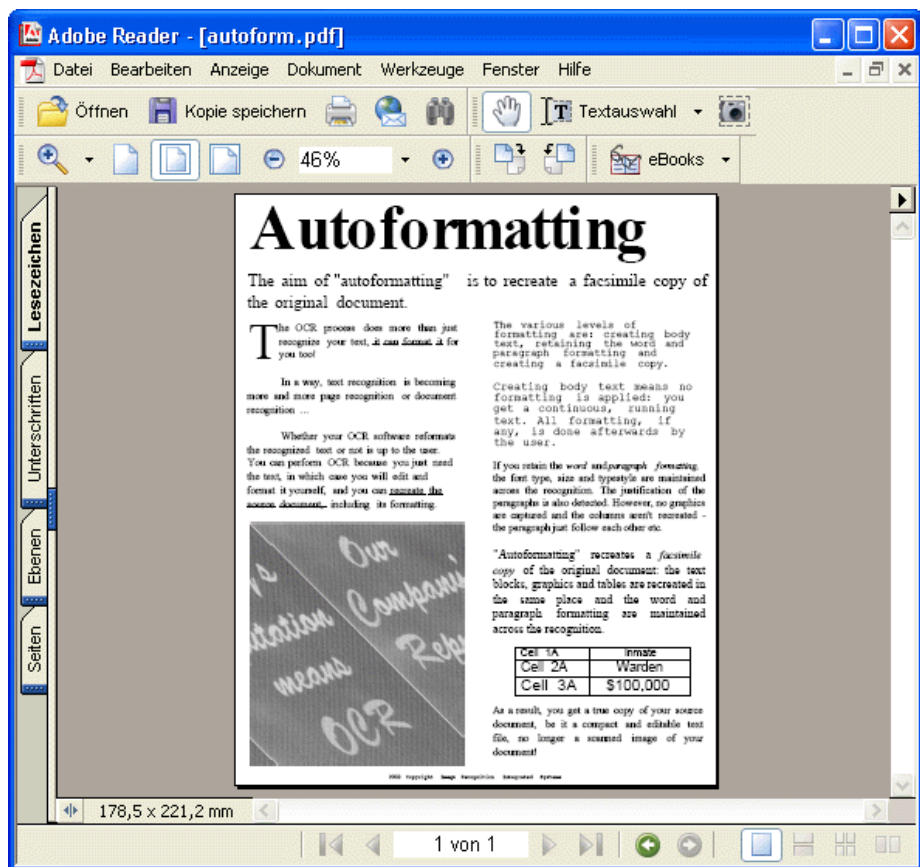
Gehen wir auf ein Format näher ein: **Adobe Acrobat PDF**. Readiris ermöglicht die Erzeugung von zwei Arten von PDF Dokumenten - PDF Text und PDF Bild-Text.



Worin besteht der Unterschied zwischen beiden Formaten? Wenn Sie das Format „PDF Text“ wählen, erzeugt Readiris eine PDF Datei, die die Textergebnisse enthält. (Grafiken können auftauchen, aber nur dann, wenn

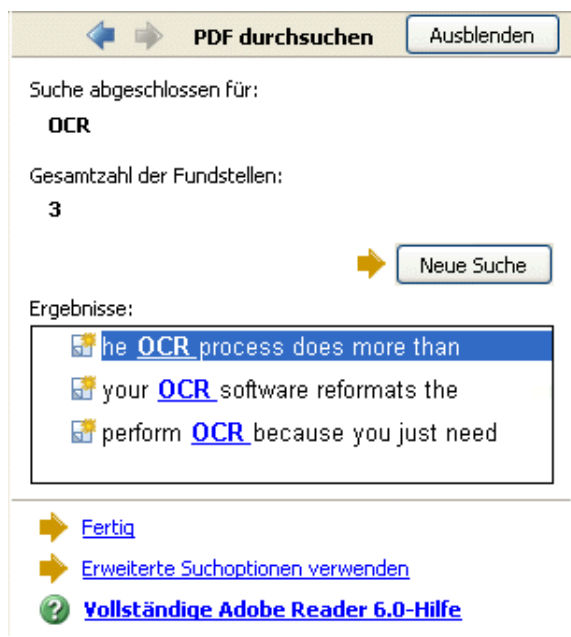


Grafikbereiche auf der Seite vorhanden sind - Fotografien, Verzerrungen usw.)  
Mit anderen Worten: Das Bild der Seite ist *nicht* in der einschichtigen PDF Datei enthalten!

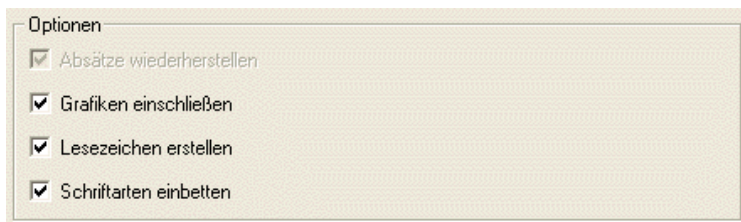




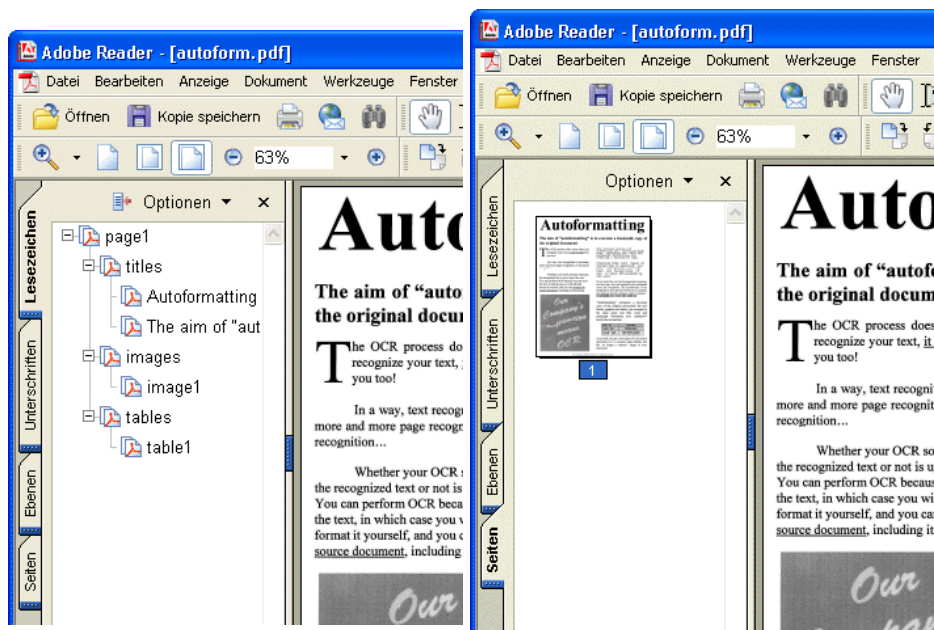
Das Format „PDF Bild-Text“ liefert ein anderes Ergebnis: Readiris erzeugt eine durchsuchbare PDF Datei, die den erkannten Text *und* das Bild der Seite enthält. Das Bild der Datei ist oberhalb des Textes in einer zweischichtigen PDF Datei enthalten. Verwenden Sie die „Suchen“-Funktion von Adobe Reader, und dies wird sehr bald deutlich!



Klicken Sie auf die „Format“ Schaltfläche, um zwei Optionen kennen zu lernen, die das Acrobat PDF Format betreffen: „Lesezeichen Erstellen“ und „Schriftarten Einbetten“.



Die Option „Lesezeichen Erstellen“ sorgt dafür, dass ein **Lesezeichen** („Bookmark“) für jedes Dokumentelement erzeugt wird - für Grafiken wie für Textblocks und Tabellen. Für Textbereiche verwendet Readiris einen intelligenten Algorithmus, der einen Titel, eine „Zusammenfassung“ für jeden Bereich erstellt. Die Tabellen und Grafiken werden einfach durchnummeriert. (Eine weitere Navigationshilfe in PDF Dokumenten, die sogenannte **Piktogramme** („Thumbnails“), können dynamisch von der Adobe Reader Software erstellt werden!)



Der Befehl „Schriftarten Einbetten“ bettet die Schriftarten in die PDF Dateien ein. Durch die Einbettung von Schriftarten wird eine Schriftartenersetzung beim Lesen und Drucken des erkannten Dokuments verhindert. Dies stellt sicher, dass alle Leser - unabhängig von der Konfiguration ihres Computers - immer den Text in dessen Originalschriftarten sehen. Die eingebetteten Schriftarten erhöhen allerdings (in geringem Maße) die Dateigröße der erkannten Dokumente!

---

## ... ODER LESEN

---

Betrachten wir es kurz aus einem anderen Blickwinkel. Da Readiris vollständige Unterstützung für das Adobe Acrobat PDF Format bietet, können Sie PDF Dateien nicht nur erzeugen, sondern auch *lesen*!

Das „**Repurposing**“ oder Ändern der Zweckbestimmung **von PDF Dokumenten** stellt ggf. einen wichtigen Einsatzbereich von Readiris dar. Hierfür gibt es mehrere Gründe. Zum einen können Sie so Bilder in Text umwandeln: Sie öffnen bildgestützte PDF Dokumente, führen die Erkennung durch und speichern das OCR-Ergebnis als Textdokument ab (in einem der unterstützten Textformate). Textdateien sind editierbar, Bilddateien nicht.

Zum anderen können Sie bildgestützte PDF Dateien in textgestützte PDF Dateien konvertieren. In diesem Fall führen Sie die Erkennung an der „nur-Bild“ PDF Datei durch und speichern das OCR-Ergebnis... als textgestütztes PDF Dokument! Textgestützte PDF Dateien können durchsucht und bearbeitet werden, „nur-Bild“ PDF Dateien nicht.

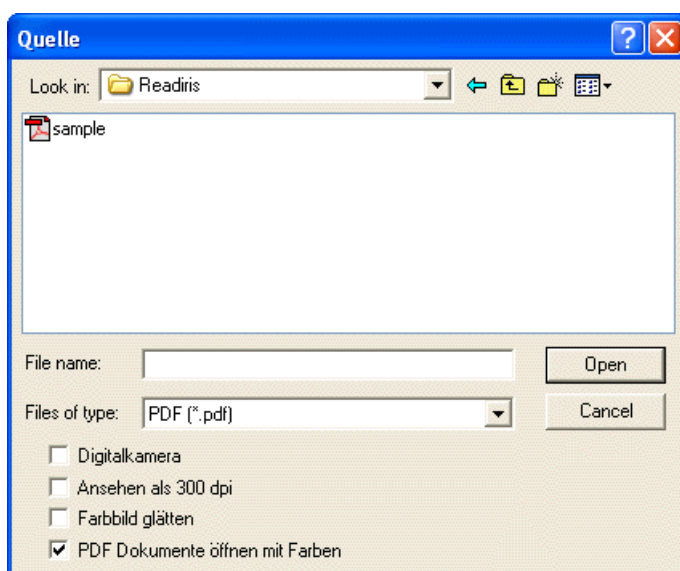
Und schließlich bietet die Umwandlung von PDF Dateien eine Möglichkeit, PDF Inhalte „freizuschalten“. Sie können die Erkennung an „nur-Lese“ PDF Dokumenten durchführen, deren Text normalerweise nicht zugänglich ist. In ungeschützten PDF Dateien kann der Inhalt ausgelesen (kopiert und als Text-Datei abgespeichert) werden, während der Inhalt von „nur-Lese“ Dateien nicht extrahiert werden kann. Diese Dokumente können nur eingesehen und gedruckt werden!

Eine wichtige Nuance: Readiris öffnet keine kennwortgeschützten PDF Dokumente, auch wenn Readiris alle anderen PDF Sicherungsmechanismen aushebelt!

Handeln Sie wie gewöhnlich: Laden Sie die PDF Dateien in den Speicher, wie Sie vorgescannte Bilder öffnen - Faxe, Schnappschüsse Ihrer Digitalkamera usw. Klicken Sie auf die Schaltfläche „Stop“ oder drücken Sie die Taste Escape, um den Ladevorgang zwischen zwei Seiten abubrechen.



Es gibt jedoch noch eine Besonderheit bei PDF Dateien. Sie können sie als Farbdokumente oder aber als Schwarzweiß-Dokumente öffnen! Diese Option wird angeboten, weil die Konvertierung von Farbdokumenten wesentlich langsamer ist.



## MEHRERE SEITEN ERKENNEN

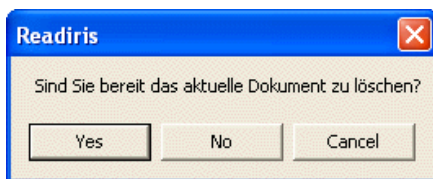
---

Nach dem OCR-Vorgang wird das gescannte Bild mit den definierten Bereichen zur weiteren Verarbeitung erneut angezeigt.

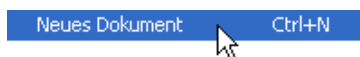
Sie können jetzt den gescannten Text in Ihrem Textverarbeitungsprogramm oder Ihrem Texteditor öffnen, ihn in Ihr Desktop-Publishing-Programm oder eine andere textbasierte Software importieren. Vergleichen Sie den Text mit dem Bild, das im Readiris Fenster angezeigt wird.

Wie aber speichern Sie den Text von zusätzlichen Seiten? Mit anderen Worten: Wie verarbeiten Sie mehrseitige Dokumente? Eigentlich ist es ganz einfach: Sie erkennen die Seiten nacheinander und speichern die Ergebnisse in ein und dieselbe Datei! (Sie sollten sich vergewissern, dass diese Datei nicht geöffnet ist; andernfalls können Sie diese Datei nicht schreiben!) Vergessen Sie nicht, auch das Font-Wörterbuch in den Zufügen-Modus zu setzen, um das interaktives Lernen mühelos fortzusetzen.

Sobald Sie Seiten scannen (oder Bilddateien öffnen) innerhalb eines Dokuments, müssen Sie entscheiden, ob ein neues Dokument erstellt werden soll, oder das aktuelle Dokument ergänzt werden soll.



Antworten Sie „nein“, um die Seiten zum aktuellen Dokument hinzuzufügen, oder „ja“, um ein neues Dokument zu erstellen. Diese Antwort hat dieselbe Wirkung wie der Befehl „Neues Dokument“ im Menü „Datei“.



Es gibt jedoch eine effizientere Methode zur Erkennung mehrerer Seiten, als die Seiten einzeln einzuscannen und dann seitenweise eine OCR-Erkennung durchzuführen: Die direkte Verarbeitung **mehrseitiger Dokumente!**

Um ein aus mehreren Seiten bestehendes Dokument in einem Durchgang einzuscannen, aktivieren Sie den Dokumenteneinzug Ihres Scanners anhand der Option „ADF“ über die „Scanner“ Schaltfläche.

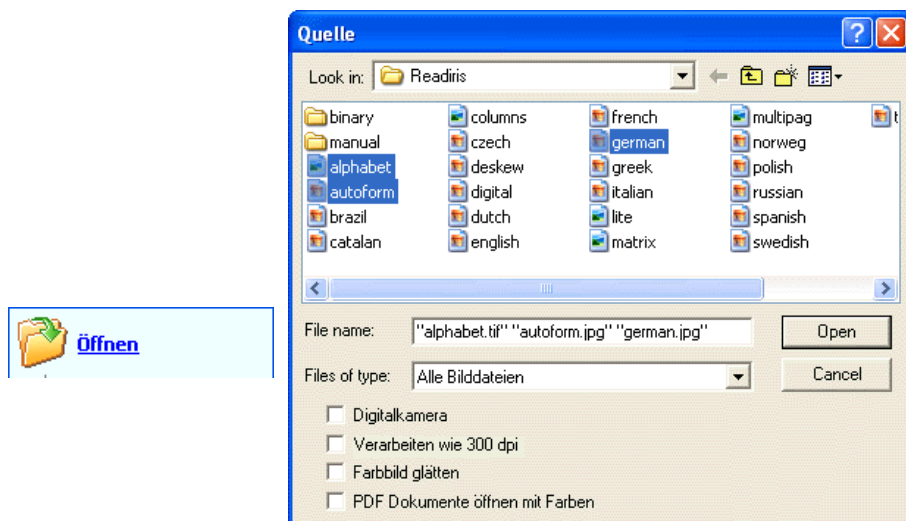


- ☐ Querformat
- ☒ ADF
- ☐ Umkehren
- ☐ Digitalkamera
- ☐ Verarbeiten wie 300 dpi
- ☐ Farbbild glätten

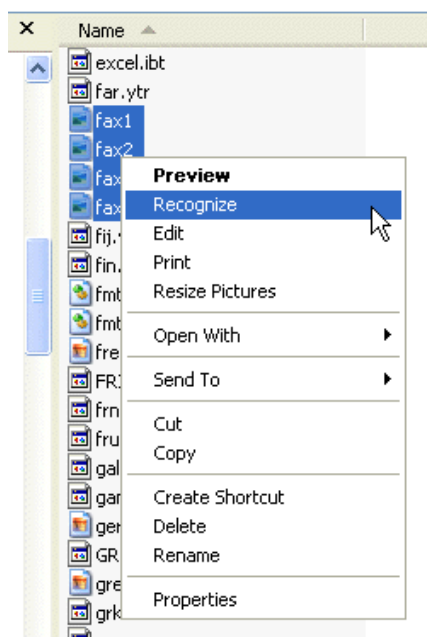
Legen Sie die Seiten Ihres Dokuments in den automatischen Dokumenteneinzug und beginnen Sie den Scanvorgang: Alle Seiten werden gescannt, bis der Dokumenteneinzug leer ist.

Sie können ebenfalls mehrere vorgescannte Bilder *öffnen*. Um mehrere Bilder zu laden, wählen Sie das erste Bild aus und halten Sie die Strg Taste gedrückt, während Sie weitere Bilder auswählen. Um eine ununterbrochene Reihe von Bildern auszuwählen, wählen Sie das erste Bild und dann bei gedrückter Umschalt Taste das letzte Bild.





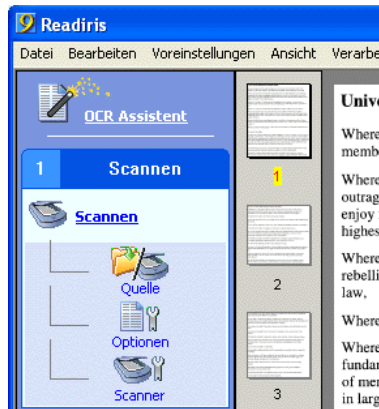
Dasselbe Ergebnis erzielen Sie, wenn Sie im Windows Explorer mehrere Bilder auswählen, die rechte Maustaste drücken und dann im „Kontext“ Menü den Befehl „Erkennen“ wählen. Sie können diese Operation wiederholen: Alle Bilder, die Sie an Readiris senden, werden an das aktuelle Dokument angehängt, bis Sie den Befehl „Neues Dokument“ wählen.



Sie können sogar mehrere vorgescannte Bilder aus dem Windows Explorer in das Readiris Fenster *ziehen*! Auch hier gilt: Alle Bilder, die Sie in das Readiris Fenster ziehen, werden an das aktuelle Dokument angehängt, bis Sie den Befehl „Neues Dokument“ wählen.

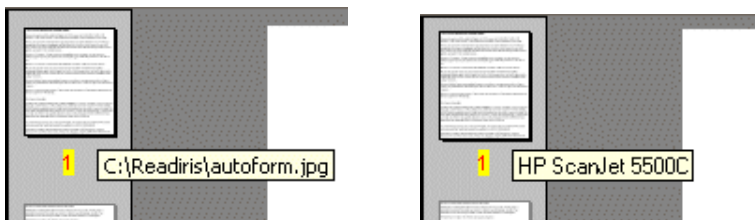
Readiris sortiert die Bilder automatisch: Bild 001.tif kommt vor 002.tif, das vor 003.tif kommt usw.

In der an der linken Seite eingeblendeten **Seiten-Toolbar** - die angezeigt wird, sobald Seiten verarbeitet werden - sind die verschiedenen Seiten des Dokuments dargestellt. Klicken Sie mit der rechten Maustaste in der Seiten-Toolbar, um auf die Seiten-Befehle zuzugreifen.



Die aktuelle Seite ist in der Seiten-Toolbar hervorgehoben und steht in der Readiris Titelleiste.

Die Seiten-Toolbar bietet einen QuickInfo: Der Benutzer kann den Mauszeiger auf einer Seiten-Ikone ruhen lassen, um zu erfahren, welches Bild in den Speicher geladen worden ist. (Wenn ein mehrseitiges Bild geöffnet worden ist, gibt es natürlich nur eine Datei für alle Bilder.) Wenn der Benutzer mehrseitige Dokumente *scannt*, gibt das QuickInfo lediglich das Scanner-Modell an.



Oder sie möchten vielleicht einen Überblick Ihres Dokumentes erzeugen? Sie können die gescannte **Bilder** schnell **drucken** mit dem Befehl "Bilder Drucken" im Menü „Datei“.

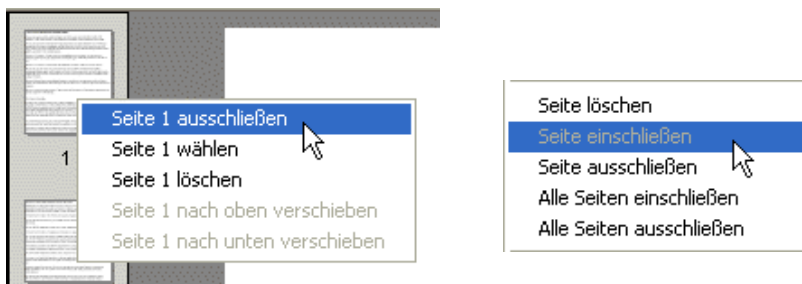


Bilder drucken... Ctrl+P

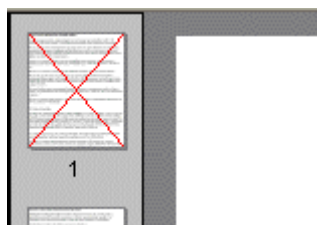
Sie können die aktuelle Seite oder alle Seiten drucken. Wählen Sie die Anzahl Seiten oder Miniaturen („Thumbnails“), die auf einer Seite gedruckt werden sollen.



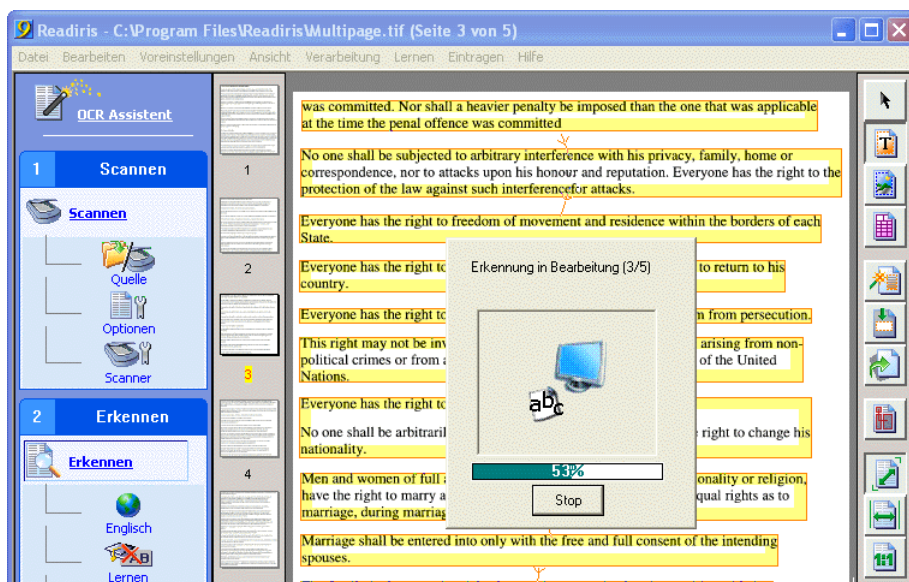
Sie müssen jedoch auch nicht alle Seiten drucken: Mit der Seiten-Toolbar (und mit dem Menü „Bearbeiten“) können Sie Seiten (zeitweilig) ausschließen. Klicken Sie mit der rechten Maustaste auf einen Seiten-Thumbnail und wählen Sie den Befehl „Seite Ausschließen“ im „Kontext“ Menü, oder zeigen Sie eine Seite an und wählen Sie den Befehl „Seite Ausschließen“ im Menü „Bearbeiten“, um sie vom Druckvorgang (bzw. Erkennungsvorgang) auszuschließen. Wählen Sie den Befehl „Seite Einschließen“, um diese Seite wieder zu berücksichtigen. Für noch flexibleres Arbeiten beinhaltet das Menü „Bearbeiten“ entsprechende Befehle für *alle* Seiten.



Die Miniaturen der ausgeschlossenen Seiten werden durchgestrichen. Bitte beachten Sie, dass die aktuelle Seite immer gedruckt werden kann, selbst wenn Sie zurzeit „deaktiviert“ ist!



Laden Sie das Beispielfeld MULTIPAGE.TIF und beginnen Sie den Erkennungsvorgang. Die Seiten werden nacheinander angezeigt, die Readiris Titelleiste zeigt die Seitennummer an.

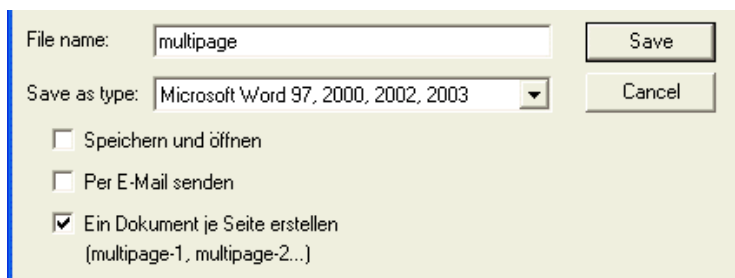


Wenn interaktives Lernen aktiviert ist, durchlaufen Sie seitenweise die Erkennungs- und Lernphase. Der Wörterbuchmodus „Neu“ wird für die erste Seite verwendet, der Modus „Zufügen“ für alle folgenden Seiten.

Wenn Sie auf die Schaltfläche „Ende“ klicken, werden alle vom System getroffenen Entscheidungen von nun an ohne benutzerseitige Bestätigung akzeptiert. Mit anderen Worten, das interaktive Lernen wird für *alle* Seiten abgebrochen, der OCR-Vorgang des Dokuments wird im automatischen Modus fortgesetzt.

Das Ergebnis der Erkennung mehrseitiger Dokumente wird in einer einzigen Ausgabedatei gespeichert. (Wenn das Erkennungsergebnis an eine Ziellanwendung gesendet wird, werden mehrere Seiten in einem einzigen Dokument erzeugt.)

Dies ist zumindest dann der Fall, wenn die Option „Ein Dokument je Seite Erstellen“ beim Speichern des erkannten Dokuments deaktiviert ist. Diese Option sorgt dafür, dass jede Seite eines mehrseitigen Dokuments in eine separate Datei gespeichert wird. Wenn der Benutzer als Dateiname „Text.doc“ wählt, werden die Dateien „Text-1.doc“, „Text-2.doc“ usw. genannt.



## **MEHRSEITIGE DOKUMENTE BEARBEITEN**

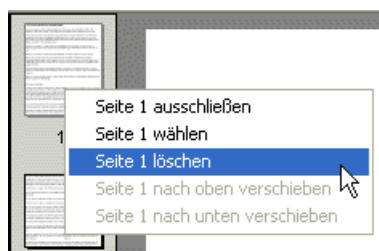
Der Benutzer kann mehrseitige Dokumente bearbeiten, insbesondere um Scanfehler zu berichtigen: Er kann Seiten aus dem Dokument löschen und Seiten an eine andere Position im Dokument verschieben.

Zuerst die Navigation: Um *zu einer Seite zu gehen*, klicken Sie auf deren Ikone in der Seiten-Toolbar oder führen Sie den Mauszeiger über den Thumbnail, drücken Sie die rechte Maustaste, um das „Kontext“ Menü zu öffnen, und wählen Sie den Befehl „Seite Wählen“. Um zur vorigen Seite zu gehen, können Sie die Taste Bildschirm-nach-oben drücken, um zur nächsten Seite zu gehen, können Sie die Taste Bildschirm-nach-unten drücken. Drücken Sie die Pos1 Taste, um zur ersten Seite zu gelangen; drücken Sie die Ende Taste, um zur letzten Seite zu gelangen. Sie können ebenfalls die entsprechenden Befehle im Menü „Ansicht“ verwenden.



Erste Seite	Home
Vorherige Seite	PageUp
Nächste Seite	PageDown
Letzte Seite	Ende

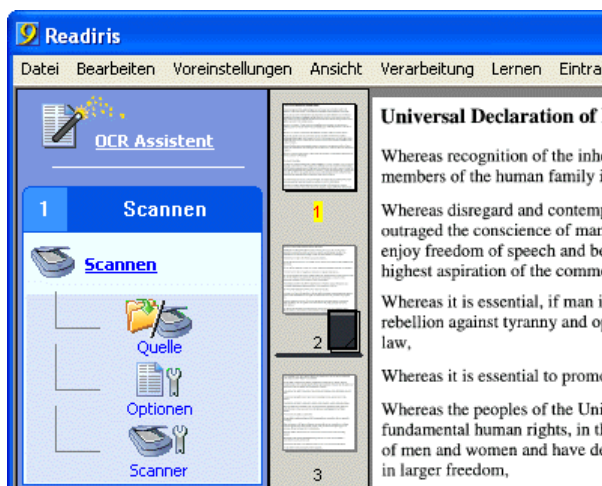
Nun zur Bearbeitung des Dokuments: Um *eine Seite* aus dem Dokument zu *löschen*, führen Sie den Mauszeiger über der Ikone der Seite, drücken Sie die rechte Maustaste und wählen Sie den Befehl „Seite Löschen“. Wir erinnern daran, dass Sie Seiten zeitweilig vom Erkennungsvorgang (und vom Druckvorgang) ausschließen können (ohne sie zu löschen): Die entsprechenden Befehle finden Sie in der Seiten-Toolbar (und im Menü „Bearbeiten“).



Um eine *Seite nach oben* zu *verschieben*, wählen Sie den Befehl „Seite nach oben Verschieben“, um eine *Seite nach unten* zu *verschieben*, wählen Sie den Befehl „Seite nach unten Verschieben“.

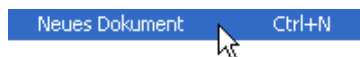
Um eine *Seite* an eine andere Position im Dokument zu *verschieben*, ziehen Sie ihre Ikone an die neue Position.





## EIN NEUES DOKUMENT STARTEN

Sie können den Befehl „Neues Dokument“ im „Datei“-Menü wählen, um das aktuelle Dokument zu schließen.



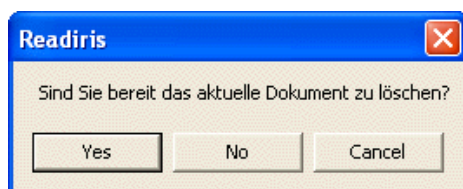
Dieser Befehl macht „reinen Tisch“. Jedes im Speicher befindliche Dokument (das eine einzige oder mehrere Seiten enthält) wird gelöscht. Sie können nun ein neues Dokument erstellen.

Sie können jedoch auch ein Dokument aus dem aktuellen Dokument heraus erstellen. Solange der OCR-Vorgang nicht ausgeführt worden ist, geht das System davon aus, dass Sie Seiten zum aktuellen Dokument hinzufügen möchten. Sie können beispielsweise alle Seiten im automatischen Dokumenteneinzug des Scanners einscannen, den Einzug erneut füllen und von vorn beginnen. Alle gescannten Seiten bilden ein einziges Dokument. Sie können auch eine Reihe



Seiten scannen und mehrere Bilddateien - beispielsweise Faxe - hinzufügen. Diese Seiten bilden wiederum ein einziges Dokument, Sie müssen lediglich mit der „Quelle“ Schaltfläche die Bildquelle ändern.

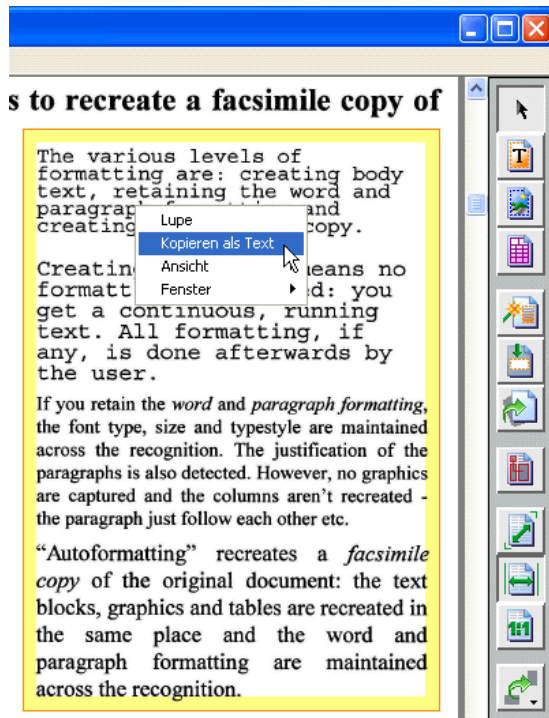
Wenn der OCR-Vorgang bereits ausgeführt wurde und Sie den Scanvorgang (oder das Laden von Bildern) erneut starten, fragt das Programm nach, ob ein neues Dokument angelegt oder das aktuelle Dokument ergänzt werden soll.



## TEXTZONEN ERKENNEN

---

Wir wissen nun, wie Seiten erkannt und wie mehrseitige Dokumente verarbeitet werden. Können wir aber einen Teil einer Seite ebenso komfortabel erkennen? Ja, das können wir! Klicken Sie die rechte Maustaste und wählen Sie den Befehl „Kopieren Als Text“ im „Kontext“ Menü: Das Textfenster unter dem Mauszeiger wird erkannt und in die Zwischenablage gesetzt.



Die aktuellen Systemeinstellungen - Sprache, Schrifttyp, usw. - gelten weiterhin. Das OCR-Ergebnis wird als „kontinuierlicher“, unformatierter Text in die Zwischenablage gesetzt.

## DIE AUSSICHT DER TEXTAUSGABE

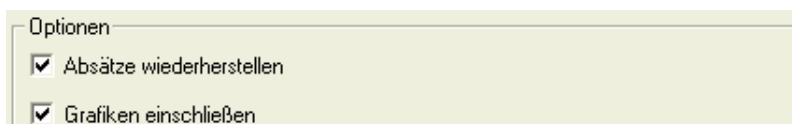
Den Text zu speichern oder zu exportieren, ist mehr als nur einen Ausgabemodus zu wählen, oder einen Dateinamen für die Ausgabedatei zu definieren. Sie wählen



auch ein Dateiformat und bestimmen das Erscheinungsbild des erkannten Texts. Kurz, Sie sollten schon, bevor Sie den Prozess starten, entscheiden, wo Sie den Text hinbringen wollen und was Sie damit tun wollen.

Sie können den Anschein von Ihrer Textausgabe mit einigen Optionen des „Format“ Befehls beeinflussen.

Der **Textfluß** des Ausgabedokumentes kann direkt durch die Option „Absätze Wiederherstellen“ beeinflußt werden.



Mit dieser Option kann Readiris automatisch die Absätze ermitteln: Readiris setzt nur einen festen **Zeilenumbruch** am Ende eines Absatzes - wie beim Wordprozessor normal - und nicht am Ende einer Zeile. Die Option „Absätze Wiederherstellen“ gehört zu den Standardeinstellungen.

Ein Beispiel zur Verdeutlichung. Die ersten drei Zeilen in einem Absatz lauten: „Der neue Präsi-“, „dent stand auf dem Balkon.“ und „Seine Frau kam dazu.“. Die Absatzkontrolle schreibt den Satz wie folgt: „Der neue **Präsident** stand auf dem Balkon. **Seine** Frau kam dazu.“ Die getrennten Silben des Wortes „Präsident“ wurden wieder zusammengefügt. Außerdem wurde ein Leerraum am Ende des Satzes zugefügt. Somit wurde das Aussehen des Textes beeinflußt und ein natürlicher Textfluß erstellt.

Ansonsten hätte Readiris eine feste Zeilenwechselmarke („Wagenrücklauf/ Neue Zeile“) am Ende der Zeile und nicht am Ende des Absatzes gesetzt.

(Die Schaltfläche „Format“ enthält noch mehr Optionen. Mehr darüber später.)

## SCANNEREINSTELLUNG

---

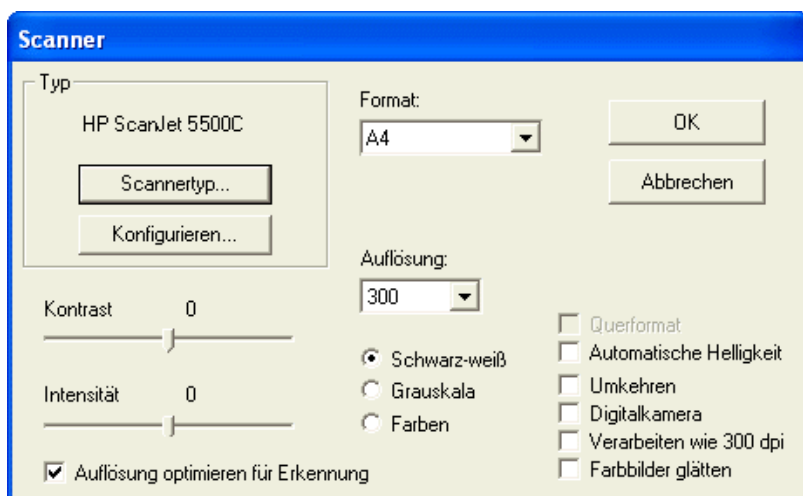
Jetzt werden wir den Scanner einstellen. Wichtige Voraussetzung ist, dass der Scanner und die erforderlichen Treiber richtig installiert wurden.

Wenn Ihr Scanner zusammen mit Ihrer Readiris Software lizenziert wurde, brauchen Sie hier wahrscheinlich nichts mehr zu machen. Ihr Scanner ist dann oft bereits unter Readiris installiert.

Klicken Sie auf die Schaltfläche „Scanner“ in der Haupttoolbar.



Klicken Sie auf das Schaltfläche „Scannertyp“, um das **Scannermodell** zu bestimmen.

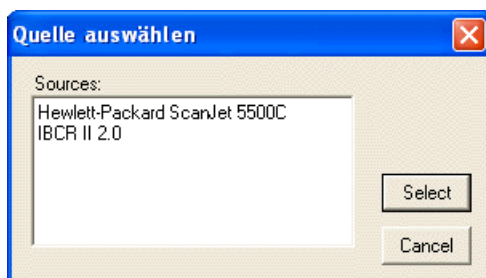


Wenn Sie „<Image>“ (Bild) als „Scanner“ selektieren, dann dienen vorgescannte Bilder immer als Bildquelle. In diesem Fall brauchen Sie die Schaltfläche „Quelle“ nicht mehr zu aktivieren, um die Festplatte als Bildquelle zu aktivieren.

Die Schaltfläche „Konfigurieren“ ist nur aktiviert, wenn der Scanner es erlaubt. Es ermöglicht den Zugriff zu einigen fortschrittlichen Scanparametern: Bei Twain



Scannern können Sie die Twain Quelle selektieren, indem Sie auf den Knopf „Konfigurieren“ klicken. (Sie können ebenfalls den Befehl „Quelle Auswählen“ im Menü „Datei“ anklicken.)



Nachdem der Scanner gewählt wurde, können Sie im gleichen Fenster die Scanauflösung, das Seitenformat und die Seitenorientierung, die Helligkeit und den Kontrast bestimmen und andeuten, wenn Sie den automatischen Dokumenteingabe Ihres Scanners verwenden möchten. Bei Scannern, die mit Twain kompatibel sind, werden diese Parameter oft innerhalb der Twain-Schnittstelle bestimmt.

Stellen Sie die **Helligkeit** und, wenn vorhanden, den **Kontrast** ein.

Durch Aktivierung der Option „Querformat“ geben Sie an, dass die gewählte Seitenorientierung breit ist anstatt schmal. Die Seitenorientierung kann nur auf reduzierten Seitenformaten angewendet werden: Mit einem A4 Flachbettscanner können A5 Texte (halb so groß) gescannt werden im Hoch- und Querformat. Die A4 Seite kann jedoch nur in einer Richtung gescannt werden!



Schließlich ermöglicht es die Option „Umkehren“, im Schwarzweiß-Scanmodus **Negativbilder** zu erzeugen. Sie möchten diese Option aktivieren zur Verarbeitung von ganzen Seiten mit weißem Text auf einem schwarzem Hintergrund.

## **BRINGEN SIE FARBE IN IHRE TEXT-SCANS!**

---

Readiris unterstützt Schwarzweiß-, Graustufen- und Farbbilder gleichermaßen. Sie können also den **Farbmodus** wählen, der Ihren Anforderungen am besten entspricht. Um Schwarzweiß-Grafiken in die erkannten Dokumente einzubinden, scannen Sie im Schwarzweiß-Modus, um Schwarzweiß-Fotos einzubinden, scannen Sie im Graustufen-Modus und um Farbbilder einzubinden, scannen Sie im Farbmodus.

Aber warum sollten Sie die Bittiefe von Bildern während des Scanvorgangs herabsetzen? Selbstverständlich werden Graustufen- und Farbbilder langsamer eingelesen und benötigen mehr RAM als Schwarzweißbilder.

Scannen in Graustufen und Farben ist nicht nur nützlich, um die Grafiken in ausreichender Qualität zu speichern, sondern in bestimmten Fällen auch nützlich oder erforderlich, um gute OCR-Ergebnisse zu erzielen. Wenn Text auf farbigem Hintergrund gedruckt ist, kann es bei einem Farbscan Farbtonunterschieden bekommen, die in Schwarzweißbildern fehlen. Wenn zwischen Text und Hintergrund nur ein geringfügiger Kontrast besteht, kann der Hintergrund zu „Rauschen“ führen, das die Erkennung schwierig oder gar unmöglich macht!

Denken Sie beispielsweise an schwarzen Text, der auf einem dunklen Hintergrund gedruckt ist: Wenn Sie ein solches Dokument im Schwarzweiß-Modus scannen, können Sie die Hintergrundfarbe ggf. nicht „fallen lassen“, ohne auch die Textinformationen zu verlieren, wie auch immer Sie die Helligkeit regeln...



**MASAYOSHI SON**, 42, president and CEO, is the master Net empire builder. His conglomerate holds stakes in 300 Internet companies in the U.S., Japan, Europe, and other Asian countries. Today, Softbank manages about \$4 billion in venture capital funds for global investments.

**YASUMITSU SHIGETA**, 35, has invested in more than 70 Web or mobile Net-based ventures in Japan and the U.S., including Tumblweed Communications and Phone.com. Shigeta is also developing new businesses that take advantage of the growth of the Internet and mobile communications.

**MASAYOSHI SON**, 42, president and CEO, is the master Net empire builder. His conglomerate holds stakes in 300 Internet companies in the U.S., Japan, Europe, and other Asian countries. Today, Softbank manages about \$4 billion in venture capital funds for global investments.

**YASUMITSU SHIGETA**, 35, has invested in more than 70 Web or mobile Net-based ventures in Japan and the U.S., including Tumblweed Communications and Phone.com. Shigeta is also developing new businesses that take advantage of the growth of the Internet and mobile communications.

Readiris erzeugt eine Schwarzweiß-Fassung jedes Graustufen- und Farbbilds. Dank intelligenter Programmroutinen können selbst schwierige Fälle gelöst werden - und so wird unseres „schwieriges“ Bild digitalisiert!

**MASAYOSHI SON**, 42, president and CEO, is the master Net empire builder. His conglomerate holds stakes in 300 Internet companies in the U.S., Japan, Europe, and other Asian countries. Today, Softbank manages about \$4 billion in venture capital funds for global investments.

**YASUMITSU SHIGETA**, 35, has invested in more than 70 Web or mobile Net-based ventures in Japan and the U.S., including Tumblweed Communications and Phone.com. Shigeta is also developing new businesses that take advantage of the growth of the Internet and mobile communications.

Um ein gescanntes Bild schwarzweiß zu sehen, deaktivieren Sie die Option „Dokument Farbig Anzeigen“ im „Ansicht“ Menü.

✓ Dokument farbig anzeigen Ctrl+O





Wenn diese Option aktiviert ist, sehen Sie keine Schwarzweißbilder an Ihrem Bildschirm - selbst wenn Sie tatsächlich Monochrombilder scannen! Dies ist darauf zurückzuführen, dass die Option „Qualitativ Hochwertige Anzeige“ im Menü „Ansicht“ die Bilder für eine optimale Lesbarkeit am Bildschirm optimiert.

✓ Qualitativ hochwertige Anzeige



Dieses spezielle hochauflösende Anzeigeverfahren wandelt Schwarzweißbilder in Graustufenbilder um.

### Reading dot matrix documents

You can read dot matrix document without changing the font mode. The software detects whether "normal" text or dot matrix printouts are being read.

Far out in the uncharted backwaters of the unfashionable end of the Western Spiral arm of the Galaxy lies a small unregarded yellow sun. Orbiting this at a distance of roughly ninety-two million miles is an utterly insignificant little blue green

### Reading dot matrix documents

You can read dot matrix document without changing the font mode. The software detects whether "normal" text or dot matrix printouts are being read.

Far out in the uncharted backwaters of the unfashionable end of the Western Spiral arm of the Galaxy lies a small unregarded yellow sun. Orbiting this at a distance of roughly ninety-two million miles is an utterly insignificant little blue green

Graustufen- und Farbbilder werden geglättet und weichgezeichnet.

### A word about OCR

The aim of OCR is to automatically enter printed text documents in a very effective and low cost way. Although the first research and development on Optical Character Recognition (OCR) began more than 30 years ago, this technology is still unknown by most of the people who could use it for their document entry applications.



## A word about OCR

The aim of OCR is to automatically enter printed text documents in a very effective and low cost way. Although the first research and development on Optical Character Recognition (OCR) began more than 30 years ago, this technology is still unknown by most of the people who could use it for their document entry applications.

Daher ist eine Vergrößerung nicht erforderlich, selbst nicht auf Laptops mit LCD-Bildschirm oder Desktop-Rechnern mit einem billigen 14" Bildschirm. Die hohe Anzeigequalität ist standardmäßig aktiviert, kann jedoch auf hochauflösenden Computerbildschirmen überflüssig sein.

## ANDERE GERÄTE, ANDERE AUFLÖSUNG

Unabhängig vom Scanmodus wählen Sie für normale Anwendungen eine **Auflösung** von 300 dpi, für eine kleinere Schriftart (unter 10 Punkt) und bei einer schlechten Qualität des Dokumentes wählen Sie eine höhere Auflösung von 400 dpi.

Readiris ist in der Lage, Zeichen mit einer **Größe** von 6 bis 72 Punkte (0,21 bis 2,54 cm.) zu lesen.

6 Punkt

# 72 Punkt

Readiris kann auch **Initialen** („Drop Letters“) erkennen. „Initialen“ sind große Buchstaben, die über mehreren Zeilen gehen. (Sie können natürlich nie größer sein als 72 Punkt!)

**R**eadiris erkennt  
"Initialen" - das  
sind große  
Buchstaben die über  
mehreren Zeilen gehen.

Da optimale OCR eine Auflösung von 300 bis 400 dpi voraussetzt, warnt Readiris Sie, wenn Sie Bilder mit einer Auflösung unter 200 dpi oder über 800 dpi vorlegen. Readiris kann jedoch zu detailreiche Scans für Sie korrigieren! Aktivieren Sie hierzu die Option „Auflösung Optimieren für Erkennung“ in den Scaneinstellungen. Wenn die Bildauflösung Ihrer Scans über 600 dpi liegt, wird die Auflösung für den OCR-Vorgang reduziert.

☒ Auflösung optimieren für Erkennung

Es gibt andere Methoden, diese Warnung zu unterdrücken: Wenn Sie **Faxe** einlesen - die eine Auflösung von 100 oder 200 dpi aufweisen -, wenn Sie Bilder mit einer Digitalkamera erzeugen - deren Auflösung unbekannt ist - und wenn Sie Bilder öffnen, in deren Datei-Header eine falsche Auflösung angegeben ist.

Um solche Bilder problemlos zu verarbeiten, aktivieren Sie die Option „Verarbeiten wie 300 dpi“. Diese Einstellung gilt sowohl für direkte Scanvorgänge wie auch für das Öffnen vorgescannter Bilder.

☐ Umkehren  
☐ Digitalkamera  
☒ Verarbeiten wie 300 dpi  
☐ Farbbild glätten

☐ Digitalkamera  
☒ Verarbeiten wie 300 dpi  
☐ Farbbild glätten  
☐ PDF Dokumente öffnen mit Farben

Wenn Ihre Bilder von einer **Digitalkamera** statt von einem Scanner stammen, müssen Sie unbedingt eine Sonderoption aktivieren. Dieser Parameter gilt wiederum für direkte Scanvorgänge wie auch für vorgescannte Bilder.



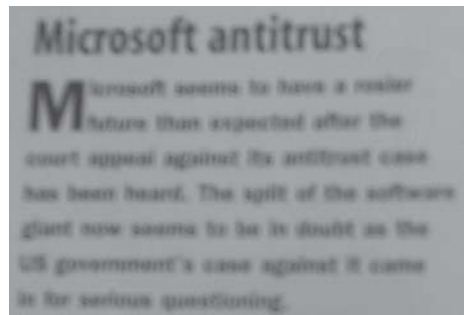
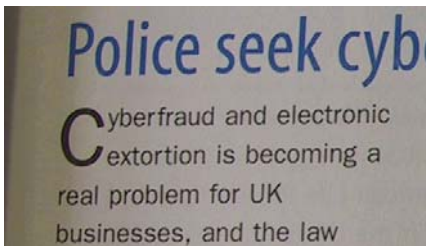
<input checked="" type="checkbox"/>	Umkehren
<input type="checkbox"/>	Digitalkamera
<input checked="" type="checkbox"/>	Verarbeiten wie 300 dpi
<input type="checkbox"/>	Farbbild glätten

<input checked="" type="checkbox"/>	Digitalkamera
<input checked="" type="checkbox"/>	Verarbeiten wie 300 dpi
<input type="checkbox"/>	Farbbild glätten
<input type="checkbox"/>	PDF Dokumente öffnen mit Farben

Hierdurch verbessern Sie das Bild, bevor es erkannt wird. Bei Digitalkameras sind spezifische Herausforderungen zu bewältigen: Sie erzeugen Bilder mit geringer Auflösung - selbst wenn Sie die Kamera sehr nahe an das Dokument halten - und die Bildauflösung ist auf jeden Fall unbekannt.

Dies sind die Einzelheiten, die Sie bei der erfolgreichen Erkennung von Bildern, die mit einer Digitalkamera aufgenommen wurden, beachten sollten.

Zum ersten sollten Sie die höchstmögliche Bildauflösung wählen. Erzeugen Sie zum Beispiel Bilder im Format 2.048 x 1.536, auch wenn ebenfalls Bilder im Format 1.024 x 768 und 640 x 480 unterstützt werden. Zum anderen sollten Sie den „Makro“-Modus der Kamera aktivieren, um Nahaufnahmen anzufertigen - was bei der Aufnahme von Dokumenten immer der Fall ist. (Dieser Modus wurde für Aufnahmen von Blumen, Insekten usw. entwickelt.) Andernfalls sind die Bilder unscharf und nicht lesbar.

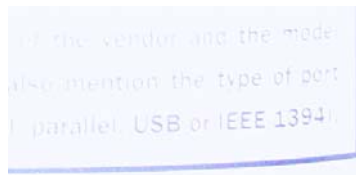


Wählen Sie keine oder eine möglichst geringe Kompression: Eine umfangreiche Kompression reduziert die Schärfe des erfassten Texts. Zoomen Sie manuell, um Ihr Dokument zuzuschneiden - einige Kameras werden mit einer Software zum

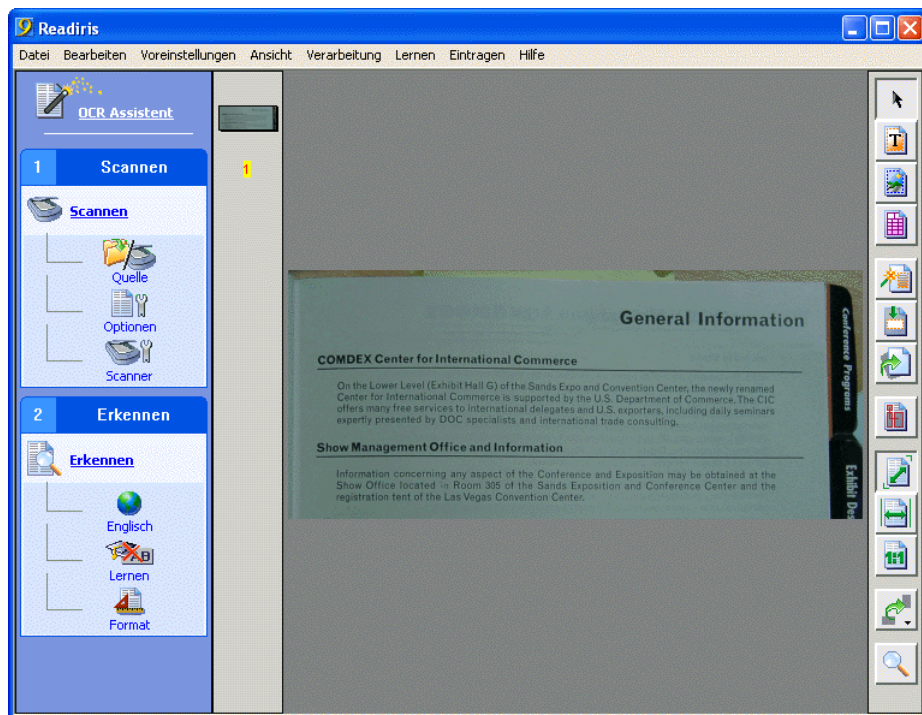
Aneinanderfügen von Bildern („photo stitching“) geliefert; Sie sollten diese Software jedoch nicht für das Dokument-Capturing verwenden.

Halten Sie die Kamera direkt über das Dokument, um das Dokument nicht aus einem schrägen Winkel zu fotografieren. Sie sollten vermeiden, dass die Kamera oder Ihre Hände Schatten auf das Dokument werfen! Erzeugen Sie unverwackelte Bilder. Bei Bedarf montieren Sie die Kamera auf einen Standfuß.

Desaktivieren Sie das Blitzlicht, wenn Sie von Glanzpapier aufnehmen, andernfalls ist das Bild ggf. zu hell. Ganz allgemein sollten Sie die Helligkeit und den Kontrast an das Umgebungslicht anpassen - Tageslicht, Lampenlicht, Neonlicht usw. (Einige Kameras können kalibriert werden, indem man ein weißes Dokument aufnimmt.)



Zum Ausprobieren können Sie das Bild DIGITAL.JPG im Readiris Ordner öffnen und die Erkennung ausführen.



## STANDARD EINSTELLUNGEN SPEICHERN

Wenn alle Scanparameter richtig eingestellt wurden, klicken Sie auf den Befehl „Standard-Parameter Speichern“ im Menü „Datei“. Die laufenden Einstellungen werden als Standard **Einstellungen** gespeichert. Bei künftigen Aufrufen arbeitet Readiris mit den Einstellungen weiter.

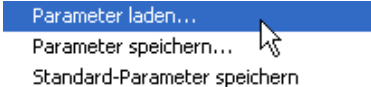


Die Einstellungsdateien enthalten mehr als nur die **Scannereinstellungen**. Sie können auch bestimmen, ob Sie die interaktive Lernfunktion benutzen werden, welche Sprache im Dokument verwendet wird, welche Art der Ausgabe verwendet wird - zum Beispiel Text an WordPad senden - usw. Kurz, *alle* operativen Einstellungen von Readiris werden in den Einstellungsdateien gespeichert.

## VORDEFINIERT EINSTELLUNGEN SPEICHERN

---

Bei jedem Start werden die Standard Einstellungen geladen. Um zu vermeiden, dass Sie die eingestellten Parameter ständig anpassen müssen, können Sie auch unterschiedliche Einstellungen speichern. Die Befehle sind „Parameter Speichern“ und „Parameter Laden“. Sie befinden sich im Menü „Datei“.



Ein Beispiel: Wenn Sie regelmäßig deutsche Dokumente mit einem spezifischen Layout scannen müssen, ist es zu empfehlen, eine Ausgabedatei zu erstellen, die auf diesen speziellen Dokumententyp Anwendung findet. Sie müssten dann „Deutsch“ als Dokumentensprache selektieren. Damit Sie die Fenster nicht immer neu definieren müssen, ist es ratsam, eine Vorlage zu laden. Schalten Sie den Lernmodus, aus aber aktivieren Sie ein Benutzerwörterbuch im „Verwenden“-Modus, weil die gleichen Schriftformen systematisch wieder angewendet werden usw.

Um eine Übersicht über die eingestellten Parameter zu bekommen, ist es unnötig die verschiedene Menübefehle zu überlaufen. Sie brauchen nur den Befehl „Information“ im Menü „Datei“ zu aktivieren.



**Einstellungen**

Scanner	
Typ	HP ScanJet 5500C
Auflösung	300 dpi
Format	A4
Modus	Schwarz-weiß
Querformat	Aus
Text	
Format	Microsoft Word 97 / 2000 / 2002 / 2003
Paragraph	Ein
Layout	Originaldokument wiederherstellen
Dokument	
Schriftart	Automatisch
Sprache	Englisch
Seite	
Auflösung	300 dpi

OK

## DOKUMENTE SCANNEN

Nachdem der Scanner nun eingerichtet ist, möchten wir mit dem Scannen von Dokumenten beginnen. Es gibt einige Faktoren, die Sie beachten sollten.

Zunächst sollten Sie auf die **Zeilenschräglage** achten. Obgleich die Seitenanalyse und die Erkennung schräglagentolerant sind, kann es schwierig sein, eine Seite korrekt zu unterteilen und per OCR zu erkennen, wenn die Schräglage zu groß ist. Eine geringfügige Zeilenschräglage (weniger als  $0,5^\circ$ ) kann ignoriert werden, da die Genauigkeit des OCR-Vorgangs hierdurch nicht beeinträchtigt wird.

Die Option „Seitenschräglage Korrigieren“ (unter der Schaltfläche „Optionen“ oder im Menü „Voreinstellungen“) legt fest, ob die Schräglage von Seiten, die in Schräglage eingescannt wurden, korrigiert wird - eine geringfügige Zeilenschräglage wird ignoriert. Diese Option ist standardmäßig deaktiviert.





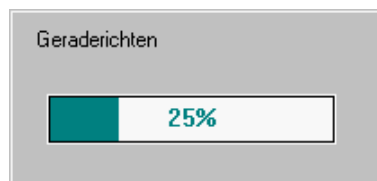
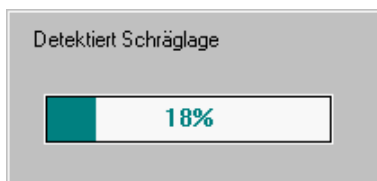
Seitenschräglage korrigieren

Wurde diese Option fälschlicherweise deaktiviert, verwenden Sie die Schaltfläche „Schräglage Seite Korrigieren“ in der Bild-Toolbar (oder die entsprechende Option im Menü „Verarbeitung“), um Seiten zu begradigen, die schräg eingescannt wurden.

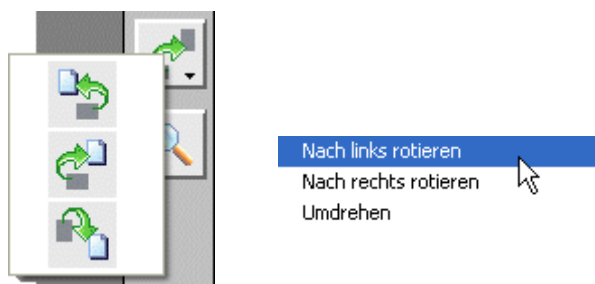


Schräglage Seite korrigieren

Die Korrektur der Schräglage dauert einige Sekunden: as Bild wird analysiert, um den Winkel festzustellen, das Farb- oder Graustufenbild und dessen Schwarzweiß-Fassung werden korrigiert und die Seitenanalyse wird erneut ausgeführt.



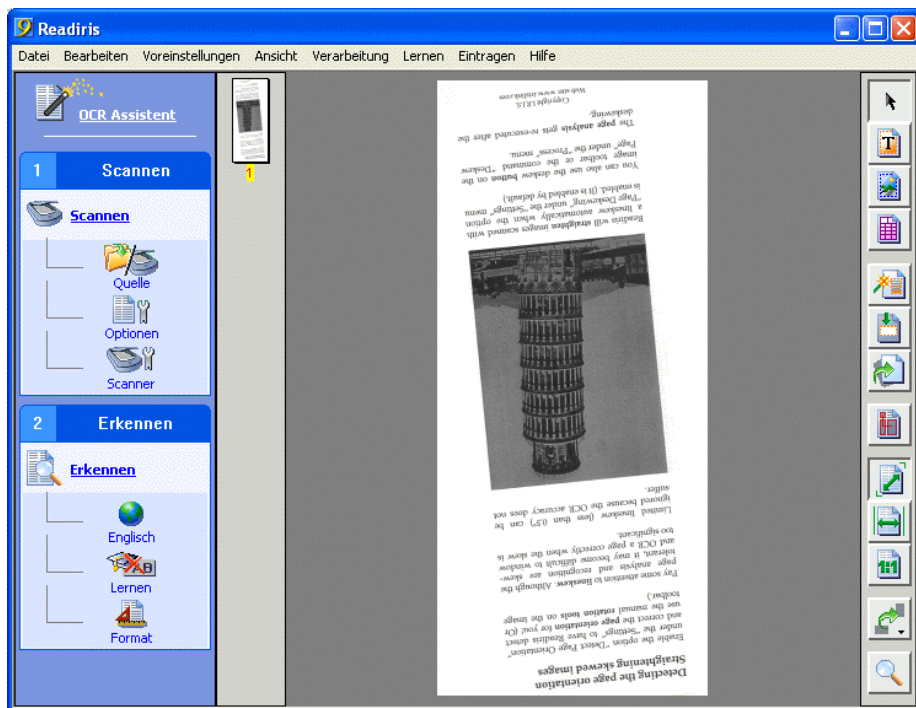
Gegebenenfalls müssen Sie auch die **Seitenorientierung** korrigieren. Verwenden Sie hierzu die Drehung-Schaltflächen der Toolbar. (Die entsprechenden Befehle stehen auch im Menü „Ansicht“.) Es gibt drei Drehrichtungen: Nach links, nach rechts und Drehung um 180°. Die Rotation erfordert einige Sekunden, weil das Bild selbst konvertiert wird und nicht nur das Display auf dem Schirm.



Readiris kann jedoch falsch ausgerichtete Seiten für Sie korrigieren. Aktivieren Sie die Option „Seitenorientierung Entdecken“ (mit der Schaltfläche „Optionen“ oder im Menü „Voreinstellungen“), damit Readiris die Seitenausrichtung bei Bedarf korrigiert.



Sie können das Bild DESKEW.JPG im Readiris Ordner verwenden, um diese Funktion auszuprobieren. Aktivieren Sie die Optionen „Seitenschräglage Korrigieren“ und „Seitenorientierung Entdecken“, bevor Sie das Bild öffnen, und lassen Sie Readiris den Turm von Pisa wieder so aufrichten, wie wir ihn mögen!



## JUSTIEREN DER GESCANNTEN BILDER

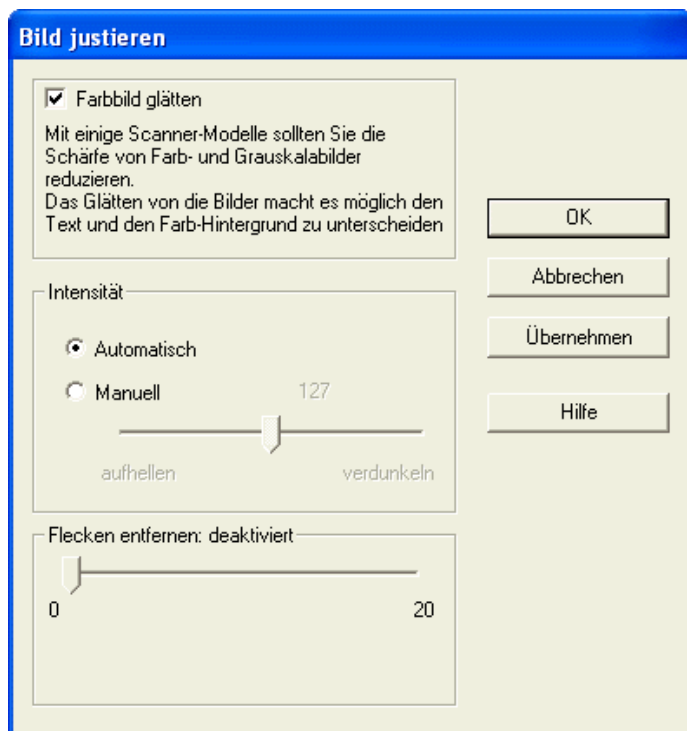
Wie bereits gesagt, konvertieren leistungsfähige Routinen Farb- und Graustufenbilder automatisch in Schwarzweißbilder. Sollte dies noch erforderlich sein, kann der Benutzer das Bild im Hinblick auf den OCR-Vorgang optimieren. Wählen Sie den Befehl „Bild Justieren“ im Menü „Verarbeitung“.



Bild justieren... Ctrl+J



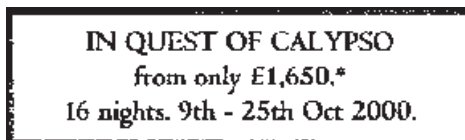
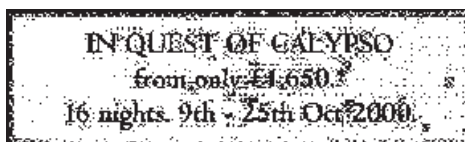
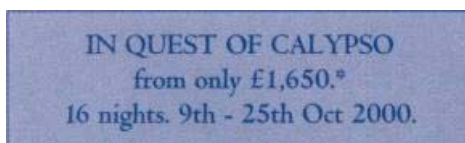
Wenn Sie auf diesen Befehl zugreifen, wird automatisch die Schwarz-weiß-Fassung angezeigt. (So, als hätten Sie die Option „Dokument Farbig Anzeigen“ deaktiviert!) Hier müssen wir näher auf einige komplizierte Konzepte eingehen.



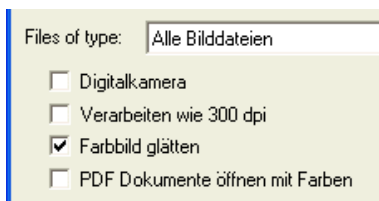
Die Option „Farbbild Glätten“ macht Graustufen- und Farbbilder homogener, indem relative Intensitätsunterschiede „abgeflacht“, „geglättet“ werden. Dies führt

zu einem schärferen Kontrast zwischen dem Vordergrund (dem Text) und dem Hintergrund (Farbe, Verzierungen usw.).

Diese **Vorverarbeitungsfunktion** kann höchst technisch und schwer verständlich erscheinen, spielt jedoch durchaus eine wichtige Rolle: Bei einigen Scannermodellen wird diese Herabsetzung der Schärfe benötigt, um Farb- und Graustufenbilder zu erkennen. Die Glättung ist manchmal die einzige Methode, Text vom farbigen Hintergrund zu unterscheiden! Hierunter sehen Sie ein Bild, das ohne Bildglättung ganz einfach unleserlich ist.



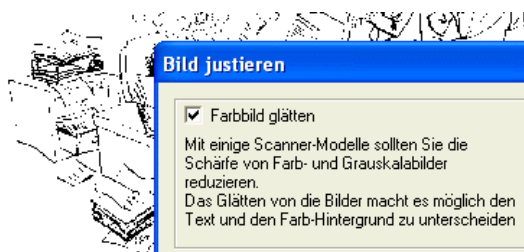
Die Bildglättung kann ebenfalls aktiviert werden, wenn Sie vorgescannte Bilder in den Speicher laden!





Und jetzt die Intensität. Unter „Intensität“ versteht man eigentlich den Schwarzweiß-Schwellenwert. Die Einstellung „Automatisch“ legt den Schwarzweiß-Schwellenwert automatisch fest. Legen Sie bei Bedarf einen anderen Schwellenwert fest, indem Sie das Schwarzweißbild dunkler oder heller machen. Wenn Sie das Bild dunkler machen, werden in der Schwarzweiß-Fassung mehr Pixel schwarz gefärbt; Wenn Sie das Bild aufhellen, werden in der Schwarzweiß-Fassung mehr Pixel weiß gefärbt.

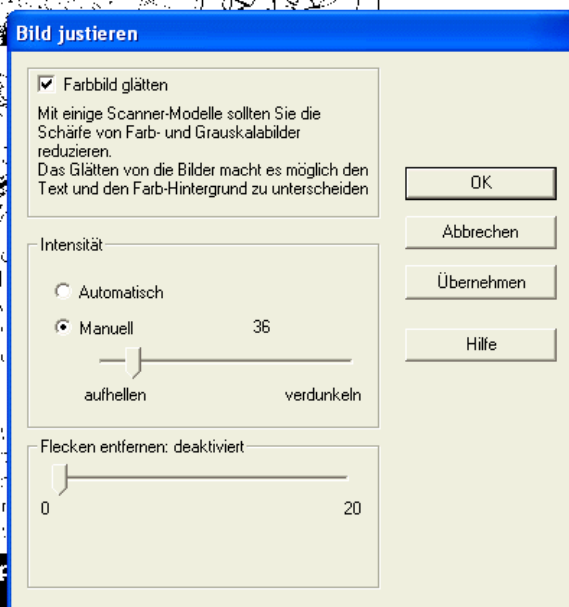
Bitte beachten Sie, dass keine Bildanpassung ausgeführt wird, bevor Sie auf „Übernehmen“ klicken! Wenn Sie auf „OK“ klicken, wird die Anpassung vorgenommen *und* das Fenster geschlossen. Hier ein Beispiel, in dem wir das Schwarzweißbild wesentlich aufgehellt haben - dabei stand die OCR-Präzision allerdings nicht im Vordergrund!



Der Scanner liest das Dokument und übersendet die Bilddaten an den Computer. Die Aufgabe des Computers ist es, den Hintergrund des Dokuments zu entfernen und die Textinformationen zu erkennen.

Dazu benutzt das System intelligente Algorithmen, um neue Lösungen bei schwierigen Fällen zu finden. Dies macht die Faserarbeit einfacher und die Software stets in der Lage, neue Lösungen zu finden.

Copy

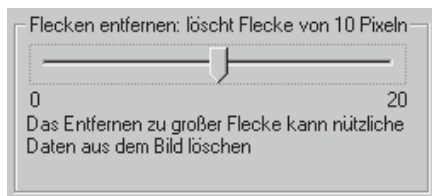


Die ersten beiden Optionen betreffen Farb- und Graustufenbilder, die letzte Option, „Flecken Entfernen“, ausschließlich Schwarzweißbilder. Als „Flecken entfernen“ bezeichnet man die Entfernung von „Parasiten-Pixels“ aus Schwarzweißbildern.

**If computers can't  
adapt easily, then  
maybe the people  
using them can.**

**If computers can't  
adapt easily, then  
maybe the people  
using them can.**

Stellen Sie sicher, dass Sie keine zu großen Flecken löschen, da andernfalls Pünktchen auf dem „i“, Teile von Punktmatrix-Buchstaben usw. verschwinden!



Die beste Methode zur Optimierung von Bildern für den OCR-Vorgang besteht darin, das Einstellungsfenster so am Bildschirm zu positionieren, dass Sie die Auswirkungen der Einstellungen, die Sie vornehmen, korrekt beurteilen können. Passen Sie die Parameter ein und klicken Sie jeweils auf „Übernehmen“, bis das Bild klar und scharf ist.

## **LASSEN SIE DEN OCR ASSISTENTEN FÜR SICH ARBEITEN**

Beginnen wir nun, Dokumente zu erfassen. Nachdem der Scanner eingestellt wurde, wollen wir damit anfangen, Dokumente zu erfassen. Damit Sie nicht durch



alle Parameter müssen, benutzen wir den **OCR Assistenten**, die bequemste Art, Seiten zu erkennen.

Klicken Sie auf „OCR Assistent“ in der Haupttoolbar (oder wählen Sie den Befehl „OCR Assistent“ im Menü „Verarbeitung“).



Der Assistent führt Sie bequem durch den Erkennungsprozess: Beantworten Sie einige einfache Fragen und Sie bekommen schnelle und einfache Ergebnisse mit Readiris.



Beachten Sie, dass der OCR Assistent bei jedem Starten von Readiris aktiviert wird. Sie können dieses vermeiden, indem Sie die Option „Assistent beim



Programmstart Aktivieren“ im ersten Fenster des Assistenten deaktivieren (und auch mit der entsprechenden Option im Menü „Voreinstellungen“).

## READIRIS ERSTELLT DAS LAYOUT

---

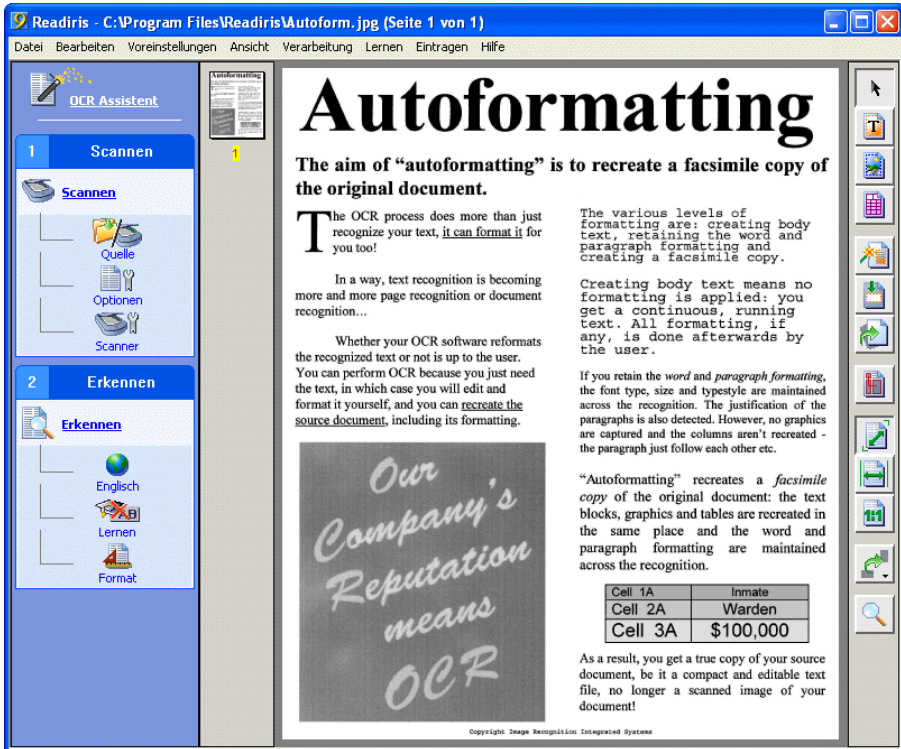
Der OCR-Assistent automatisiert den Erkennungsprozess weitgehend; „automatische“ OCR darf jedoch *nicht* mit automatischem Formatieren verwechselt werden! Der automatische Erkennungsmodus ermöglicht den automatischen Erkennungsprozeß. „Autoformatierung“ bedeutet, dass Readiris eine **Faxkopie** des gescannten Dokumentes erstellt. Die Wort-, Absatz-, und Seitenformatierung vom Originaldokument werden übernommen.

Ähnliche Schriftarten (mit und ohne Serifen, proportional und feste Breite, normal und verdichtet) werden wie im Originaldokument verwendet. Schriftgröße und Schriftauszeichnungen (fett, kursiv, unterstrichen, hochgestellt und tiefgestellt) werden ebenfalls bei der Erkennung aufrechterhalten. Tabulatoren und Einzüge (links, zentriert, rechts und Blocksatz) von jedem Textblock werden wiederhergestellt. Gleiches gilt für Aufzählungszeichen und Nummerierung. Alle E-Mail-Adressen und URLs von Webseiten werden erkannt und in der Ausgabe als Hyperlinks wiederhergestellt. Die Position der Spalten, Textblöcke und Grafiken entspricht dem Originaldokument.

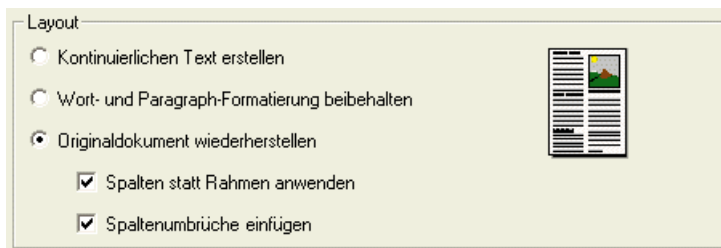
Mit anderen Worten: Readiris ist in der Lage, eine originalgetreue Kopie Ihres Originaldokuments als eine kompakte Textdatei und nicht wie eine Bilddatei zu speichern!

All das bedeutet, dass die Fenstersortierung nur *teilweise* arbeitet, wenn „Autoformatierung“ aktiviert wurde. Sie können Zonen löschen oder hinzufügen, aber eine neue Reihenfolge der Zonen wird einfach ignoriert!

Ein Beispiel zur Verdeutlichung. Öffnen Sie das Bild AUTOFORM.JPG im Readiris Ordner.

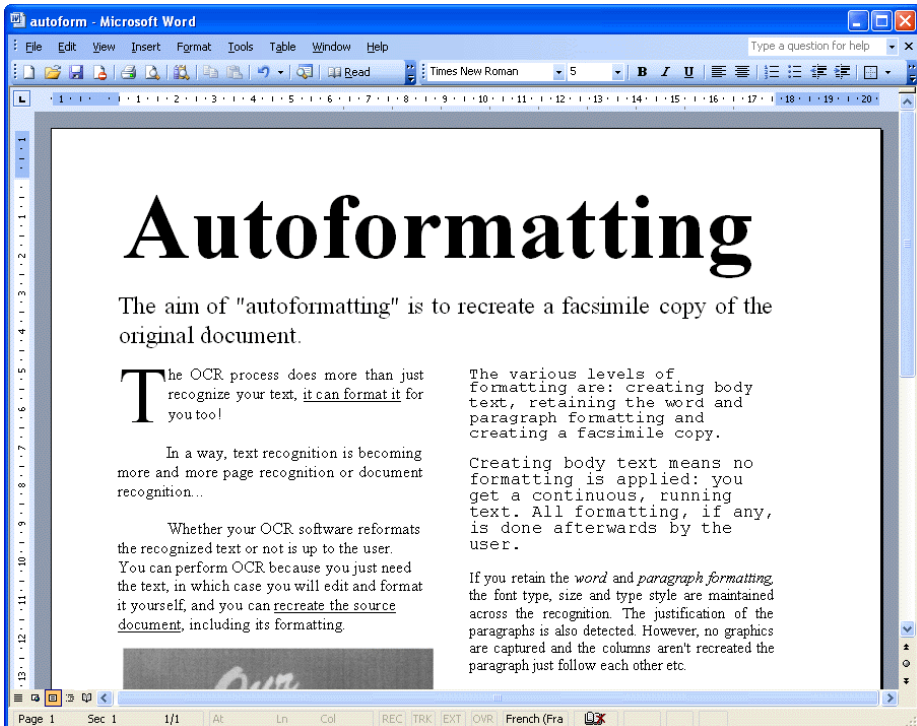


Klicken Sie auf die Schaltfläche „Format“ auf der Hauptfunktionsleiste und wählen Sie, das OCR-Ergebnis in Microsoft Word zu senden, oder selektieren Sie „RTF“ (Rich Text Format) oder das Word (DOC) Format. Wählen Sie weiter „Originaldokument Wiederherstellen“ als Layoutoption. (Die Option „Absätze Wiederherstellen“ wird automatisch aktiviert, um den Zeilenumbruch zu aktivieren.)

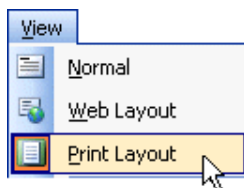


Ob die Layoutrekonstruktion verfügbar ist oder nicht, hängt vom gewählten Ausgabemodus ab. Einige „arme“ Formate, die „einfachen“ Text erzeugen, z.B. Text (ANSI), MS-DOS Text (ASCII) usw., unterstützen *keine* fortgeschrittenen Formatierungs-codes und ermöglichen somit keine automatische Formatierung. Das Adobe Acrobat PDF Format hingegen wurde geschaffen, um das Aussehen Ihrer Dokumente widerzuspiegeln: PDF Dokumente implizieren naturgemäß eine automatische Formatierung.

Wenn der erkannte Text mit einem Textverarbeiter geöffnet wird, sieht er ohne Eingreifen des Verwenders so aus.



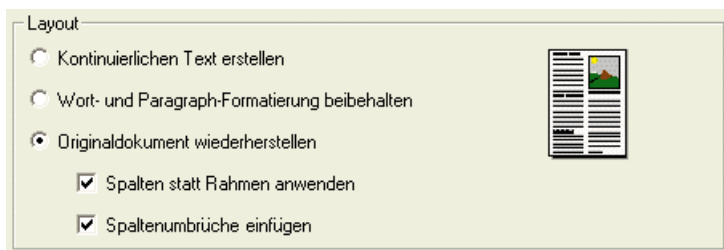
Um das Dokument in der richtigen Ansicht zu sehen, müssen Sie den „WYSIWIG“-Modus, meist „Seitenlayout“ genannt, aktivieren. Wenn Sie aber das erkannte Dokument direkt in Microsoft Word exportieren, wird die Seitenlayout- oder Seitenansicht-Ansicht automatisch aktiviert.



Also, Readiris kann Texte erkennen und formatieren. OCR ist nicht nur eine Texterkennung. OCR wird mehr und mehr eine **Seiten-** oder **Dokumentenerkennung!**

## SPALTEN BITTE, KEINE RAHMEN!

Die Formatierungsoption „Spalten statt Rahmen Verwenden“ bestimmt, *wie* die automatische Formatierung vorgenommen wird: Textblöcke, Tabellen und Grafiken können entweder in Rahmen oder aber in editierbaren **Spalten** untergebracht werden.



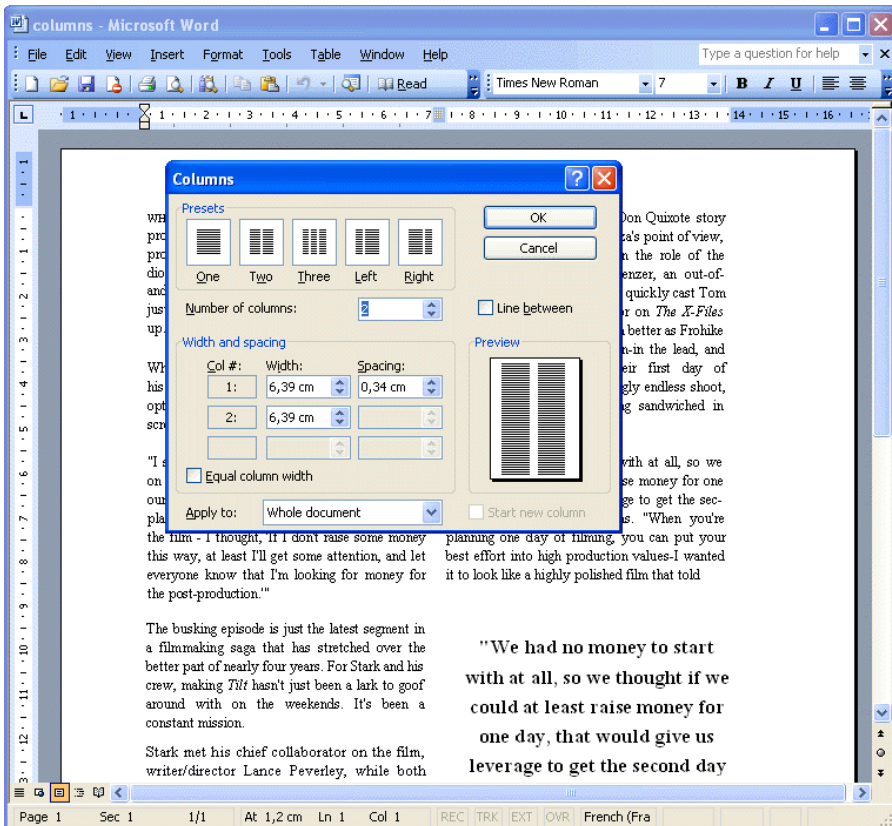
Rahmen („Frames“) sind getrennte „Containers“, die zur Positionierung mehrerer Textblöcke, Grafiken und Tabellen auf einer Seite verwendet werden. In Spalten fließt der Text natürlich von einer Spalte zur nächsten. In Spalten gegliederte Texte sind wesentlich einfacher zu bearbeiten.

Wir gehen nun davon aus, dass im gescannten Dokument tatsächlich Spalten vorhanden sind: Wenn das System nämlich keine Spalten im Quelldokument



feststellt, verwendet dieser Formatiermodus ohnehin Rahmen als „Rückzugslösung“.

Zum Ausprobieren können Sie das Bild COLUMNS.TIF im Readiris Ordner verwenden.



Die Option „Spaltenumbrüche Einfügen“ verfeinert die Wiederherstellung von Spalten: Sie legt fest, ob Sie „harte“ Spaltenumbrüche am Ende jeder Spalte einfügen möchten oder nicht. Mit Spaltenumbrüchen bleibt jeder Text, den Sie bearbeiten, hinzufügen oder entfernen, in der jeweiligen Spalte; kein Text fließt automatisch über einen Spaltenumbruch hinaus. Jeglicher Text hinter einem Spaltenumbruch wird an den oberen Rand der folgenden Spalte verschoben!

Aktivieren Sie diese Option, wenn Sie die Spaltenumbrüche an den Positionen erhalten möchten, an denen sie im erkannten Dokument gefunden wurden - unabhängig davon, welche Bearbeitungsschritte nach der OCR-Erkennung am Text vorgenommen werden. In Zeitungen und Zeitschriften entsprechen die verschiedenen Spalten einer Seite oftmals verschiedene Artikel. Es ist ggf. nicht ratsam, Text von einer Spalte automatisch in die nächste fließen zu lassen!

Deaktivieren Sie diese Option, wenn der Textkörper in Spalten aufgeteilt ist. So stellen Sie sicher, dass der Text ganz natürlich von einer Spalte in die nächste fließt.

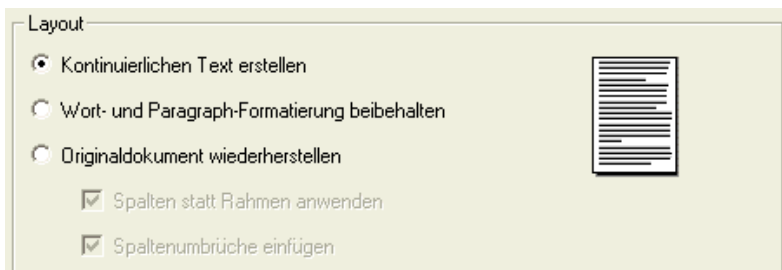
---

## TEXTFORMATIERUNG, TEIL 2

---

Es gibt auch zwei andere Layoutoptionen: „Kontinuierlichen Text Erstellen“ und „Wort- und Paragraph-Formatierung Beibehalten“.

Sowie das Symbol rechts zeigt, bedeutet „Kontinuierlichen Text Erstellen“ dass ein nicht formatierter, **kontinuierlicher Text** erstellt wird. Der Text wird ohne Formatierung gespeichert. Diese Option ist nur dann sinnvoll, wenn Sie nur den Text brauchen, aber nicht die Formatierung.



Textkörper erhalten Sie ebenfalls, wenn Sie eine Textzone schnell erkennen, indem Sie mit der rechten Maustaste auf die Textzone klicken und den Befehl „Kopieren Als Text“ wählen: Wenn die Erkennung abgeschlossen ist, können Sie den Textkörper in Ihre Textanwendung einfügen.

Die Option „Wort- und Paragraph-Formatierung Beibehalten“ ist ein Mittelding. Die **Wortformatierung** - Schriftart, Schriftgröße und Schriftauszeichnung - werden beibehalten. Das Gleiche gilt für die **Absatzformatierung** - der Randausgleich und die Einzüge werden ebenfalls beibehalten.

Verwechseln Sie diese Formatierungsoption nicht mit „kompletten“ Autoformatierung: hier werden die Absätze hintereinander gestellt, es werden keine Spalten rekreiert oder verschiedene Zonen in bestimmten Positionen kopiert.

## TEXT MEHRMALS EXPORTIEREN

---

Sie können die OCR-Ergebnisse mehrmals exportieren, ohne die Erkennung zu wiederholen! Ändern Sie das Textformat und die Formatierungsoptionen mit der Schaltfläche „Format“ und klicken Sie erneut auf „Erkennen“. Es wird nun keine OCR-Erkennung durchgeführt - außer wenn Sie neue Fenster festgelegt oder bestehende Fenster verändert haben! - sondern Readiris formatiert die OCR-Ergebnisse einfach um und speichert sie im neuen Textformat oder sendet sie an die soeben festgelegte Zielanwendung.



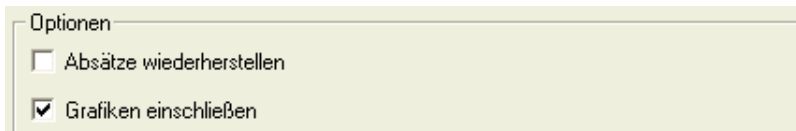


Dasselbe gilt für alle anderen Elemente, die Sie ändern: Wenn Sie eine Seite zu einem OCR-Job hinzufügen, wird nur diese Seite erkannt. Wenn Sie eine neue Textzone auf einer beliebigen Seite erstellen, wird nur diese Zone erkannt, bevor die Ergebnisse exportiert werden.

Sie können beispielsweise ein 10-seitiges Dokument erkennen und in einer Word-Datei abspeichern, die Zusammenfassung scannen und per E-Mail an einen ungedulden Kollegen senden, den Anhang - eine Tabelle - scannen und die gesamte Ergebnisse in einer HTML Datei speichern, die auf der Website Ihres Unternehmens veröffentlicht wird.

## GRAFIKEN GETRENNT SPEICHERN

In unserem Beispiel war die Grafik im erkannten Text enthalten. Ob dies der Fall ist oder nicht, hängt von der Option „Grafiken Einschließen“ ab. Ob es möglich ist, Grafiken innerhalb eines Textes zu speichern, hängt vom Ausgabemodus ab - „arme“ Textformate wie Text (ANSI) usw. können keine Grafiken enthalten!



Sie können mit Readiris jedoch auch Grafiken speichern, ohne die Texterkennung durchzuführen. Da Readiris Schwarzweiß-, Graustufen- und Farbbilder erzeugt, können Sie Schwarzweiß-Grafiken und Fotos erfassen.

Wie? Zeichnen Sie eine Grafikzone um die Illustrationen, Cartoons usw., die Sie brauchen, herum. Grafikfenster manuell zeichnen funktioniert genauso wie mit Text- oder Tabellenfenstern, wählen Sie jetzt einfach nur die Schaltfläche „Grafikfenster“.

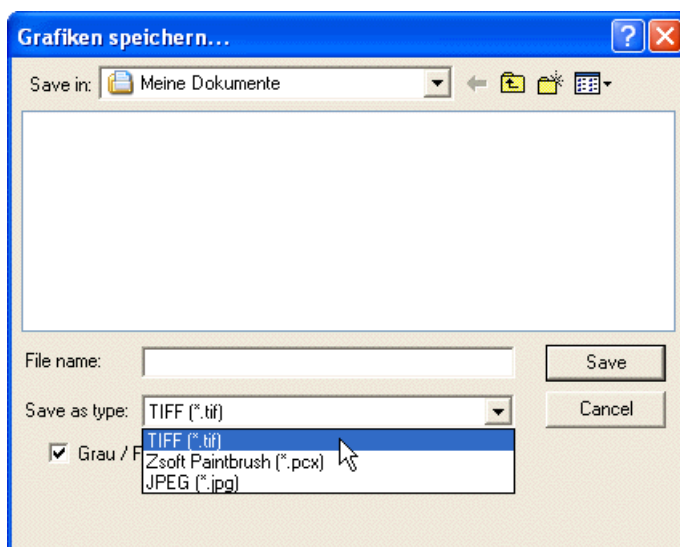


Selektieren Sie danach den Befehl „Grafiken Speichern“ im Menü „Datei“.

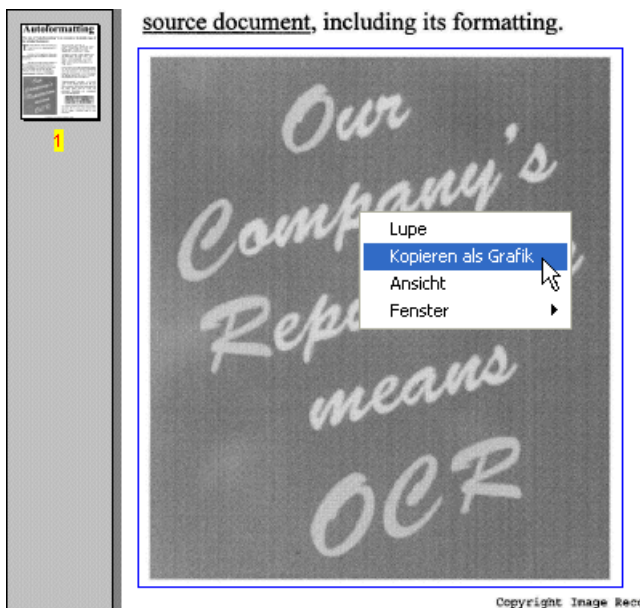
Grafiken speichern...



Legen Sie fest, welches Grafikdateiformat Sie benutzen möchten. Wählen Sie ein Format, das von Ihrer Fotoretuschiersoftware unterstützt wird. Verfügbar sind die Formate JPEG, TIFF und Paintbrush (PCX). Aktivieren Sie diese Option „Grauskala/Farben“, um die Grafiken als Farb- oder Graustufengrafik zu speichern.



Um eine Grafik in die Zwischenablage zu senden, statt sie als Bilddatei zu speichern, klicken Sie mit der rechten Maustaste auf das Grafikenfenster und wählen Sie den Befehl „Kopieren als Grafik“: Die Grafikzone unter dem Mauszeiger kann nun eingefügt werden!



## FAXE ERKENNEN UND ZURÜCKGESTELLTE ERKENNUNG

Die Speicherung von Grafiken als Bilddateien bietet Ihnen noch andere Möglichkeiten. Sie können die Bilder als *ganze* Seite speichern und die **zurückgestellte Erkennung** hinterher durchführen. Das gleiche haben wir mit den vorgescannten Bildern in diesem Handbuch gemacht!

Scannen Sie einfach das Dokument ein und wählen Sie den Befehl „Ganze Seite als Bild Speichern“ im „Datei“ Menü, um eine einzelne Seite zu speichern. Sie werden wieder aufgefordert, die gesamte Seite als TIFF oder Paintbrush Datei (PCX) zu speichern.



Wählen Sie den Befehl „Alle Seiten als Bild Speichern“, um ein mehrseitiges Dokument zu speichern. Nur ein Dateiformat ist verfügbar: mehrseitige TIFF.

Jetzt können Sie die Festplatte als Quelle selektieren und die Bilddatei mit der Schaltfläche „Öffnen“ (oder mit dem entsprechenden Befehl im Menü „Verarbeitung“) öffnen. (Wenn Sie den Befehl „Öffnen“ im Menü „Datei“ benutzen, brauchen Sie die Quelle nicht auf dem neuesten Stand zu bringen.)

Da Readiris Farb-, Graustufen- und Schwarzweißbilder gleichermaßen unterstützt, öffnet Readiris Adobe Acrobat PDF Dokumente, JPEG Bilder, Paintbrush (PCX) Bilder, DCX Faxbilder (eine mehrseitige Version des Paintbrush Formats), PNG Bilder, TIFF Bilder (nicht komprimiert, LZW, Packbits, Gruppe 3 und Gruppe 4 komprimiert), mehrseitige TIFF Bilder und Windows Bitmaps (BMP).

Diese Funktion ist sehr praktisch, wenn Sie Ihre **Faxe** zur weiteren Bearbeitung in Textdateien konvertieren möchten! Readiris verwendet besonders intelligente Routinen beim Lesen von Faxnachrichten: Die Software erkennt die typischen Fauxauflösungen - 100 x 200 dpi („normal“), 200 x 200 dpi („fein“) und 200 x 400 dpi („superfein“) - und führt automatisch eine Vorverarbeitung dieser Bilder durch, um optimale OCR-Ergebnisse zu erzielen.

Dennoch ist es eine gute Idee, Ihre Partner zu bitten, die Faxe in „feiner“ Qualität zu senden - diese Faxe liefern bessere OCR-Ergebnisse.

Vergessen Sie nicht, dass Sie mit der rechten Maustaste auf Bilder im Windows Explorer klicken und den Befehl „Erkennen“ im „Kontext“ Menü wählen können, um Bilder zu öffnen! Sie können ebenfalls Dateien per „Drag & Drop“ öffnen: Ziehen Sie Bilddateien vom Windows Explorer in den Bildbereich von Readiris oder das Readiris Anwendungssymbol, und schon werden die Dateien geöffnet.

## **TABELLENERKENNUNG**

---

Bis jetzt haben wir uns mit der Erkennung von Texten und Faxen, sowie mit der Speicherung von Grafiken befaßt. Jetzt werden wir eine Tabelle bearbeiten. Nehmen Sie eine Zeichentabelle und scannen sie oder öffnen Sie die Vorlage TABLES.JPG im Readiris Ordner.

Tatsächlich enthält das Bild TABLES.JPG zwei Tabellen, und das ist kein Zufall! Die Seitenanalyse gibt sie als Tabellenfenster an, und Readiris wird sie für Sie rekonstruieren, indem es die Tabellen Zelle für Zelle in Ihr Spreadsheet einfügt oder indem es das Tabellen-Objekt in Ihre Textverarbeitungsdateien einfügt.

Gehen wir die verschiedenen Lösungen durch, beginnend mit den „umrahmten“ Tabellen - sie haben Ränder um die einzelnen Zellen herum.



**Reading Tables**

Readiris recognizes tabular data and recreates them cell by cell in worksheets or as table objects inside wordprocessor files.

To insert tables as table objects, you must retain the word and paragraph formatting or recreate the source document; see the "Format" button on the main toolbar.

The page analysis detects "**gridded**" and "**ungridded**" tables. "Gridded" or "framed" tables have borders around the cells - as does the example below. The borders of the table cells get recreated.

Performance test optical media				
CD-ROM Digital Versatile Disk	Average access time (msec)	CPU utilization (%)	Video clip playbacks (frames dropped)	Sequential read 16 KB (K bps)
CD-ROM 4x speed	442	4.2	10 8	612
CD-ROM 12x speed	137	20.9	5 4	1,586
CD-ROM 24x speed	80	58.2	3 2	2,258
CD-ROM 32x speed	60	72.1	- -	2,987
DVD	58	78.9	- -	3,143

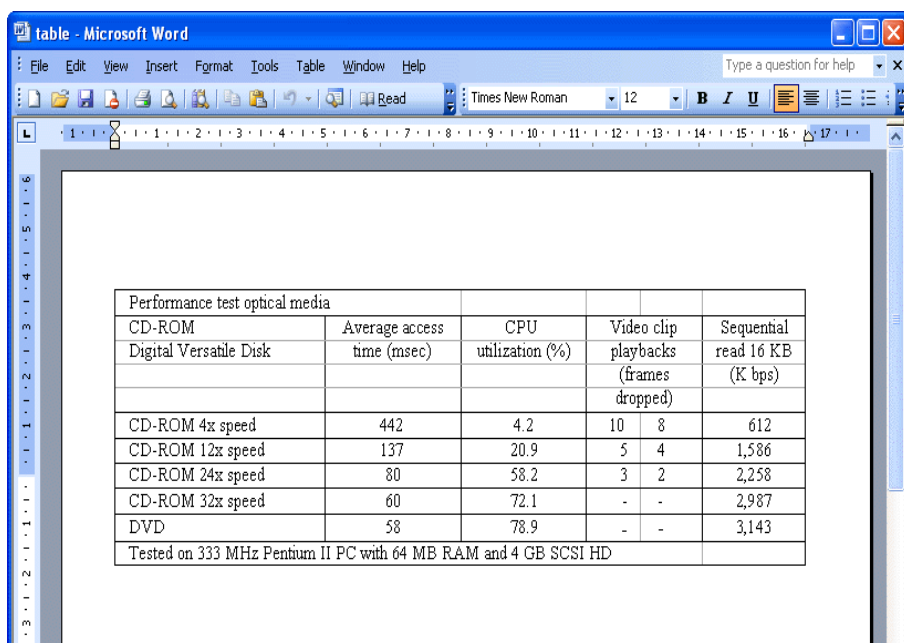
Tested on 333 MHz Pentium II PC with 64 MB RAM and 4 GB SCSI HD

"**Ungridded**" tables don't have any borders around the cells. When the columns of ungridded tables are too widely spaced, the page analysis may not detect a table window to avoid confusion with columnized text blocks.

When your tables exclusively contain **numeric characters**, enable the numeric reading mode with the "Language" button on the main toolbar for increased accuracy.

Finally, you can send your tables of figures directly to Microsoft Excel by selecting the spreadsheet as target application - refer to the "Format" button on the main toolbar

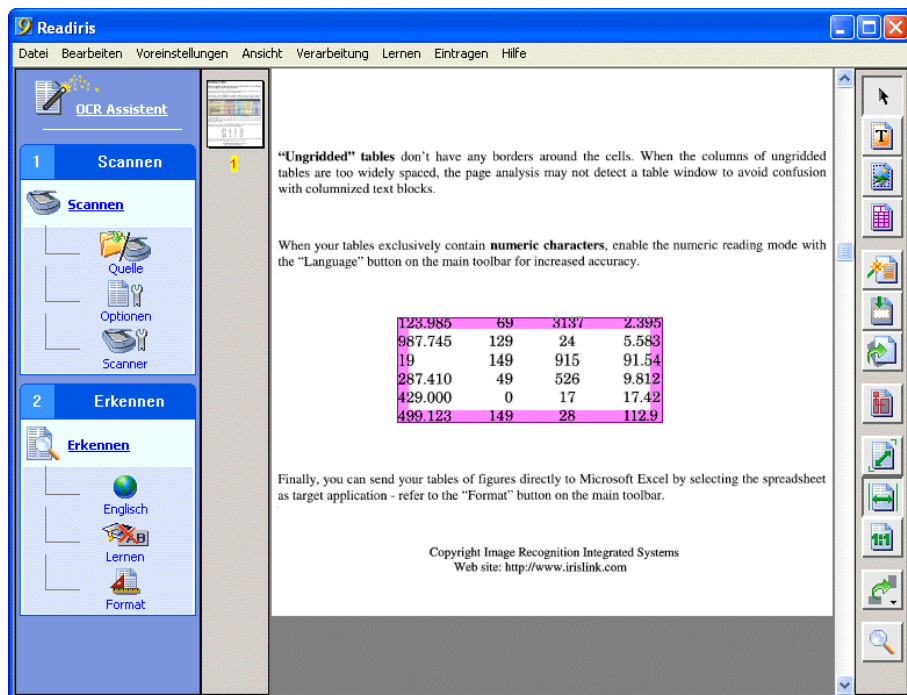
Führen Sie die Erkennung durch, nachdem Sie die Layoutoption „Wort- und Paragraph-Formatierung Beibehalten“ oder „Originaldokument Wiederherstellen“ aktiviert haben, und die Tabelle wird wiederhergestellt. Öffnen Sie Ihre Textverarbeitung, um das Ergebnis zu sehen. Die Zellen und Ränder wurden einzeln von Readiris wieder hergestellt! (Sie konnten natürlich auch die Textparagrafen in Ihre Textdatei miteinschliessen.)



The screenshot shows a Microsoft Word window titled "table - Microsoft Word". The menu bar includes File, Edit, View, Insert, Format, Tools, Table, Window, and Help. The toolbar shows various icons for file operations and formatting. The status bar at the bottom indicates "Times New Roman", "12", and "B I U". The table is centered on the page and contains performance test data for optical media.

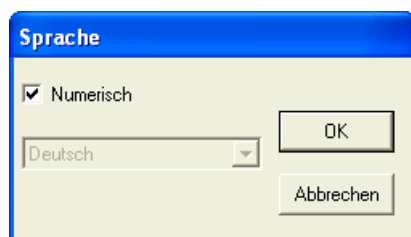
Performance test optical media					
CD-ROM	Average access	CPU	Video clip		Sequential
Digital Versatile Disk	time (msec)	utilization (%)	playbacks		read 16 KB
			(frames		(K bps)
			dropped)		
CD-ROM 4x speed	442	4.2	10	8	612
CD-ROM 12x speed	137	20.9	5	4	1,586
CD-ROM 24x speed	80	58.2	3	2	2,258
CD-ROM 32x speed	60	72.1	-	-	2,987
DVD	58	78.9	-	-	3,143
Tested on 333 MHz Pentium II PC with 64 MB RAM and 4 GB SCSI HD					

Und jetzt die „nicht umrahmte“ Tabellen - Sie haben keine Ränder um die Zellen herum. Beachten Sie, dass die Seitenanalyse die Tabelle trotzdem erkennt!



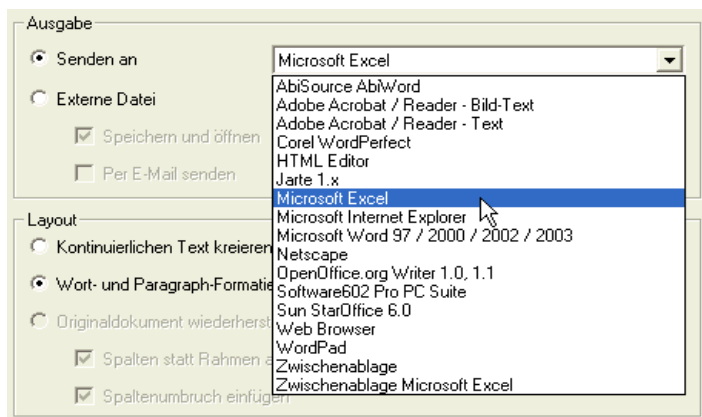
Um eine optimale Scanqualität zu bekommen, sollten Sie die Erkennung auf **numerischen Symbole** mit der Schaltfläche „Sprache“ begrenzen. (Der numerische Modus ist nicht ausschließlich numerisch, er enthält die Zeichen 0 bis 9, +, \*, /, %, , (Komma), . (Punkt) (,), -, =, \$, £, ¥ und das € Symbol.)





Da Sie das nur tun können, wenn die Tabelle keine alphabetische Symbole enthält - sonst werden die Textteile nicht richtig erkannt -, kann der numerische Modus jetzt aktiviert werden. Dieser konnte aber nicht für die erste Tabelle aktiviert werden.

Diesmal werden wir das OCR-Ergebnis direkt in das Spreadsheet Microsoft Excel senden. Wir wählen Excel als gewünschte Anwendung mit der Schaltfläche „Format“.

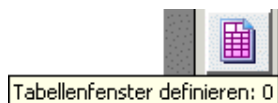


Das Spreadsheet startet automatisch und die Ergebnisse sehen so aus: Die typische Tabellenstruktur mit Reihen und Spalten wird wieder hergestellt. Sie können die Daten direkt weiterverarbeiten.



	A	B	C	D	E
1	123.985	69	3137	2.395	
2	987.745	129	24	5.583	
3	19	149	915	91.54	
4	287.410	49	526	9.812	
5	429.000	0	17	17.42	
6	499.123	149	28	112.9	
7					

Sie könnten auf „nicht umrahmte“ Tabellen stoßen, die die Seitenanalyse nicht als Tabellen erkennt, weil die Spalten zu weit auseinander stehen. Readiris versucht Verwechslungen mit Text in Spalten zu vermeiden. Um ein Tabellenfenster manuell zu kreieren, klicken Sie auf die Schaltfläche „Tabellenfenster“ in der Bild-Toolbar und gehen Sie wie üblich vor; das QuickInfo des Befehls gibt wie üblich die Anzahl Tabellenfenster an.



## ONLINE-HILFE BEKOMMEN

Hiermit beschließen wir unsere allgemeine Übersicht von Readiris. Es ist möglich, dass die neuesten Informationen noch nicht in diesem Handbuch besprochen wurden. Wir empfehlen Ihnen deshalb, das Online-Hilfe-System zu konsultieren, wenn Sie zusätzliche Informationen über Readiris brauchen.

Selektieren Sie das Menü „Hilfe“. Der Befehl „Hilfethemen“ oder die Taste F1 gestattet Ihnen den Zugriff auf viele Online-Hilfe-Informationen.



Die anderen Befehle des Menüs „Hilfe“ erklären, wie Sie Produktunterstützung bekommen und wie Sie I.R.I.S. kontaktieren können, bietet einen direkten Zugriff zur I.R.I.S. Homepage usw.

