

# Readiris™ PRO 10

## BENUTZER- HANDBUCH



Readiris Corporate

© 2005 I.R.I.S. Alle Rechte vorbehalten

OCR Technologie von I.R.I.S.

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

ICR und Strichcodelesetechnologie von I.R.I.S.

BCR und Feldanalyse-Technologie von I.R.I.S.

© 2005 I.R.I.S. Alle Rechte vorbehalten

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

---

Herzlichen Glückwunsch für den Erwerb von Readiris! Dieses Softwarepaket wird bei der Erfassung von Texten, Tabellen, Grafiken, Visitenkarten, Strichcodes und sogar Handschriften zweifellos von großem Nutzen für Sie 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 unterstützt bis zu 117 Sprachen: es werden alle amerikanischen und europäischen Sprachen unterstützt, die Sprachen in Mitteleuropa, die baltischen Sprachen, Griechisch und die kyrillischen („russischen“) Sprachen. (Optional können Sie Dokumente in Hebräisch und vier asiatische Sprachen lesen - Japanisch, vereinfachtes und traditionelles Chinesisch und Koreanisch.) Auch vermischte Alphabete sind für Readiris kein Problem: die Software erkennt „westliche“ Wörter, die in griechischen, kyrillischen, hebräischen und asiatischen Dokumenten auftauchen - meist nicht transkribierbare Eigennamen, Markennamen usw. werden in westlichen Zeichen 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 realen 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 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. Readiris kann sogar das Hintergrundfoto einer gescannten Seite in das Ergebnisdokument einfügen! Und da Readiris Graustufen- und Farbscans unterstützt, können Sie beliebige Grafiken einlesen, ob Lineart, Schwarzweißfotos 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.

Auch Strichcodes, die auf einer gescannten Seite auftauchen, werden gelesen, und das funktioniert auch für handgeschriebenen Text - handgeschriebene Texte können erkannt werden, wenn diese in deutlich voneinander getrennten „Blockbuchstaben“ geschrieben sind.

Sie können sogar Visitenkarten mit Readiris erkennen: scannen Sie Ihre Visitenkarten ein, führen Sie an ihnen eine Texterkennung durch und konvertieren Sie sie in eine Adressen-Datenbank. Denken Sie doch nur an die letzte Ausstellung zurück, als Sie mit einem Haufen Visitenkarten zurückgekommen sind und Ihre Sekretärin zwei Tage gebraucht hat, all diese Karten einzugeben!

Die Kartendaten werden automatisch aus dem Bild ausgelesen, und die erkannten Daten werden spezifischen Datenbank-Feldern zugeordnet. Readiris macht umfassenden Gebrauch von einer Wissens-Datenbank, wodurch es die erforderliche Intelligenz erwirbt, um den Vor- und Nachnamen, eine Stadt und ihr Bundesland, eine Telefon- und eine Telefaxnummer usw. zu unterscheiden. Die daraus resultierenden Daten können sofort an ein Adressenverwaltungsprogramm wie Microsoft Outlook (Express) oder eine beliebige vCard-kompatible Anwendung gesendet werden.

Readiris unterstützt eine Vielzahl beliebter Scanner: zahlreiche Flachbettscanner, Sheetfedscanner und Kombigeräte („MFPs“ bzw. „Multifunctional Peripherals“) sowie Digitalkameras können genutzt werden. Readiris unterstützt ebenfalls den



Twain Scanning-Standard und einige Scan-Plattformen. Durch das Scannen mit Zeitverzögerung können auch dann mehrseitige Dokumente gescannt werden, wenn der Scanner nicht mit einem Blatteinzug ausgestattet ist. (Readiris Corporate unterstützt Hochgeschwindigkeits-Scanner und führt eine Stapelerkennung an umfangreichen Bildsammlungen durch: gescannte Stapel können mit Hilfe von leeren Seiten in separate Dokumente getrennt werden, ein automatisches Lesen von Strichcodes stellt eine ordnungsgemäße Indexierung von Dokumenten sicher, die nun in elektronischer Form vorliegen.)

## **INHALTSÜBERSICHT**

---

Zeit sparen, Neu schreiben vermeiden! .....	III
Inhaltsübersicht .....	VI
Warenzeichen und Urheberrechte .....	VIII

### **Kapitel 1: Installation**

Systemvoraussetzungen .....	1-1
Installation von Readiris .....	1-1
Entfernen der Readiris Software .....	1-4
Readiris Entfernungsprogramm .....	1-4
Windows „Installieren/Entfernen“-Assistent .....	1-5
Installieren von Software-Optionen .....	1-6
Installierte Dateien .....	1-8
„Lese Mich“ Datei und Dokumentation .....	1-8
Handschriftformular .....	1-8
Scanner Treiber .....	1-8
Vergessen Sie nicht registrieren zu lassen! .....	1-9
Produktunterstützung bekommen .....	1-11
Mit I.R.I.S. in Verbindung treten .....	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 .....	2-11

Eins, Zerlegung eines gescannten Bildes .....	2-14
Anderthalb, Fenster sortieren .....	2-17
Zwei, Manuelle Fensterbestimmung eines Bildes .....	2-20
Drei, Fenstervorlagen speichern .....	2-24
Mit Readiris um die Welt! .....	2-26
Benutzerlexika steigern die Linguistik .....	2-31
Bei Bedarf ändert Readiris die Sprachen .....	2-34
Gemischtsprachige Dokumente einlesen .....	2-36
Bestimmen der Dokumenteigenschaften .....	2-38
Readiris wird immer intelligenter! .....	2-40
Lernen .....	2-42
Nicht Lernen .....	2-42
Löschen .....	2-43
Rückgängig .....	2-43
Ende .....	2-44
Abbrechen .....	2-44
Die Rolle der Zeichenwörterbücher .....	2-44
Das Ergebnis direkt in Ihre Anwendung senden .....	2-47
Das Ergebnis in einer Textdatei speichern .....	2-51
Portable Dokumente erstellen... ..	2-55
... oder lesen .....	2-62
Mehrere Seiten erkennen .....	2-65
Mehrseitige Dokumente bearbeiten .....	2-75
Ein neues Dokument starten .....	2-78
Textzonen erkennen .....	2-79
Die Aussicht der Textausgabe .....	2-81
Scannereinstellung .....	2-82
Bringen Sie Farbe in Ihre Text-Scans! .....	2-84
Andere Geräte, andere Auflösung .....	2-87
Standard Einstellungen speichern .....	2-91
Vordefinierte Einstellungen speichern .....	2-92
Dokumente scannen .....	2-93
Justieren der gescannten Bilder .....	2-97
Wenn der OCR Assistent für Sie arbeitet... ..	2-102
Readiris erstellt das Layout .....	2-103
Spalten bitte, keine Rahmen! .....	2-108
Textformatierung, Teil 2 .....	2-112



Text mehrmals exportieren .....	2-113
Grafiken getrennt speichern .....	2-113
Speichern farbiger Hintergründe .....	2-116
Grafiken durch und durch .....	2-119
Faxe erkennen und zurückgestellte Erkennung .....	2-121
Dokumentstapel erkennen .....	2-122
Einen „überwachten“ Ordner festlegen .....	2-124
Organisieren von Stapeln .....	2-126
Tabellenerkennung .....	2-130
Handschriftlichen Text erkennen .....	2-135
Lesen von Strichen und Lücken .....	2-139
Visitenkarten erkennen .....	2-141
Visitenkarten scannen .....	2-142
Den Lesemodus für Visitenkarten aktivieren! .....	2-147
Visitenkarten erkennen .....	2-151
Online-Hilfe bekommen .....	2-152

## **WARENZEICHEN UND URHEBERRECHTE**

---

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

AutoFormat, Connectionist, der IBCR-II, 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. 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.
- 64 MB RAM. 128 MB RAM ist zur Verarbeitung von Graustufen- und Farbbildern empfohlen.
- 120 MB freier Festplattenspeicherplatz. 105 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.

Readiris Corporate benötigt eine Bildschirmauflösung von 1.024 x 768.

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

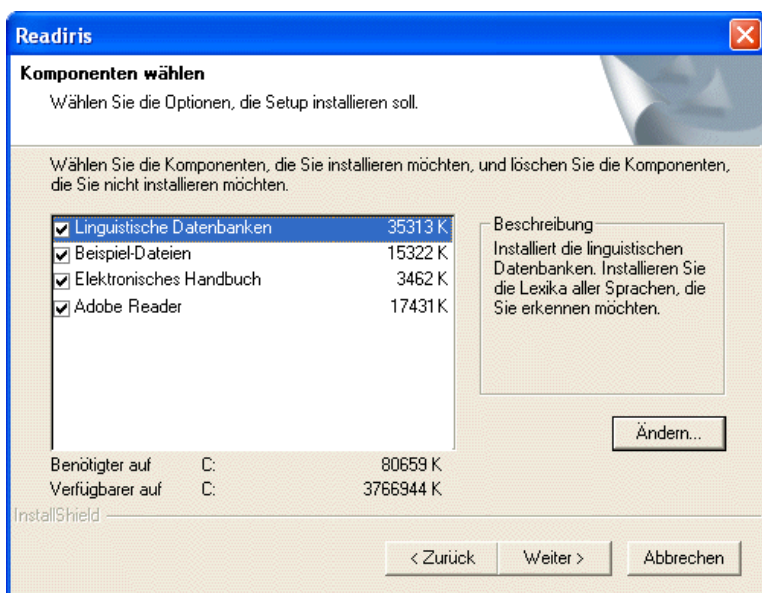


CD-ROM-Laufwerk ein und warten Sie, bis sie aktiviert wird. Folgen Sie den Anweisungen auf dem Bildschirm.

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.

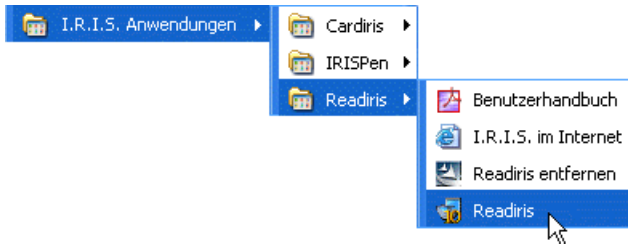
Benutzer von Windows XP, Windows 2000 und Windows NT müssen sicherstellen, dass Sie die erforderlichen **Zugriffsrechte** haben. Wenden Sie sich bei Bedarf an Ihren Systemadministrator.

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 Softwareokumentation 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.



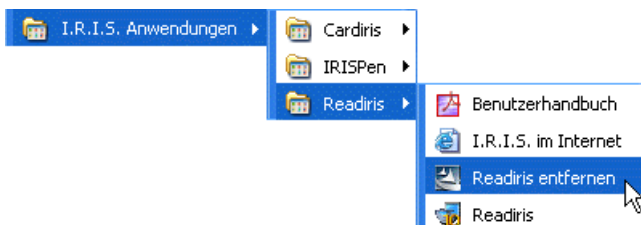
Außerdem legt das Installationsprogramm eine Readiris **Verknüpfung** auf dem Windows Schreibtisch an, mit dem Sie Readiris direkt vom Schreibtisch aus starten können.

## 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.

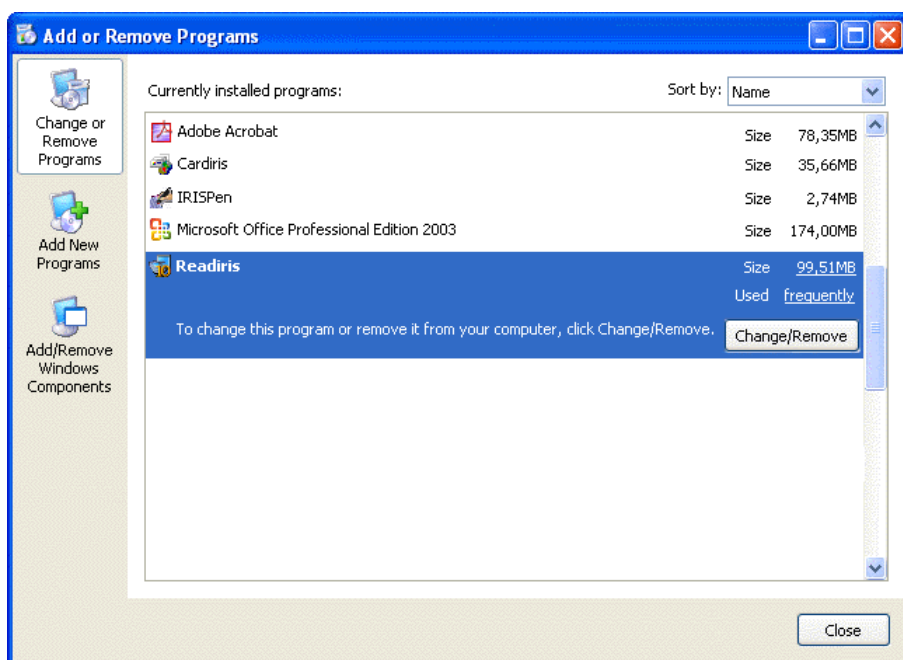
### Readiris Entfernungsprogramm

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.



## INSTALLIEREN VON SOFTWARE-OPTIONEN

Für die Readiris Software stehen zwei Software-Optionen zur Verfügung: die Software-Optionen „Asiatisches OCR-Modul“ und „Hebräisches OCR-Modul“. Mit der Option „Asiatisches OCR-Modul“ können Sie Texte in Japanisch, Traditionelles Chinesisch, Vereinfachtes Chinesisch und Koreanisch lesen.

The screenshot shows the Readiris OCR Assistant interface. The main window displays a document with Chinese text and a portrait of a man. The interface includes a menu bar, a toolbar, and a status bar.

**OCR Assistent**

1 Scannen

Scannen

Quelle

Dokumenttyp

Optionen

Scanner

2 Erkennen

Erkennen Speichern

Chinesisch (Ver)

Lernen

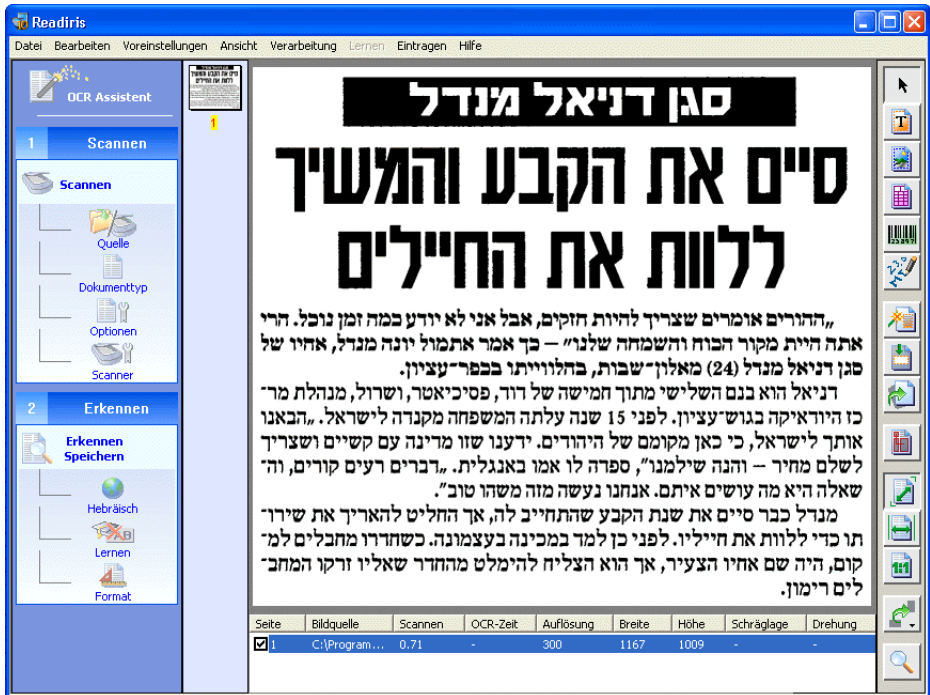
Format

再没有不可跨越的障碍  
再没有不可完成的挑战  
因为全体无间的配合  
荣誉得以壮阔成就  
造车亦是如此  
东南汽车

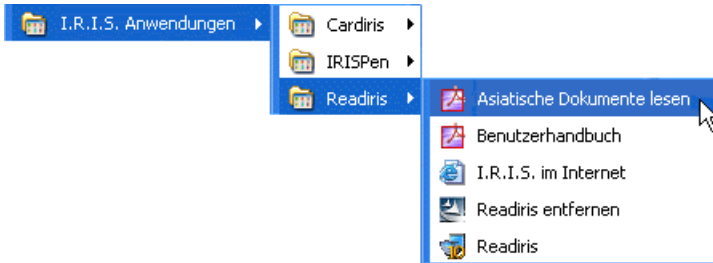
为您邀约台湾多家协力厂  
秉持 制造您身边的车的理念  
不只是国际化的专业和严格  
更科学规划，确保彼此配合严密和谐  
使得驾乘拥有更为人性的舒适与方便

Seite	Bildquelle	Scannen	OCR-Zeit	Auflösung	Breite	Höhe	Schräglage	Drehung
<input checked="" type="checkbox"/> 1	C:\Progra...	0.54	-	300	727	939	-	-

Das „Hebräische OCR-Modul“ ermöglicht, wie der Name schon sagt, das Erkennen von Dokumenten in Hebräisch.



Diese Software-Optionen werden wie gewohnt auf einer selbststartenden CD-ROM geliefert. Nach dem Installieren der Option „Asiatisches OCR-Modul“ stehen Ihnen spezielle Dokumente zur Verfügung, in denen erläutert wird, wie Dokumente in asiatischen Sprachen erkannt werden können.



## **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)

### **Handschriftformular**

TEMPLATE.PDF leeres Handschriftformular zum Ausdrucken

TEMPLATE.DOC leeres Handschriftformular zum Bearbeiten

### **Scanner Treiber**

Zögern Sie nicht, Ihren Scanner-Fabrikanten oder seinen Vertreter 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 Registrierungsmöglichkeiten: 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!



**Registrieren Sie Ihre Readiris-Lizenz**

**Warum registrieren?**

- Die Registrierung erlaubt es uns, Sie über die künftige **Produktentwicklungen** und über **weitere I.R.I.S.-Produkte** zu informieren.
- Durch die Registrierung haben Sie das Recht, kostenlose **Produktunterstützung** zu bekommen und über **Sonderangebote** informiert zu werden.
- Sie bekommen einen von Ihrer Software-Version abhängigen "**Softkey**". Diesen Softkey benötigen Sie, um Readiris nach Ablauf eines Monats weiter benutzen zu können.

**Wie kann ich...?**

**Per Post**

Schicken Sie uns Ihre **Registrierkarte** zu.

**Fax**

Bei Verlust Ihrer Registrierkarte oder wenn die Softwareokumentation lediglich in **elektronischer Form** geliefert wurde, senden oder faxen Sie den Ausdruck der [Registrierkarte](#) an I.R.I.S.

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.



**Readiris Hilfe**

Hide Back Forward Home Print Options

Contents | Index | Search

- Willkommen bei der Readiris Hilfe
- Vorstellung OCR
- Dokumente erkennen
- Handschriftlichen Text erkennen
- Strichcodes erkennen
- Wie kann ich...?
- Befehlsübersicht
- Software-Versionen und Optionen
- Produktregistrierung
- Produktunterstützung
  - Produktunterstützung bekommen**
  - I.R.I.S. erreichen
  - Produktunterstützung bekommen
- I.R.I.S.

## So erhalten Sie Produktunterstützung

Jeder registrierte Kunde bekommt **kostenlose** technische **Produktunterstützung**. ((Nach der [Registrierung](#) erhalten Sie Informationen über Sonderangebote.)

**Europa**  
Hotline: 32-10-45 13 64 (Bürozeiten) (alle EU-Hauptsprachen)  
Fax: 32-10-45 34 43

**USA**  
Hotline: 1-561-921-0847 / 800-447-4744 (Bürozeiten)  
Fax: 1-561-921-0854

**WWW**  
[www.irislink.com/support.html](http://www.irislink.com/support.html) (Hilfe zur Fehlerbehebung)  
Klicken Sie [hier](#) um Zugang zu Informationen zur Fehlerbehebung zu erhalten.

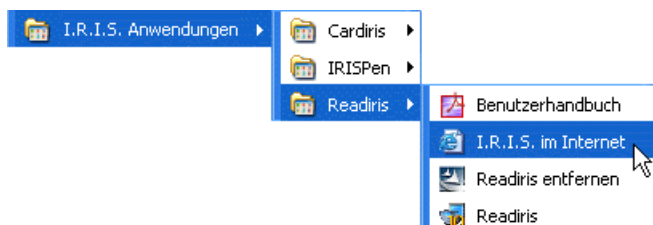
**E-Mail**  
[support@irislink.com](mailto:support@irislink.com), [support@irisusa.com](mailto:support@irisusa.com)

## MIT I.R.I.S. IN VERBINDUNG TRETEN

Sie können sich ebenfalls an I.R.I.S. wenden, um weitere Informationen über andere Lösungen aus unserer Produktpalette zu erhalten, 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 Startbildschirm und das Kommando „I.R.I.S. im Internet“ im „Hilfe“ Menü von Readiris.



# 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.

**Readiris**

Die Identifikationsnummer auf diesem PC ist:

315212345D59427F894PI222451AC85127I

Zur Aktivierung dieser Anwendung brauchen Sie einen Schlüssel.

Setzen Sie sich in Verbindung mit I.R.I.S. um diesen Schlüssel zu bekommen

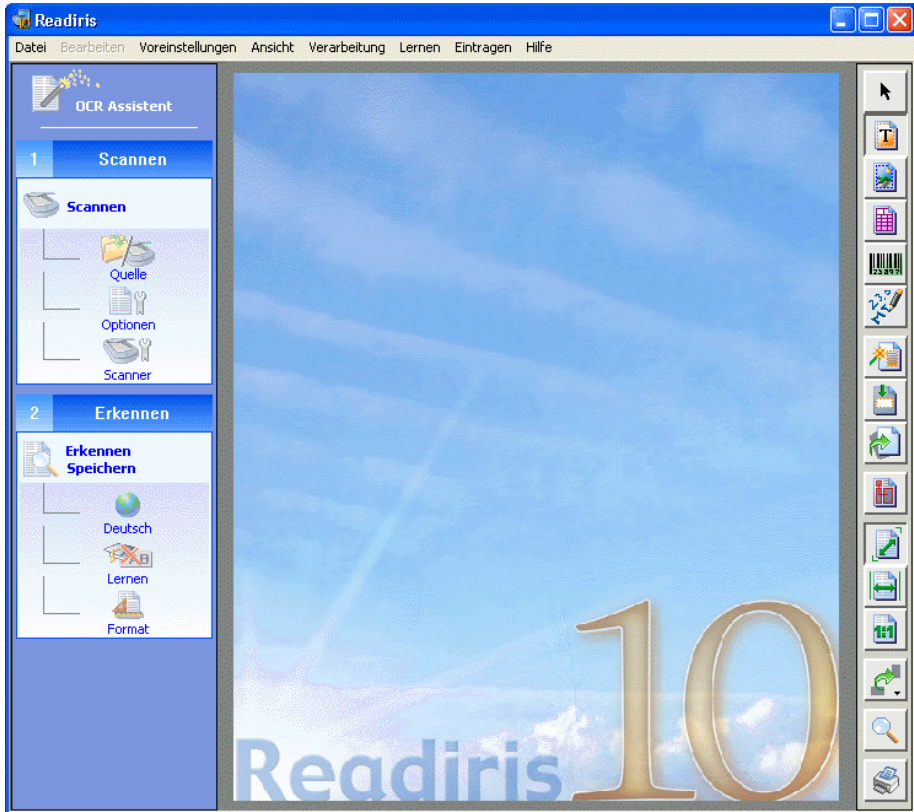
Geben Sie den Schlüssel ein:



## **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: der **Tooltip** 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).



Zweitens wird anschließend das **Dokumentpanel** unter dem gescannten Bild angezeigt. In diesem Panel werden **statistische Daten** und Informationen über alle gescannten Seiten angezeigt - Quelle und Auflösung des Bildes, für Scannen und Erkennen benötigte Zeit usw. (Auch im Dokumentpanel stehen Tooltips zur Verfügung...)

Seite	Bildquelle	Scannen	OCR-Zeit	Auflösung	Breite	Höhe	Schräglage	Drehung
<input checked="" type="checkbox"/> 1	C:\Program Files\Readiris\multipage.tif	1.09	-	300	2000	2388	-	-
<input checked="" type="checkbox"/> 2	C:\Program Files\Readiris\multipage.tif	1.03	-	300	2000	1888	-	-
<input checked="" type="checkbox"/> 3	C:\Program Files\Readiris\multipage.tif	0.98	-	300	1912	2004	-	-

## 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

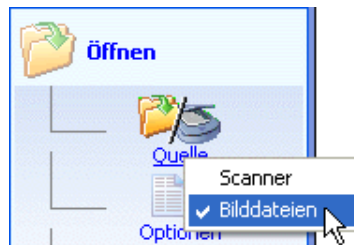
Ihnen den Start auch ohne Scanner ermöglichen. Diese Bilder schauen wir uns an.

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

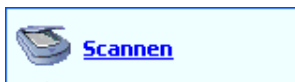
Farb-, Graustufen- und Schwarzweißbilder werden gleichermaßen unterstützt. In Readiris können Sie Adobe Acrobat PDF-Dokumente, DCX-Faxnachrichten (eine Mehrseitenversion des Paintbrush-Formats), DjVu-Bilder (\*.djv, \*.djvu), JPEG-Bilder, JPEG 2000-Bilder (\*.j2c, \*.jp2), PNG-Bilder, TIFF-Bilder (unkomprimiert, LZW, PackBits, Group 3, Group 4 und JPEG-komprimiert), mehrseitige TIFF-Bilder, Windows Bitmaps (\*.bmp) und ZSoft Paintbrush-Bilder (\*.pcx) 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.



Öffnen...

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.



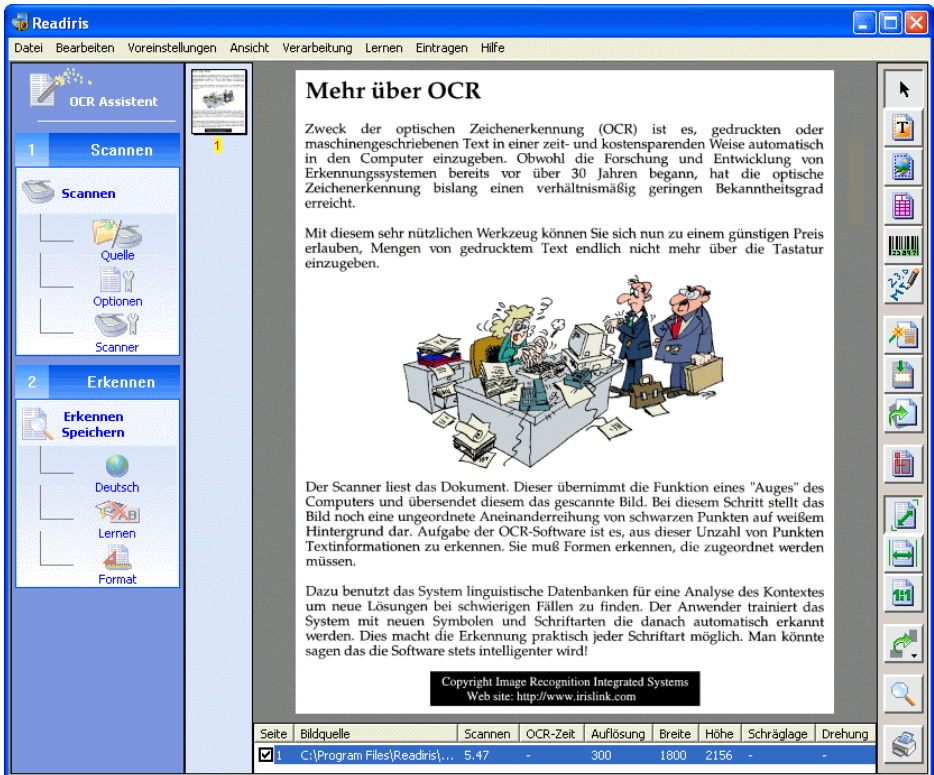
Ladet D:\Readiris\german.jpg



Konvertierung

64%

Dieses Bild wird schließlich im Bildbereich angezeigt. In der Seiten-Toolbar und im Dokumentpaneel wird angezeigt, dass eine einzelne Seite in Readiris geladen wird.

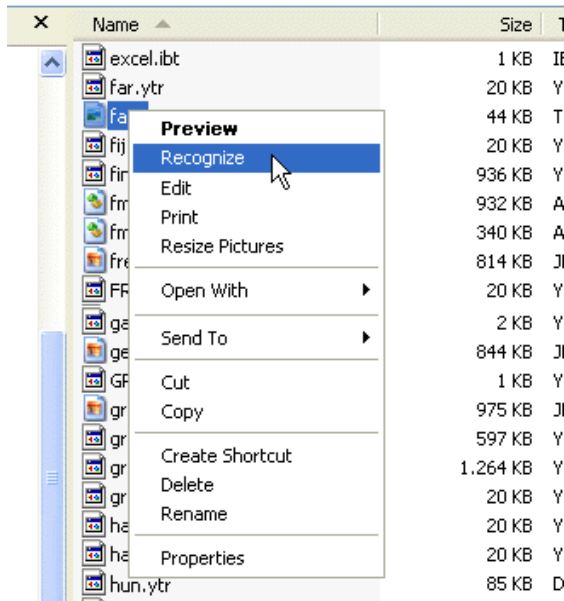


Eine dritte Möglichkeit ist, vorgescannte Bilder mit Hilfe der „**Drag & Drop**“-Funktion zu öffnen. Ziehen Sie Bilder vom Windows Explorer in den Readiris Bildbereich oder auf das Readiris Anwendungssymbol, und schon werden die Bilder geöffnet.



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**

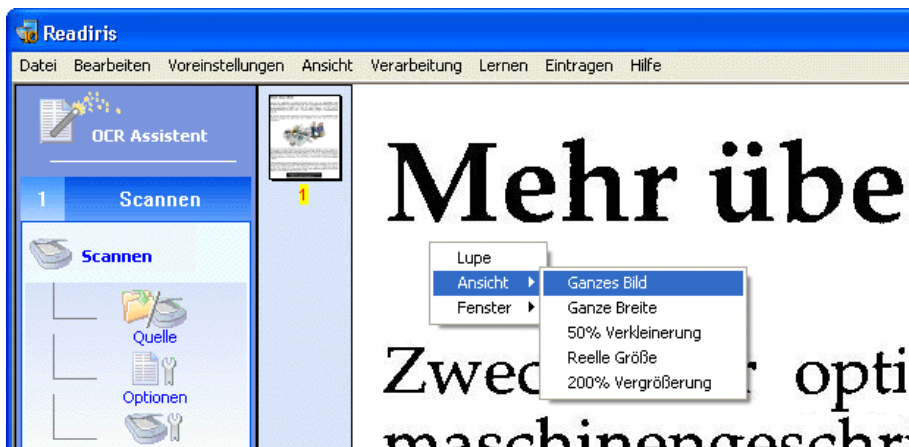
Readiris verfügt über mehrere Befehle um das gescannte Bild zu **zoomen**, zum Beispiel um die Scanqualität zu überprüfen.



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

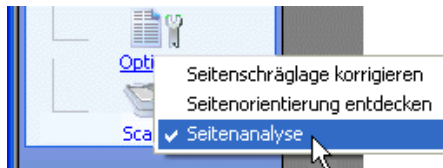




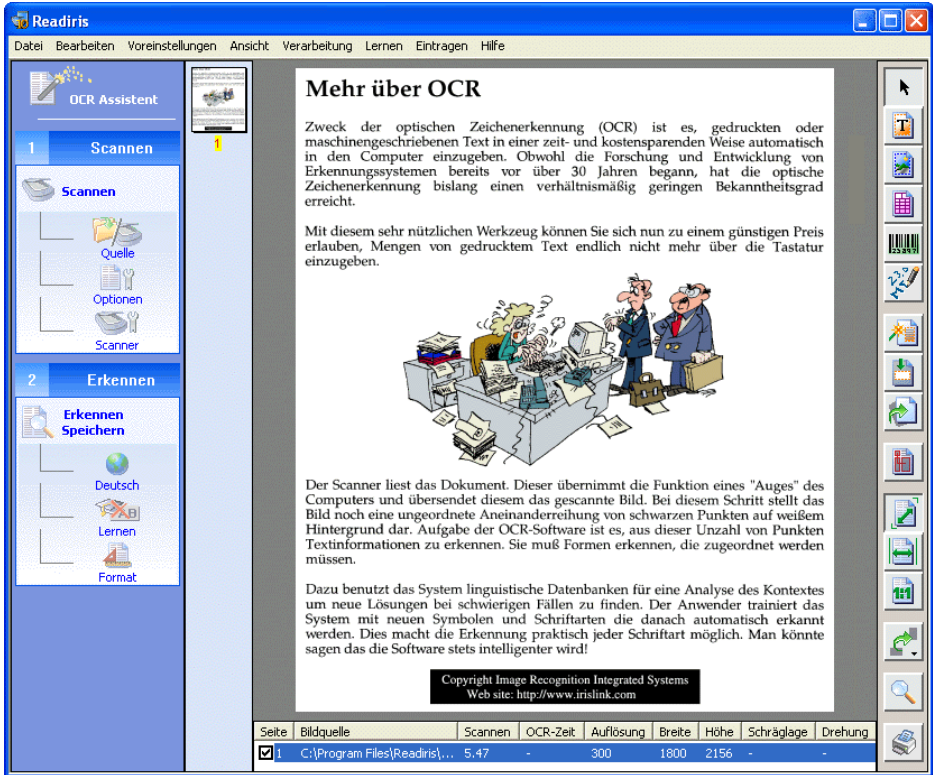
## EINS, ZERLEGUNG EINES GESCANNTEN BILDES

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

Zwei gesonderte Zonen werden in jedem Fall manuell gezeichnet: Strichkodezonen und Handschriftzonen. (Weitere Informationen über das Lesen

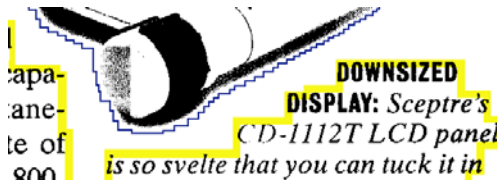
von Strichcodes und Handschriften finden Sie unter „Handschriftlichen Text erkennen“ weiter unten.)

Ein **Farbcode** gibt den Fenstertyp an: Textzonen sind orange, Grafikfenster sind violett und Tabellen sind rosa. Strichcodezonen sind grün und Handschriftzonen sind blau.

Die Zahl der Fenster wird jederzeit in dem Tooltip der Fenster-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. Sie können leicht erkennen, welche Zonen ausgewählt sind und welche nicht: die ausgewählten Fenster erscheinen in vollen Farben, nicht ausgewählte Fenster werden in einem helleren Farbton angezeigt und haben keine Nummer.



The screenshot shows the Readiris 10 software interface. The window title is "Readiris" and the menu bar includes "Datei", "Bearbeiten", "Voreinstellungen", "Ansicht", "Verarbeitung", "Lernen", "Eintragen", and "Hilfe". The main window is divided into a left sidebar and a central workspace.

**Left Sidebar:**


- OCR Assistent**
- 1 Scannen**
  - Scannen
  - Quelle
  - Optionen
  - Scanner
- 2 Erkennen**
  - Erkennen Speichern
  - Deutsch
  - Lernen
  - Format

**Central Workspace:**

**Mehr über OCR**

2 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.

3 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.

4 

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>

**Bottom Panel:**

Seite	Bildquelle	Scannen	OCR-Zeit	Auflösung	Breite	Höhe	Schräglage	Drehung
<input checked="" type="checkbox"/> 1	C:\Program Files\...	5,13	-	300	1800	2156	-	-

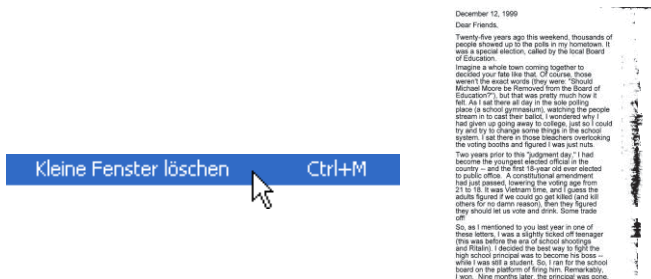
Die Seitenanalyse wird als Standardeinstellung eingestellt. Wenn die Seitenanalyse inaktiviert wurde weil Sie zum Beispiel versehentlich einige Fenster gelöscht haben, können Sie wieder aktiviert werden, indem Sie die Schaltfläche „Seite Analysieren“ in der Bild-Toolbar anklicken.





Wenn Sie asiatische oder hebräische Dokumente bearbeiten, wählen Sie die Dokumentensprache aus, *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. In hebräischen Dokumenten verläuft der Text von rechts nach links. Und wenn Sie vergessen, die richtige Sprache auszuwählen, tun Sie dies nachher. Readiris führt die Seitenanalyse automatisch erneut aus!

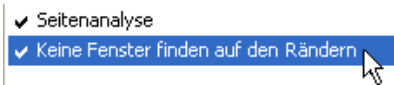
Manche Dokumente weisen zahlreichen „Streupunkte“ auf der Seite auf, das eigentliche Bild ist ggf. mit einem schwarzen Seitenrand umgeben usw. Um alle kleinen Fenster zu entfernen - wobei davon ausgegangen wird, dass sie keinen wirklichen Text enthalten - und das verbleibende Fenster neu anzuordnen,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



Eine andere, ähnliche Routine wird automatisch ausgeführt: das Ermitteln von Zonen auf den Seitenrändern. Wenn diese Routine deaktiviert ist, ignoriert die Seitenanalyse alle Zonen, die die Seitenränder berühren. Wenn der Scanner schwarze Ränder um das eigentliche Bild erzeugt, tendiert die Seitenanalyse zum Erkennen von Zonen, an Stellen, an denen sich nur „Rauschen“ befindet.

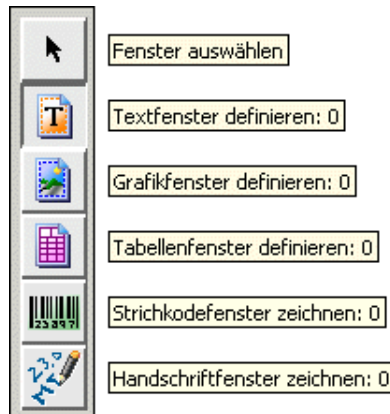


Grafikzonen auf den Seitenrändern bleiben unberührt: Fotos berühren oftmals die Seitenränder, Hintergrundgrafiken bedecken meist die gesamte Seite usw.



## ZWEI, MANUELLE FENSTERBESTIMMUNG EINES BILDES

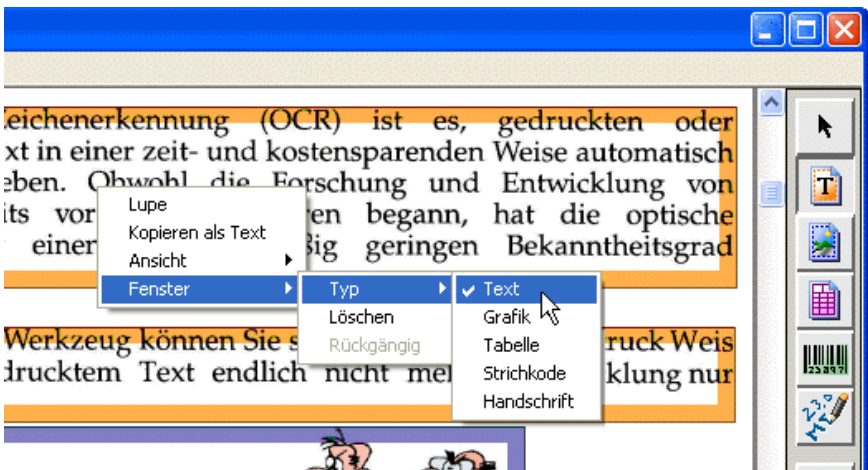
Durch die Seitenanalyse werden die Fenster automatisch bestimmt. Alternativ dazu kann die Fensterbestimmung auch manuell durchgeführt werden mit Hilfe der **Fenster-Tools** von Readiris.



(Wie bereits erwähnt, werden Strichcodefenster und Handschriftfenster immer von Hand gezeichnet: diese Fenster werden von der Seitenanalyse nicht erkannt!)

Um ein Rechteck um die betreffende Zone zu **zeichnen**, wählen Sie das entsprechende Werkzeug auf der Seiten-Funktionsleiste aus und ziehen Sie den Mauszeiger von der oberen linken Ecke zur rechten unteren Ecke des Fensters. (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.

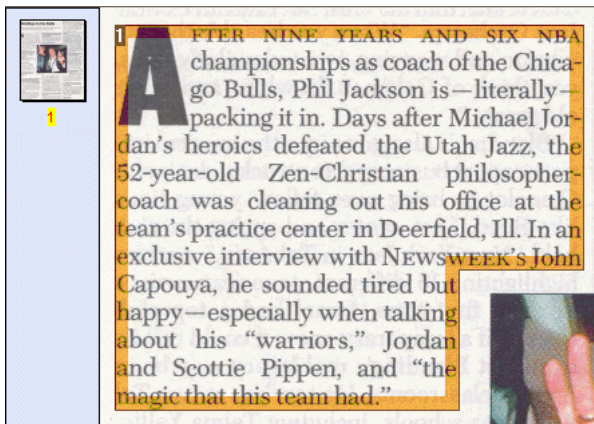


Die Fenster werden automatisch in der Reihenfolge der Erstellung sortiert: die Sortierreihenfolge wird durch Zahlen angezeigt.

Sie können auch „unregelmäßige“ Textblöcke umrahmen, indem Sie **polygonale Fenster** erstellen. Sie können polygonale Fenster erstellen durch Zusammenzufügen rechteckiger 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 oder Strichcodefenster 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.



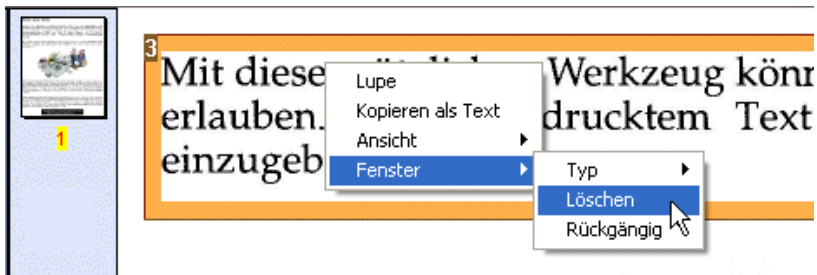
Der Scanner liest das Dokument.

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 **ändern**, 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 internes 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“.

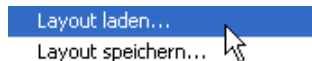
Bearbeiten	
Rückgängig	Ctrl+Z
Ausschneiden	Ctrl+X
Kopieren	Ctrl+C
Einfügen	Ctrl+V
Löschen	Delete
Kleine Fenster löschen	Ctrl+M
Alles auswählen	Ctrl+A

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 Entf 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.

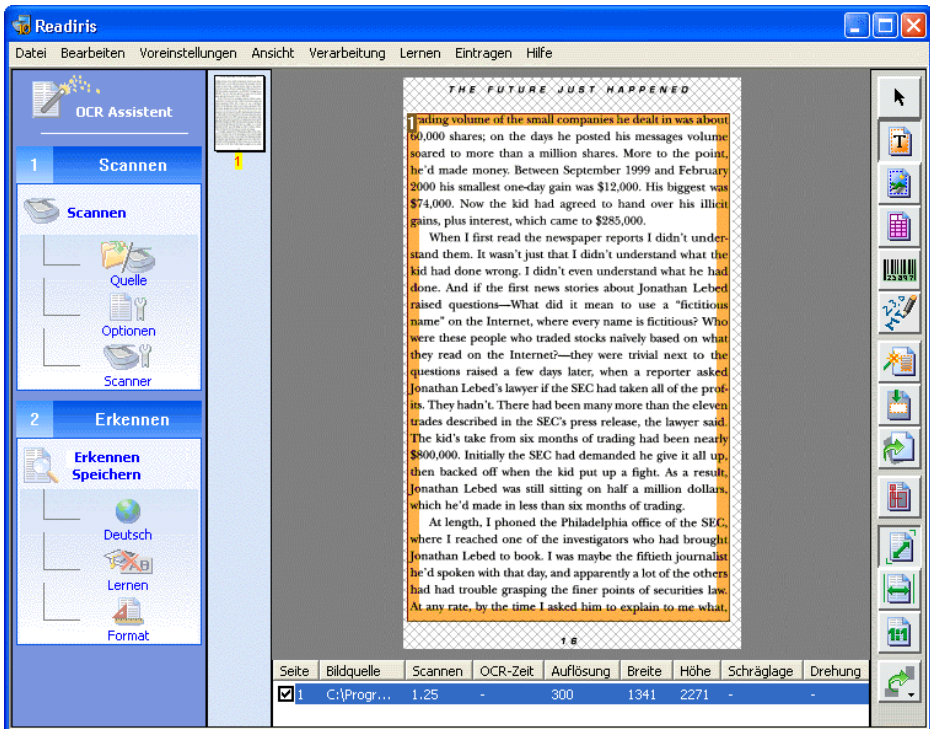
Wenn Sie eine solche Vorlage in den Speicher laden wird die Seitenanalyse automatisch inaktiviert. Die Layoutvorlage bleibt solange aktiviert bis die Seitenanalyse auf der Haupttoolbar wieder aktiviert wird. Wenn Sie ein Layout in den Speicher laden, können Sie die Option „Layout Anwenden für Alle Seiten“ wählen, um das Layout direkt auf alle Seiten des aktuellen Dokuments anzuwenden.



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!

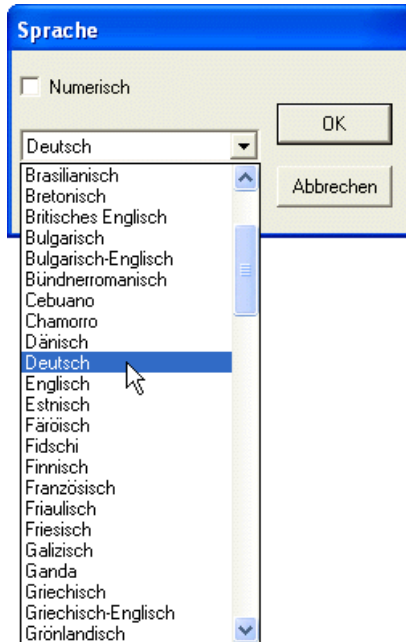
Vorausgesetzt Sie haben die Fensterbestimmung korrekt durchgeführt, dann sind Sie jetzt fast soweit, dass Sie die Zeichenerkennung aktivieren können. Wir sagen „fast“, weil wir die Sprachenauswahl und die Dokumenteinstellungen noch nicht kontrolliert haben!



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.)

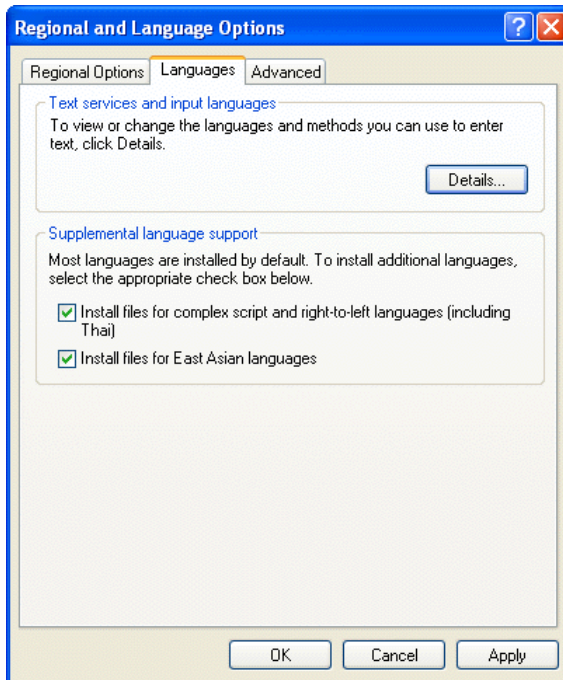
Readiris beschränkt sich nicht nur auf Deutsch: es werden bis zu 117 **Sprachen** unterstützt! Readiris unterstützt alle amerikanischen und europäischen Sprachen sowie die zentraleuropäischen, griechischen, türkischen, kyrillischen („russischen“) und baltischen Sprachen.

Optional können Sie **hebräische** und **asiatische Dokumente** lesen: das Sondermodul „Hebräisches OCR-Modul“ ermöglicht, wie der Name schon sagt, das Erkennen von Dokumenten in Hebräisch, die Software-Option „Asiatisches OCR-Modul“ ermöglicht das Erkennen von Japanisch, Vereinfachtes Chinesisch, Traditionelles Chinesisch und Koreanisch. (Vereinfachtes Chinesisch wird auf dem chinesischen Festland und in Singapur verwendet, während traditionelles Chinesisch in Hongkong, Taiwan, Macau und den chinesischen Gemeinschaften in Übersee verwendet wird.)

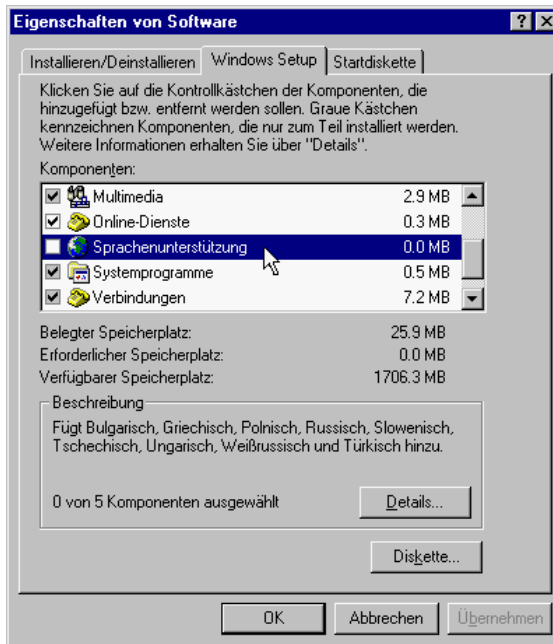
Denken Sie auch daran, dass die britischen und amerikanischen - oder sagen wir „internationalen?“ - Varianten der englischen Sprache ebenfalls differenziert werden. Dies gilt auch für Spanisch und Mexikanisch.

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 und hebräische Dokumente anzuzeigen und zu bearbeiten, können Sie eine asiatische bzw. hebräische Version des Betriebssystems Windows installieren. Sie können derartige Dokumente jedoch auch in Word 2003, Word 2002 oder Word 2000 anzeigen und bearbeiten. Office 2003, Office XP und 2000 wurden speziell für die Arbeit mit Dokumenten in einer Vielzahl von Sprachen entwickelt. Weitere Informationen zu diesem Thema finden Sie in der Readiris-Datei „Liesmich“.

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 „**Zweifelfällen**“, so wie das „O“, das als „0“ gelesen werden könnte. Ein anderes typisches Beispiel ist „l“ und „1“, 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 „1“.

Folgendes Beispiel illustriert die Schattenvariationen von „l“ und „1“. 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.

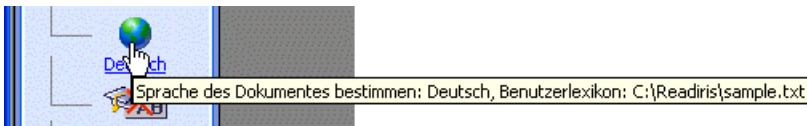
193 1950s. 1hr  
Well, Rossellini

## **BENUTZERLEXIKA STEIGERN DIE LINGUISTIK**

Sie können das linguistische „Feedback“ erweitern, indem Sie es anpassen: so leistungsfähig die Standard-Lexika auch sein mögen, Benutzer der Readiris Corporate können die **OCR-Präzision** weiter steigern, indem sie mit dem Befehl „Benutzerlexikon“ im Menü „Voreinstellungen“ Benutzerlexika laden.

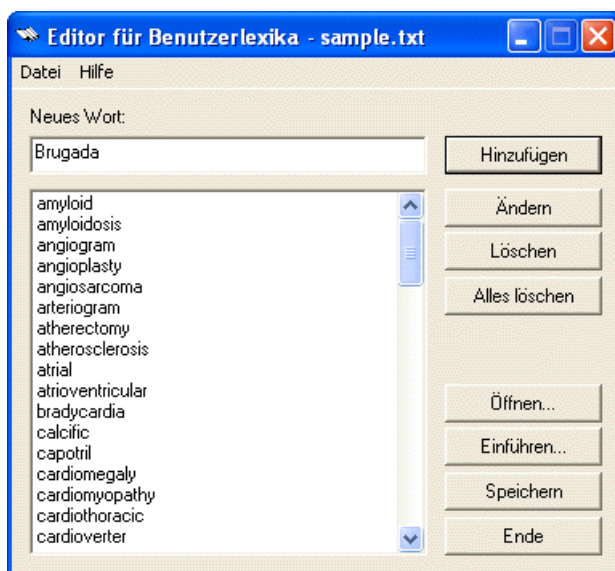


Der Tooltipp zur „Sprache“ Schaltfläche gibt an, welches Benutzerlexikon zurzeit aktiv ist.

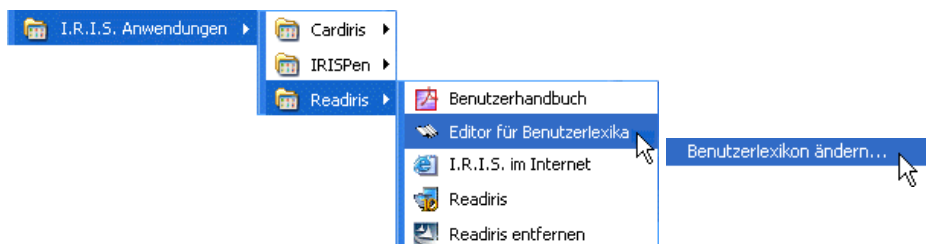


Benutzerlexika sind Wörterlisten mit Begriffen, die nicht in den „Allzweck“ Basislexikon vorhanden sind, beispielsweise technische, wissenschaftliche, rechtliche oder andere **unternehmensspezifische Begriffe**.

Readiris liegt ein Editor für Benutzerlexika bei, mit dem Sie solche **Benutzerlexika** erstellen und pflegen können. Dieses Programm ist äußerst benutzerfreundlich; bei Fragen können Sie jederzeit in der Online-Hilfe nachschlagen.



Sie finden diesen Editor im Untermenü „I.R.I.S. Anwendungen - Readiris“ sowie im „Voreinstellungen“ Menü von Readiris.





## BEI BEDARF ÄNDERT READIRIS DIE SPRACHEN

---

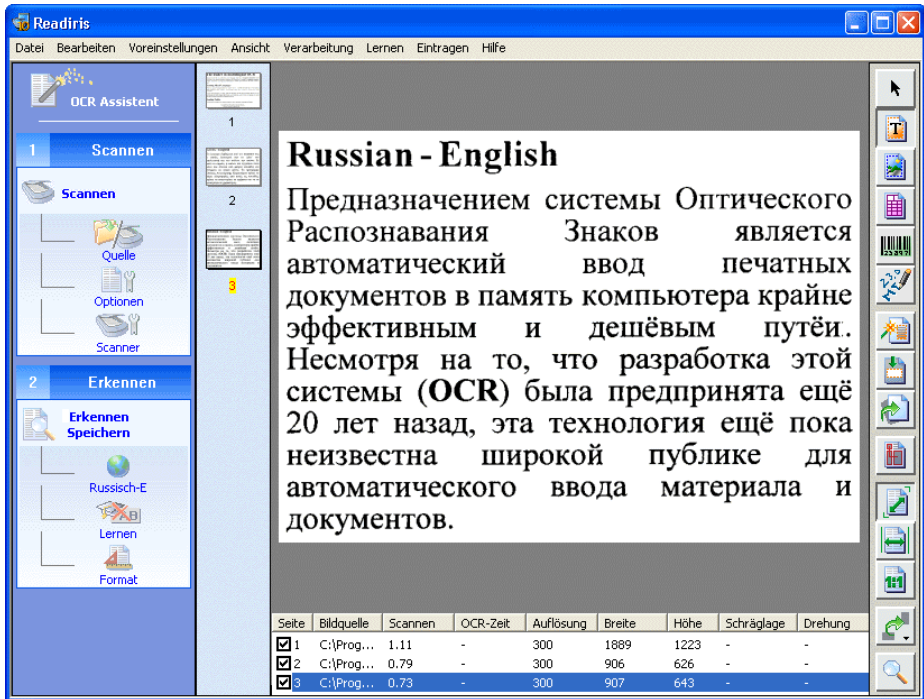
Aber es geht noch weiter: Readiris kann mitten in einem Satz die Sprache wechseln, ohne benutzerseitigen Eingriff! Wenn „westliche“ Wörter in griechischen, kyrillischen, hebräischen oder asiatischen Dokumenten auftauchen - zahlreiche nicht transkribierbare Eigennamen, Markennamen usw. werden mit den bekannten westlichen Zeichen geschrieben -, so kann Readiris automatisch zum korrekten Alphabet umschalten. Anders gesagt: Sie können ein **gemischtes Alphabet** aus griechischen, kyrillischen, hebräischen oder asiatischen und westlichen Zeichen 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!

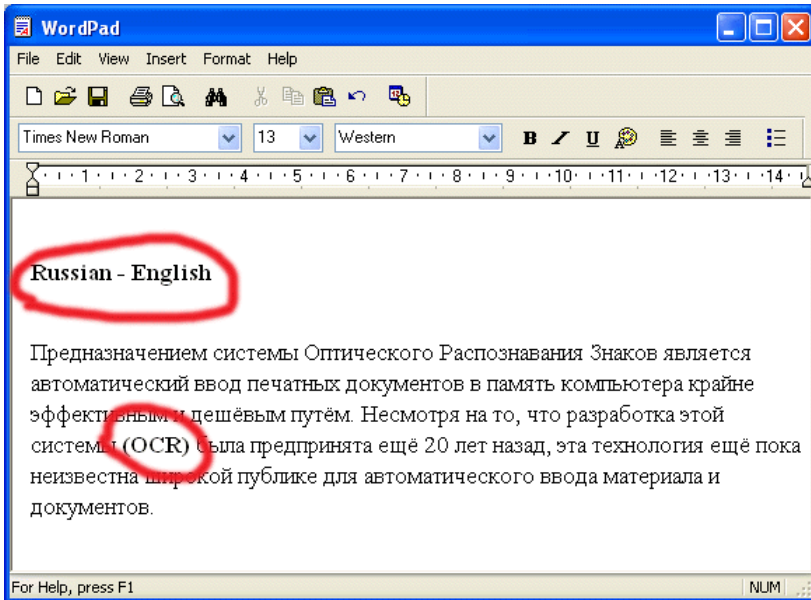


Hier sehen Sie ein Beispiel, in dem ein russischer Text einige Wörter in Englisch enthält. Wenn Sie sich selbst überzeugen möchten, öffnen Sie die Bilddatei ALPHABETS.TIF und führen Sie eine Erkennung der entsprechenden Seite aus!



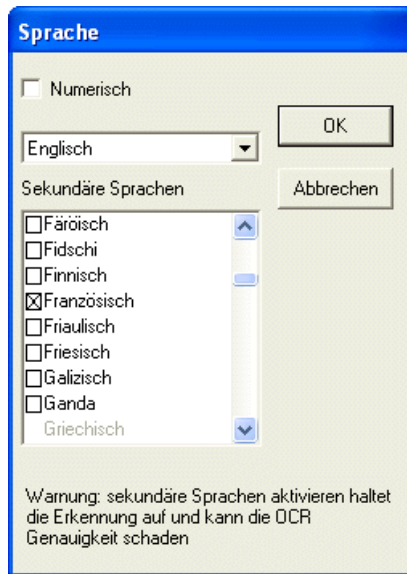


Das Endergebnis sieht aus wie folgt: Sie werden vielleicht doch noch einen kyrillischen **Zeichentyp** selektieren müssen, damit der russische Text richtig angezeigt wird.



## GEMISCHTSPRACHIGE DOKUMENTE EINLESEN

Readiris Corporate entfaltet seine volle Leistung beim Einlesen von Dokumenten, in denen mehrere Sprachen verwendet werden: mit dieser Version können Sie eine Primärsprache und (bis zu 4) **Sekundärsprachen** (aus derselben Sprachgruppe) wählen.

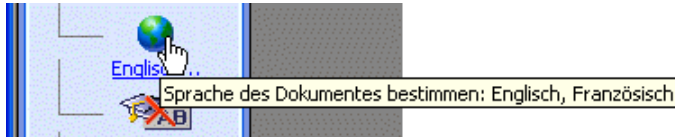


Um ein Dokument zu lesen, in dem die französische Übersetzung neben einem englischen Text aufgeführt ist, können Sie Englisch als Hauptsprache und Französisch als Sekundärsprache wählen. Dabei wird nicht nur der Zeichensatz um Buchstaben mit betonten Zeichen wie ç, é und ù erweitert, sondern beide linguistischen Datenbanken versorgen den OCR-Vorgang mit einem linguistischen „Feedback“, so dass Readiris von selbst erkennt, welche Sprache wo in Ihrem Dokument verwendet wird!

Sie können bis zu 5 Sprachen gleichzeitig auswählen. Es können nur Sprachen aktiviert werden, die zur selben Sprachengruppe gehören: Sprachen, für die eine andere Windows Codepage als die der „Hauptsprache“ verwendet wird, können *nicht* gleichzeitig aktiviert werden. Wählen Sie jedoch keinesfalls Sprachen, die nicht im Dokument verwendet werden: je größer der Zeichensatz, desto langsamer die Erkennung und desto größer die Gefahr von OCR-Erkennungsfehlern!



Falls Sie schnell erfahren möchten, welche Sprachen zurzeit ausgewählt sind: im Tooltip der „Sprache“-Schaltfläche sind die aktiven Sprachen aufgeführt.



## 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ate** 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allen automatisch erkennen, auer wenn Sie sicher sind, dass nur Punktmatrix-Dokumente eingelesen werden! (Selbstverstandlich ist „Automatisch“ der Standardwert.)



Der Schrifttyp ist im Tooltip der Schaltfläche „Erkennen-Speichern“ angegeben: Wenn dem Tooltip keine Angabe hinzugefügt ist, ist die automatische Erkennung der Druckqualität aktiviert. Wenn die Nachricht „Punktmatrix“ im Tooltip erscheint, ist der Punktmatrix-Lesemodus aktiviert.



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



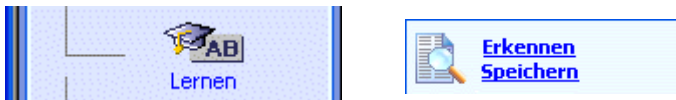
Bei *festen* Schriftarten haben alle Buchstaben die gleiche Breite. Ein „i“ nimmt genau so viel Platz ein wie ein „w“, wie in diesem Satz vorgeführt. Wir denken hier an Dokumente, die mit einer Schreibmaschine geschrieben wurden, wo eine feste Entfernung der Buchstaben von einander vorgegeben ist.

Bei der *Proportional*schrift 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 Proportionalchrift 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 auf die Schaltfläche „Erkennen-Speichern“.



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



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

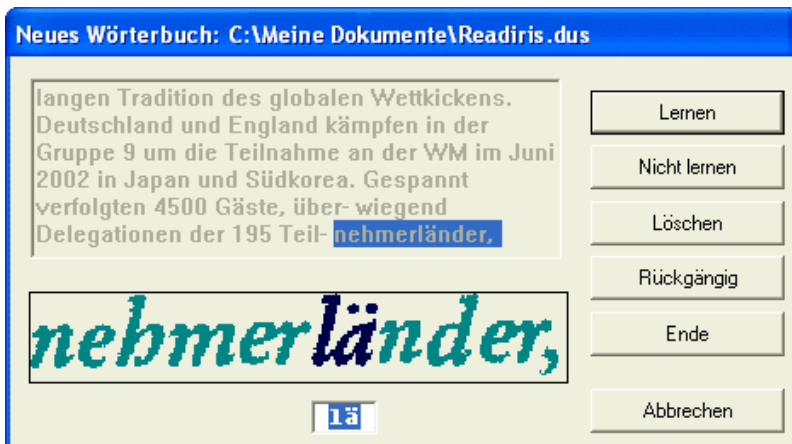
(Das interaktive Lernen findet keine Anwendung auf asiatische Dokumente: Lernen in diesen Sprachen ist zwecklos, weil es sich hier um tausende verschiedener Symbole handelt - und Sie müßten in der Lage sein, die Ideogramme mit einer westlichen Tastatur einzugeben!)

**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 „ $\pi$ “ 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 Titelbalken des Lernfenster 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 Wirking von „Zeichenwörterbucher“ 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.

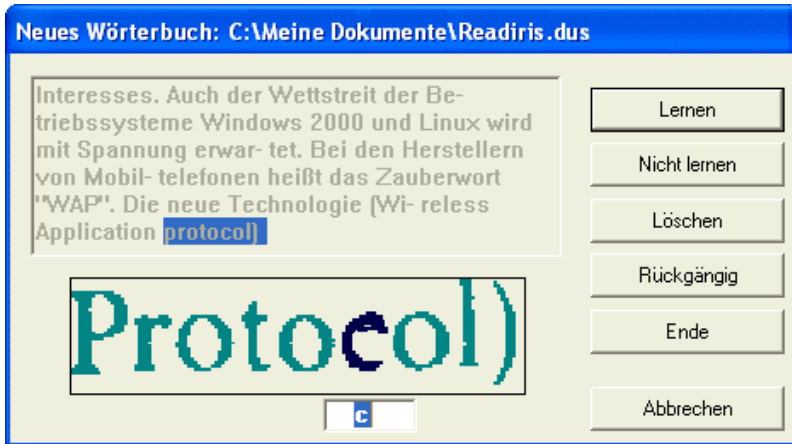
Im oben erwähnten Beispiel hält das System bei einige zusammenlaufenden Zeichen an. Wir klicken auf „Lernen“, um die Zeichen, die nicht mit anderen verwechselt werden können, zu akzeptieren.

### **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, Kaffeelecke 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**

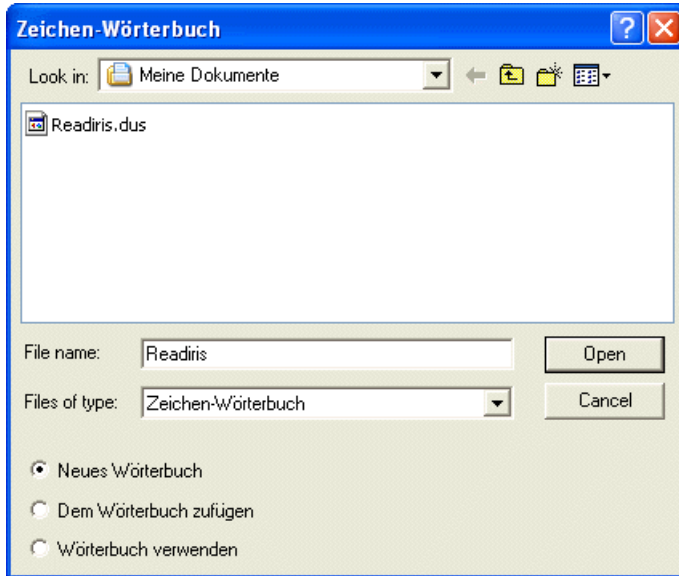
---

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.

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

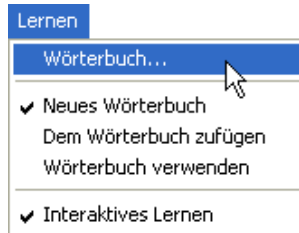
Zeichenwörterbücher sollten in den Speicher geladen werden, um die darin enthaltenen Informationen beim Erkennen ähnlicher Dokumente zu nutzen. So kann Readiris die „Intelligenz“ der Font-Wörterbücher verwenden. Man kann sagen, dass Readiris bei jeder Nutzung intelligenter wird!

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 Standardextension \*.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.)

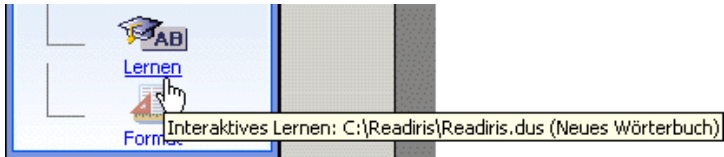
Der Modus „Zufügen“ gibt an, dass die Ergebnisse in einem bereits *existierenden* Wörterbuch gespeichert werden: das Erkennungssystem nutzt bereits die zusätzliche Intelligenz der Wörterbücher und Sie fügen neue Zeichen hinzu. Mit anderen Worten: mit dieser Option wird ein Zeichen-Wörterbuch schrittweise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

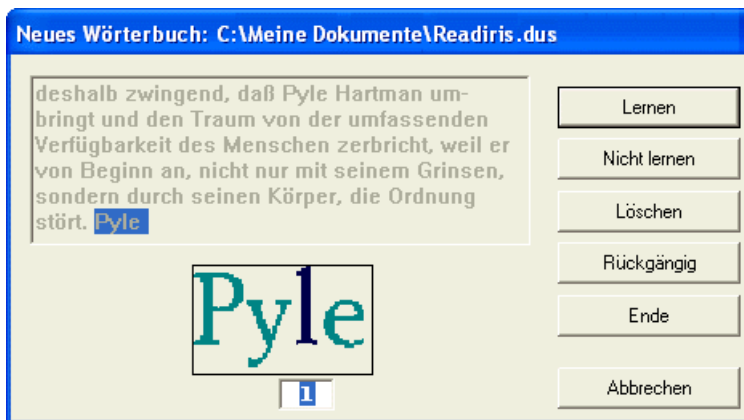
In der letzten Option „Wörterbuch Verwenden“, funktioniert das Wörterbuch nur im Lesen-Modus („Read Only“). Es werden dem Wörterbuch *keine* neuen Zeichen zugefügt. Sie verwenden es aber.

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.

Der Tooltip 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.

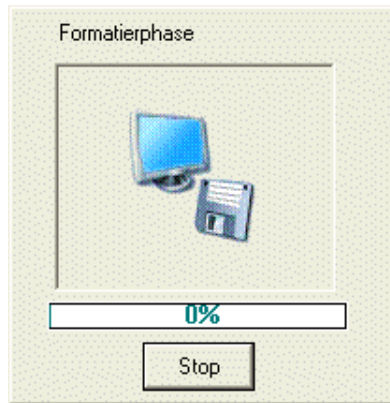


## **DAS ERGEBNIS DIREKT IN IHRE ANWENDUNG SENDEN**

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.)

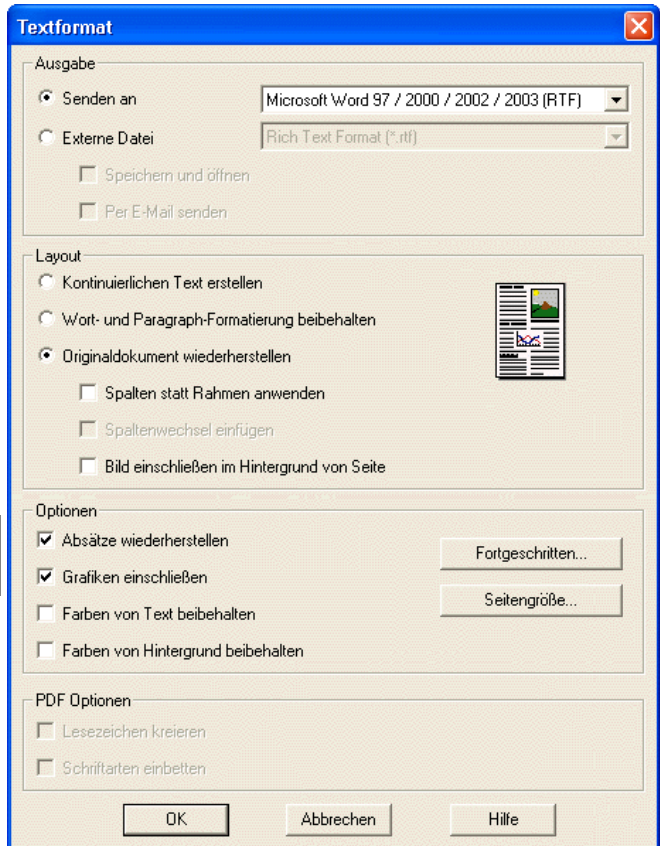


Das gescannte Bild erscheint mit den definierten Bereichen wieder auf dem Bildschirm. Es kann jetzt weiter bearbeitet werden und wird solange auf dem Bildschirm angezeigt, bis Sie eine andere Seite scannen.

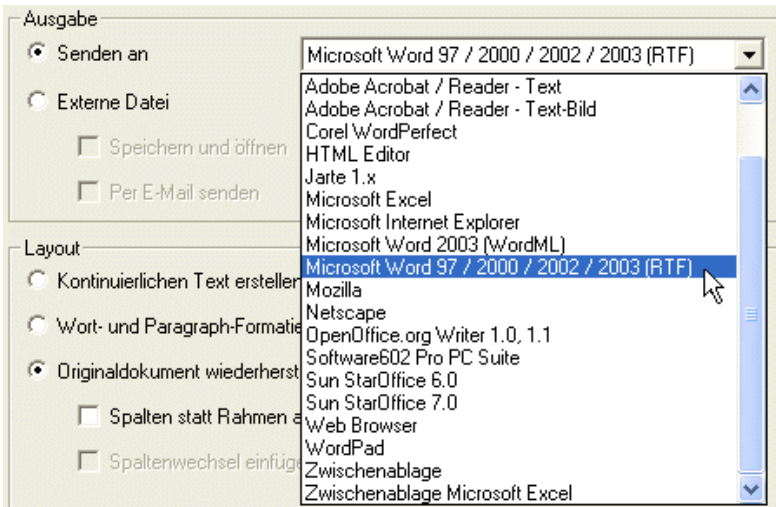
Sie haben jetzt ein Papierdokument in eine editierbare Textdatei verwandelt. Ein Vorgang der 40 Mal schneller ist, als das manuelle Abschreiben! Sie können es jetzt mit dem Bild, das Sie in Ihrem Readiris Fenster haben, vergleichen.

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.



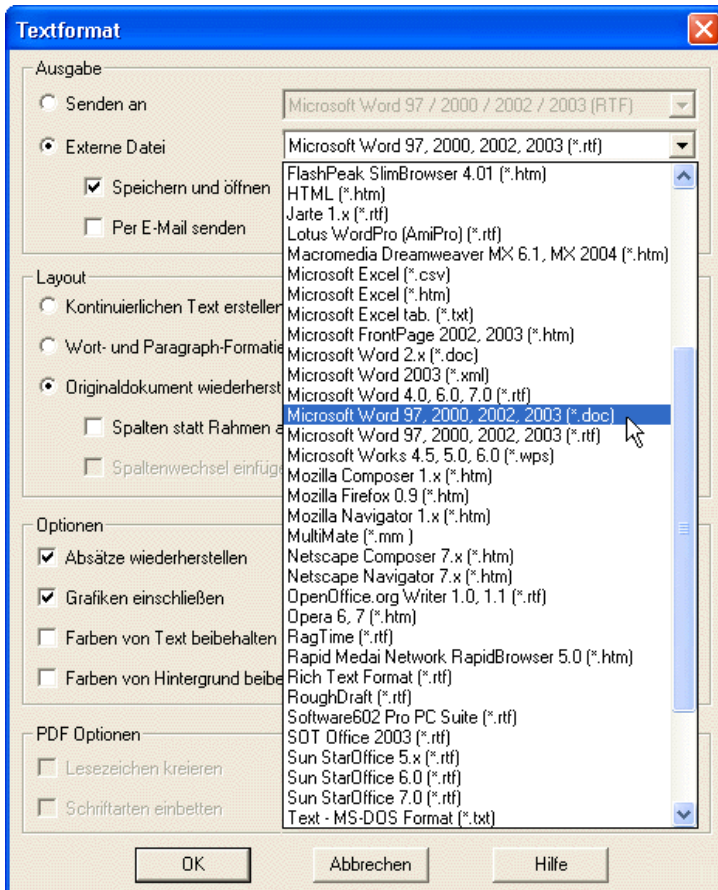


Vergessen Sie nicht, dass Sie mit der Option „Senden an“ den erkannten Text auch in die Windows **Zwischenablage** kopieren können. Es ist also nicht unbedingt notwendig, das Ergebnis zu exportieren... oder es in einer externen Datei zu speichern!

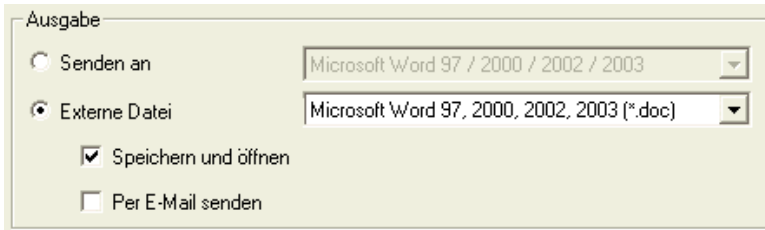
## **DAS ERGEBNIS IN EINER TEXTDATEI SPEICHERN**

---

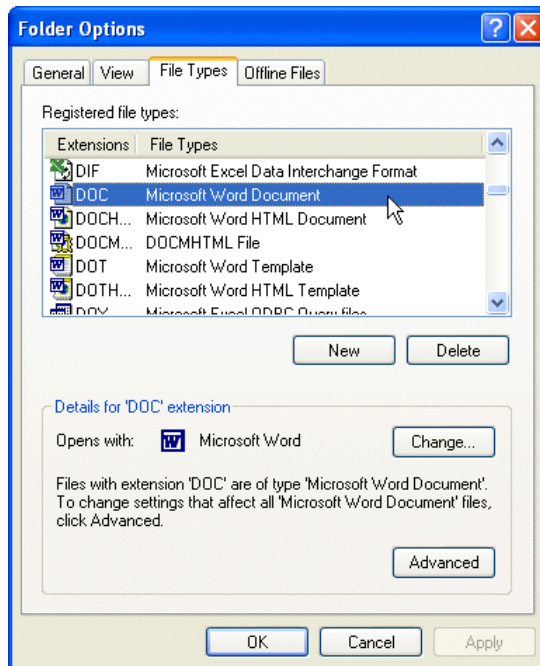
Sie können das OCR-Ergebnis tatsächlich als „externe“ Datei abspeichern. Readiris unterstützt auch hier eine Vielzahl von Dateiformaten, darunter alle bekannten Textverarbeitungen, Tabellenkalkulationen, HTML-Editoren usw. (Readiris unterstützt unter anderem WordML, das neue Textformat von Microsoft Office 2003!)



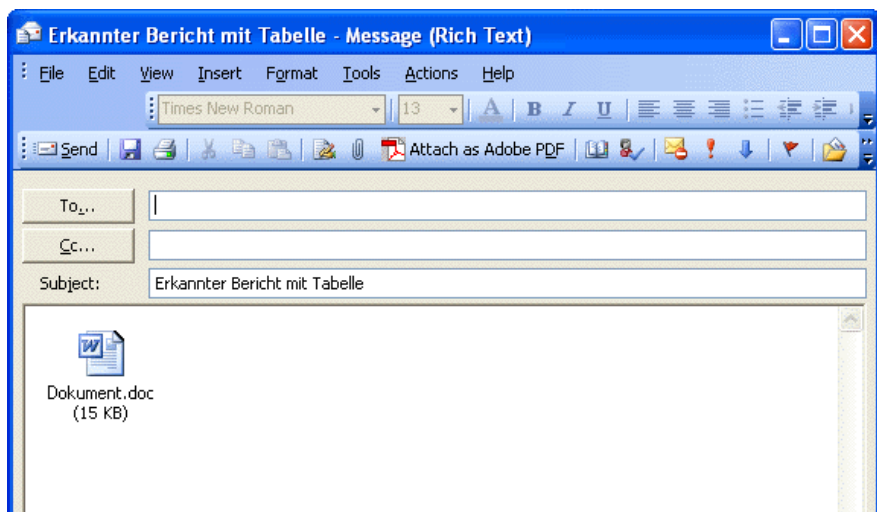
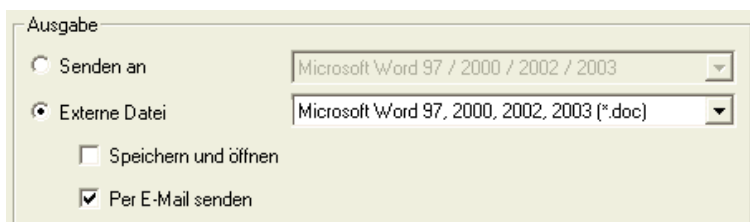
Die Option „Speichern und Öffnen“ ähnelt weitgehend der „Senden“-Funktion: das erkannte Dokument wird nach dem Speichern gleich geöffnet.



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.)

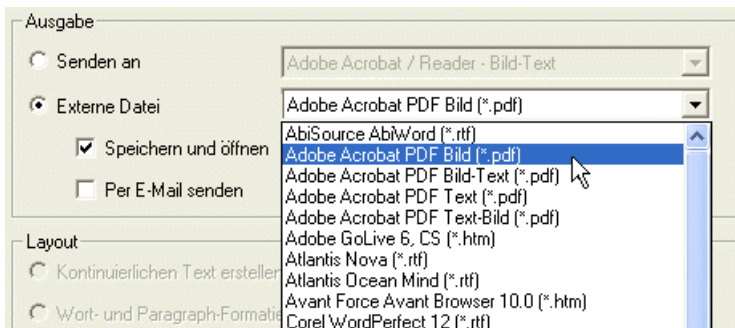
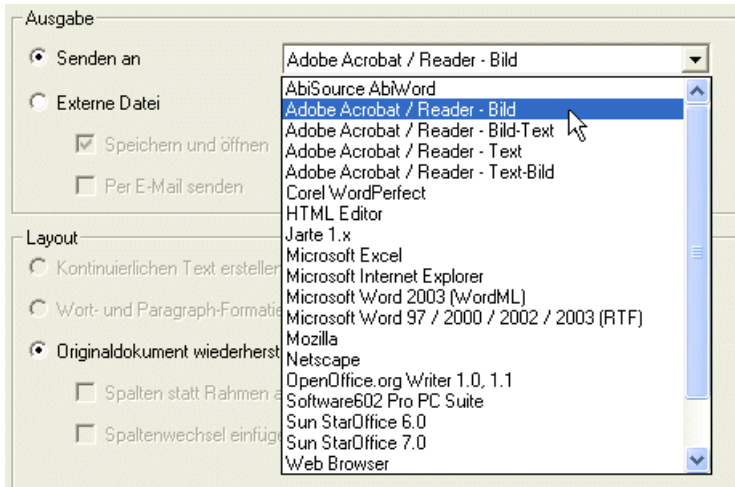


Der Befehl „Per E-Mail Senden“ erzeugt eine neue **E-Mail**-Nachricht, an die das erkannte Dokument angehängt wird. Kennen Sie eine schnellere Methode, ein Papier-Dokument rasch zu verbreiten...?



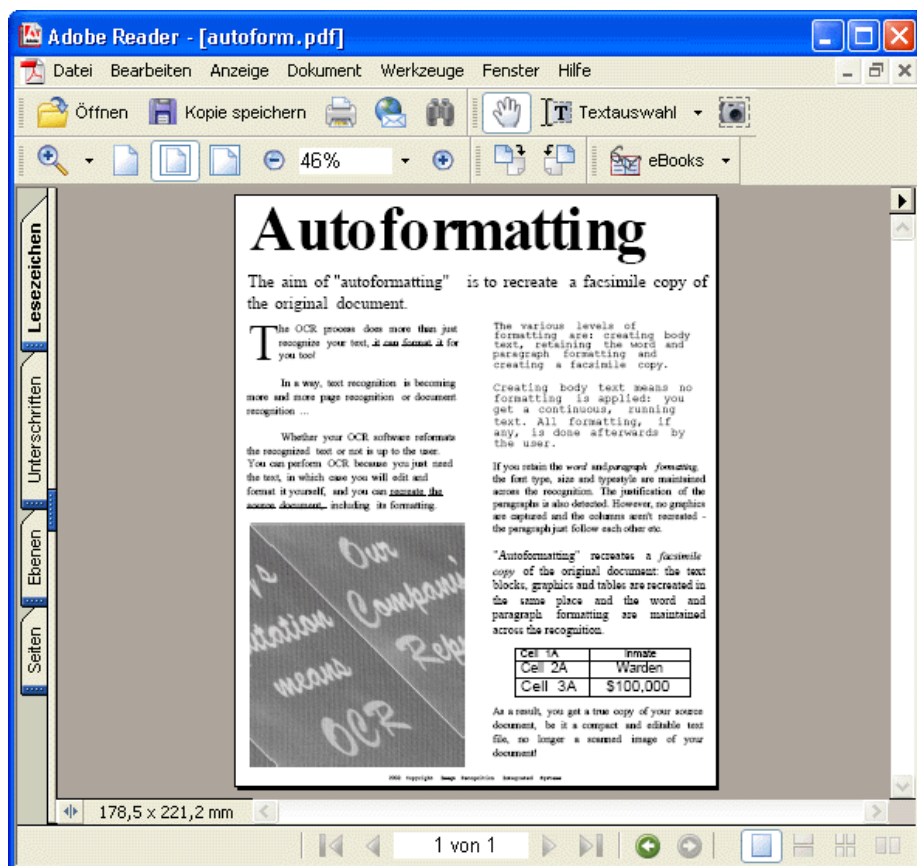
## PORTABLE DOKUMENTE ERSTELLEN...

Wir werden uns ein Format genauer ansehen: **Adobe Acrobat PDF**. Mit Readiris können Sie text- und bildbasierende PDF-Dokumente erstellen.



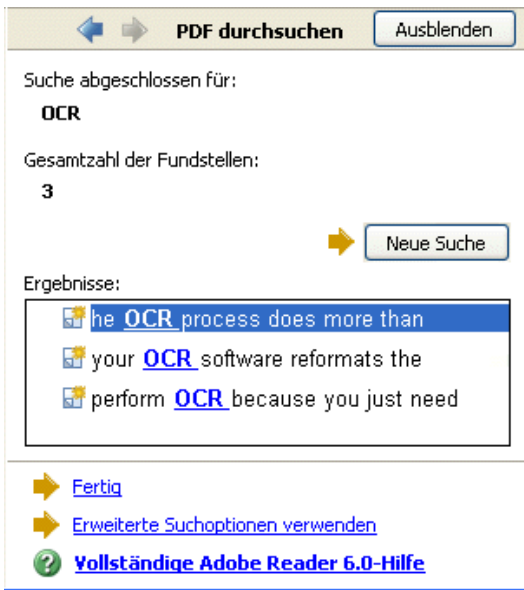
Worin besteht der Unterschied bei diesen Formaten? Wenn Sie das Format „PDF Text“ auswählen, erstellt Readiris eine PDF-Datei, die das Textergebnis enthält. (Es können Grafiken enthalten sein, jedoch nur, wenn Grafikzonen auf der Seite vorkommen - Fotografien, Verzierungen usw.) Anders gesagt: das Seitenabbild ist *nicht* in der einschichtigen PDF-Datei enthalten! Das Format

„PDF Bild“ ist ebenfalls einschichtig, es enthält jedoch nur das gescannte Abbild, keine OCR-Ergebnisse!





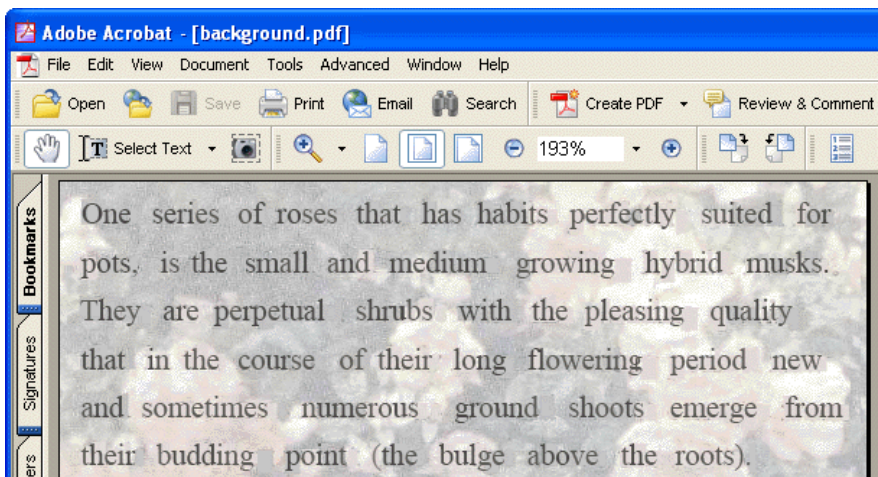
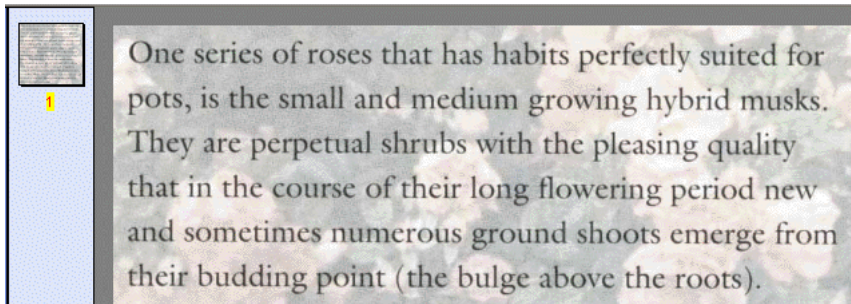
Die Formate „PDF Text-Bild“ und „PDF Bild-Text“ führen zu unterschiedlichen Ergebnissen. Readiris erstellt eine durchsuchbare PDF-Datei, die den erkannten Text *und* das Seitenabbild enthält. In PDF-Dateien „Text-Bild“ ist der Text in der zweischichtigen PDF-Datei *über* dem Seitenabbild angeordnet, in PDF-Dateien „Bild-Text“ ist der Text *unter* dem Seitenabbild platziert. Probieren Sie das Werkzeug „Suchen“ im Adobe Reader aus und Sie werden den Unterschied erkennen!



PDF-Dateien vom Typ „Text-Bild“ sind recht kompliziert: die Pixel des erkannten Textes werden gelöscht, um ein lesbares Dokument zu erhalten! Die Darstellung von erkanntem Text zum Beispiel in Schwarz über Bitmaps schwarzer Zeichen würde zu einem Text mit starker Schattierung führen... Wenn Sie sich selbst überzeugen möchten, können Sie mit dem Bild BACKGROUND.JPG eine



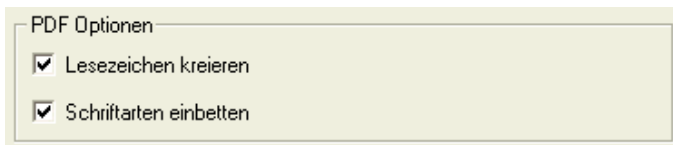
Erkennung ausführen. (Readiris Corporate bietet diese Funktion auch für andere Textformate...)



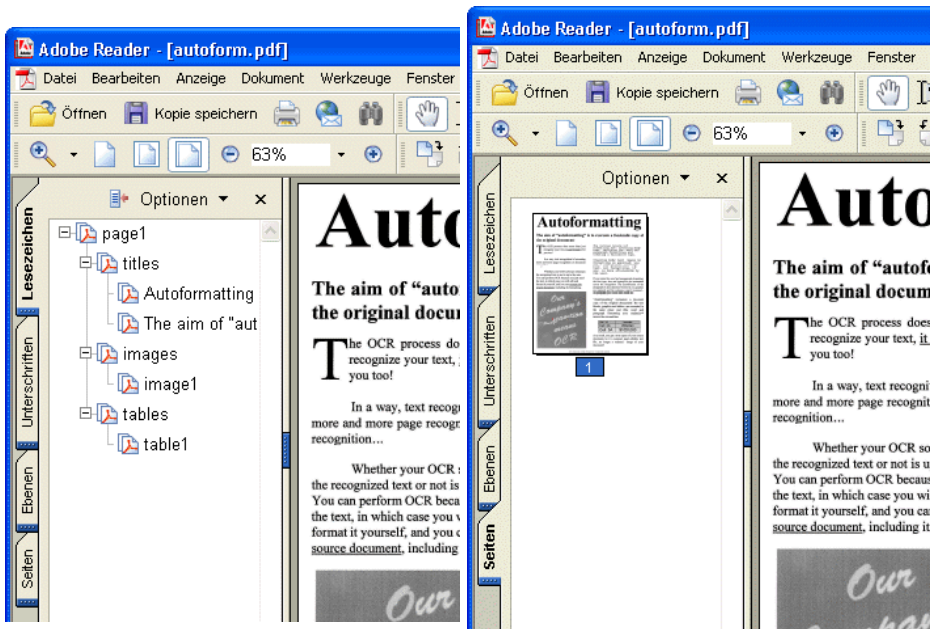
Alle auf Text basierenden PDF-Dateien stellen URLs von Websites als sichtbare Links dar: klicken Sie auf einen solchen Link, um die zugehörige Website anzuzeigen!



Klicken Sie auf die Schaltfläche „Format“, um einige Optionen zum Acrobat PDF-Format kennen zu lernen: „Lesezeichen Kreieren“ und „Schriftarten Einbetten“.

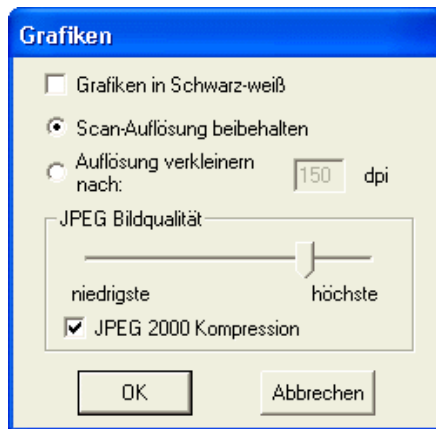
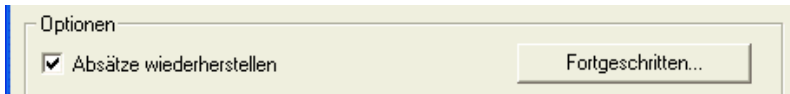


Die Option „Lesezeichen Kreieren“ sorgt dafür, dass ein **Lesezeichen** („bookmarks“) für jedes Dokumentelement erzeugt wird - für Grafiken wie für Textblocks und Tabellen. Für Textbereiche verwendet Readiris einen intelligenten Algorithmus, der einen Titel, eine „Zusammenfassung“ für jeden Bereich erstellt. 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!

Eine weitere Option für PDF-Dateien ist unter den fortschrittlichen Grafikoptionen im Dialogfeld „Format“ versteckt. Wenn Sie PDF-Dateien erzeugen, können Sie die Komprimierungsmethode für die in PDF-Dateien enthaltenen Grafiken auswählen - JPEG oder JPEG 2000. (JPEG 2000 ist die neueste, kompaktere Version des JPEG-Standards.)



## ... 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“

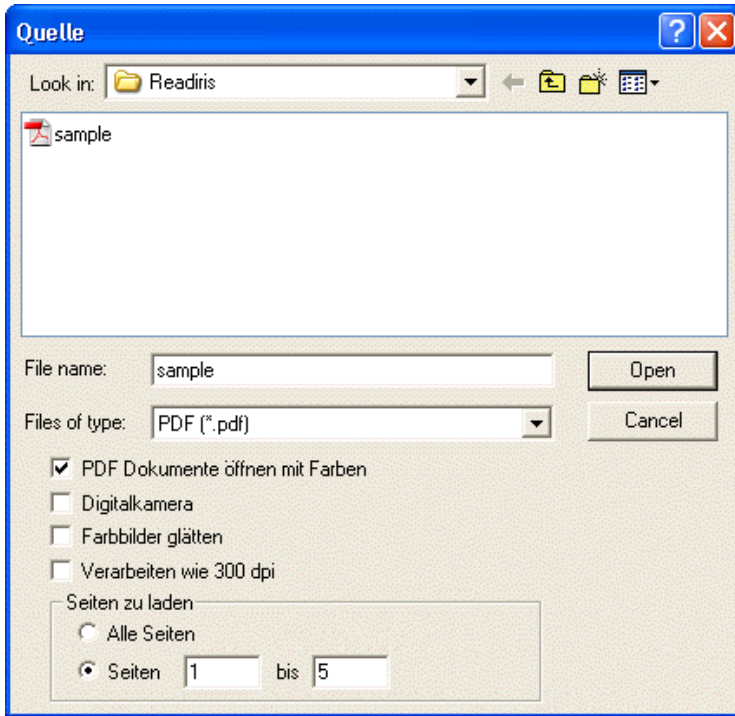
PDF Datei durch und speichern das OCR-Ergebnis... als textgestütztes PDF Dokument! Textgestützte PDF Dateien können durchsucht und bearbeitet werden, „nur-Bild“ PDF Dateien nicht.

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!

Zwei wichtige Feinheiten sind zu erwähnen: Readiris öffnet durch Kennwort geschützte PDF-Dateien nicht, auch wenn alle anderen PDF-Sicherheitsmechanismen von Readiris umgangen werden! (Genauer gesagt: „Master-Kennwörter“, durch welche die Berechtigungen für PDF-Dateien festgelegt werden, werden von Readiris umgangen, „Benutzerkennwörter“, die zum Öffnen eines PDF-Dokuments benötigt werden, werden von Readiris jedoch nicht umgangen.) Zweitens: Readiris konvertiert keine PDF-Dateien, in denen mit JPEG 2000 komprimierte Bilder enthalten sind.

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.



Zweitens möchten Sie möglicherweise festlegen, welche Seiten Sie konvertieren möchten. Wenn Sie, sagen wir mal, nur ein einzelnes Kapitel einer längeren PDF-Veröffentlichung erfassen möchten, wäre es wenig sinnvoll, das gesamte Buch in Readiris zu laden... Durch die Vorgabe eines sinnvollen Bereichs **Seiten zu laden** können Sie viel Zeit sparen! (Das gilt auch für TIFF-Bilder, die aus mehreren Seiten bestehen.)

## **MEHRERE SEITEN ERKENNEN**

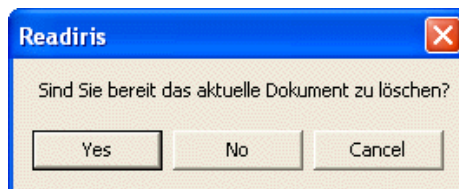
---

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

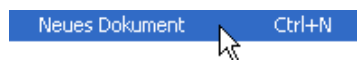
Sie können jetzt den gescannten Text in Ihrem Textverarbeitungsprogramm oder Ihrem Texteditor öffnen, ihn in Ihr Desktop-Publishing-Programm oder eine andere textbasierte Software importieren. Vergleichen Sie den Text mit dem Bild, das im Readiris Fenster angezeigt wird.

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.



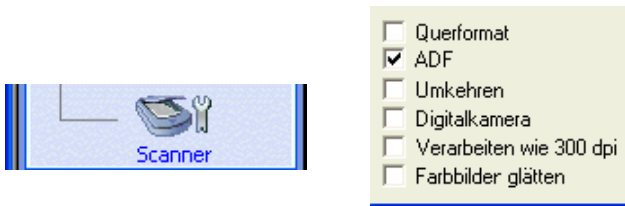
Antworten Sie „nein“, um die Seiten zum aktuellen Dokument hinzuzufügen, oder „ja“, um ein neues Dokument zu erstellen. Diese Antwort hat dieselbe Wirkung wie der Befehl „Neues Dokument“ im Menü „Datei“.





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.

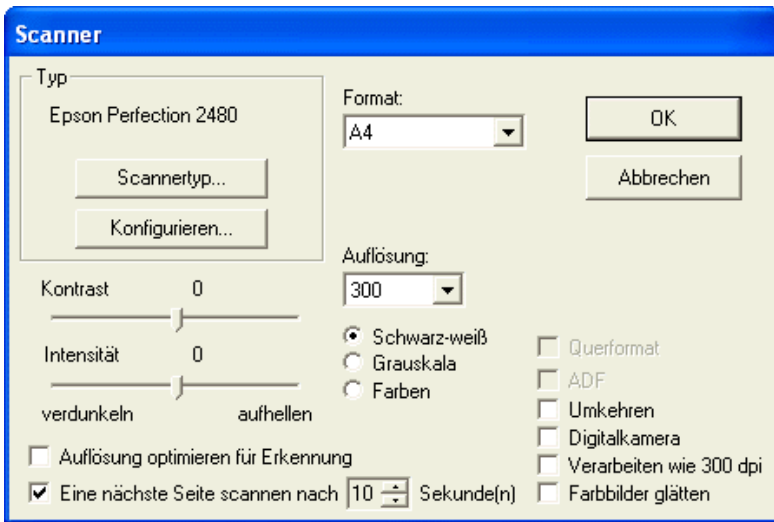


Legen Sie die Seiten Ihres Dokuments in den automatischen Dokumenteneinzug und beginnen Sie den Scanvorgang: alle Seiten werden gescannt, bis der Dokumenteneinzug leer ist.

Wenn der Flachbettscanner nicht mit einem Dokumenteneinzug ausgestattet ist, können Sie durch das **Scannen mit Zwischenraum** mehrseitige Dokumente effizient scannen. Der Scanner scannt automatisch nach Ablauf einer vom Benutzer vorgegebenen Anzahl von Sekunden eine weitere Seite; die Verzögerung ermöglicht Ihnen, die in den Flachbettscanner eingelegte Seite zu entnehmen und die nächste Seite einzulegen.

Geben Sie in den Scannereinstellungen den Zeitraum vor, den Sie zum Einlegen der nächsten Seite in das Flachbett des Scanners benötigen; klicken Sie auf die Schaltfläche „Scanner“ und geben Sie einen geeigneten Wert für die Option „Eine Nächste Seite Scannen nach x Sekunde(n)“ an.





Klicken Sie auf die Schaltfläche „Scannen“, um den Scanvorgang zu beginnen. Klicken Sie im Dialogfeld „Scannen mit Zwischenraum“ auf „Abbrechen“, um den automatischen Scanvorgang zu beenden.

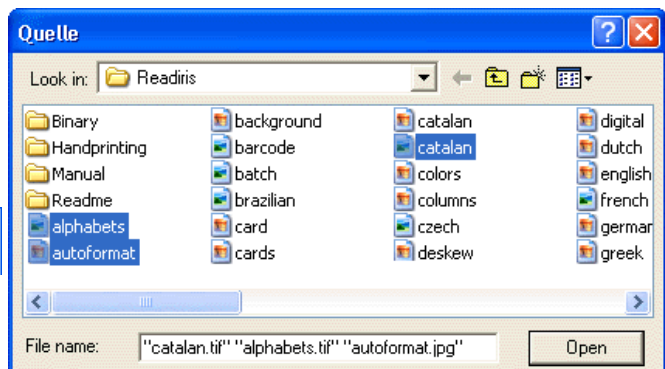


Readiris ist voll flexibel: Sie können die Verzögerung umgehen und sofort scannen und Sie können den Scanvorgang anhalten, um einen Telefonanruf anzunehmen! Klicken Sie im Dialogfeld für automatisches Scannen auf

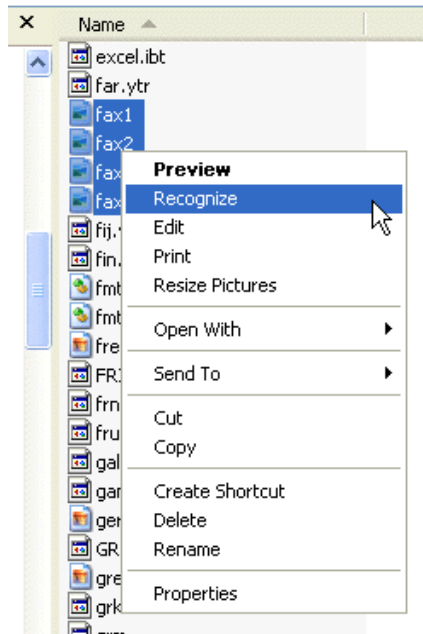


„Pausieren“, um das Scanintervall anzuhalten; wenn Sie den Scanvorgang fortsetzen möchten, klicken Sie auf „Weitergehen“. Oder klicken Sie im Dialogfeld „Scannen mit Zwischenraum“ auf „Jetzt Scannen“, um den Scanvorgang sofort zu beginnen; die Verzögerung wird umgangen!

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.

Im **Dokumentpaneel** werden **statistische Daten** und Informationen über alle gescannten Seiten angezeigt - Quelle und Auflösung des Bildes, für Scannen und Erkennen benötigte Zeit usw. Mit der entsprechenden Option im Menü „Ansicht“ kann das Dokumentpaneel verborgen bzw. wieder angezeigt werden.



✓ Dokumentpanel

Sie erfahren hier auch, welches Bild in den Speicher geladen wurde. Wenn ein aus mehreren Seiten bestehendes Bild geöffnet wurde, ist offensichtlich nur eine einzige Datei mit allen Bildern vorhanden. Wenn Sie zurzeit Dokumente *scannen*, wird im Dokumentpanel einfach nur das Scannermodell angegeben.

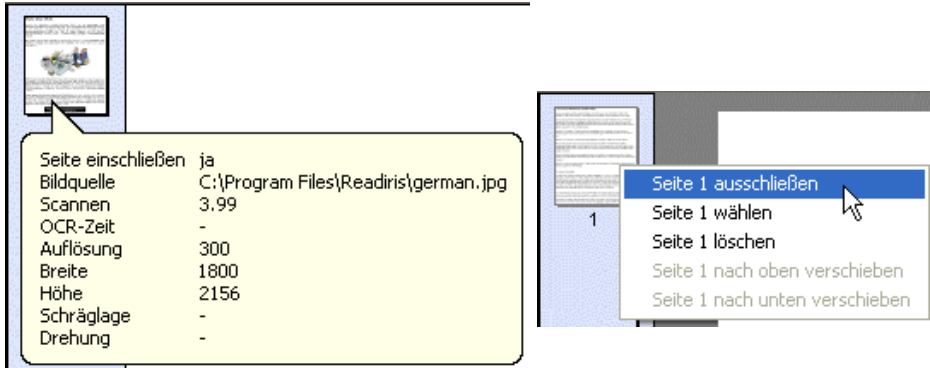
Seite	Bildquelle	Scannen	OCR-Zeit	Auflösung	Breite	Höhe	Schräglage	Drehung
<input checked="" type="checkbox"/> 1	C:\Program Files\Readiris\multipage.tif	1.09	-	300	2000	2388	-	-
<input checked="" type="checkbox"/> 2	C:\Program Files\Readiris\multipage.tif	1.03	-	300	2000	1888	-	-
<input checked="" type="checkbox"/> 3	C:\Program Files\Readiris\multipage.tif	0.98	-	300	1912	2004	-	-

Um die Breite einer Spalte zu ändern, ziehen Sie den Zeiger für Spaltenbreitenänderung. (Die *Reihenfolge* der Spalten kann nicht geändert werden.) Wenn die Spalte zu schmal ist, um die Daten darzustellen, positionieren Sie den Mauszeiger über der Spalte: die Daten werden in einem Tooltipp vollständig angezeigt!

Bildquelle	Scannen
C:\Program Files\...	2.81

Seite	Bildquelle	Scannen
<input checked="" type="checkbox"/> 1	C:\Program Files\Readiris\german.jpg	

Diese Informationen über alle Seiten können Sie auch mit dem Befehl „Information“ im Menü „Datei“ anzeigen und Sie können die Informationen zu einzelnen Seite anzeigen, indem Sie den Mauszeiger über die Miniatur einer Seite in der **Seiten-Toolbar** links positionieren. Diese Funktionsleiste wird angezeigt, sobald Seiten verarbeitet werden. Über diese Funktionsleiste können die Seitenbefehle aufgerufen werden (durch Klicken mit der rechten Maustaste).



Die aktuelle Seite ist im Dokumentpaneel und auf der Seiten-Funktionsleiste hervorgehoben und in der Titelleiste von Readiris aufgeführt.

Um eine Seite zu löschen, markieren Sie die Seite im Dokumentpaneel und drücken Sie die Taste Entf! (Sie können die Seite auch auf der Seiten-Toolbar markieren, mit der rechten Maustaste klicken und im „Kontext“-Menü den Befehl „Seite Löschen“ auswählen.

Wenn Sie einen Überblick über das Dokument benötigen, können Sie mit der Schaltfläche „Drucken“ auf der Bild-Toolbar (oder mit dem Befehl „Bilder Drucken“ im Menü „Datei“) die gescannten **Bilder** in kurzer Zeit **ausdrucken**.

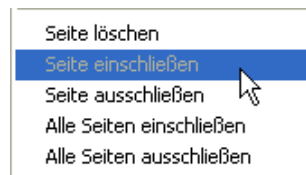


Sie können die aktuelle Seite oder alle Seiten drucken. Wählen Sie die Anzahl Seiten oder Miniaturen („thumbnails“), die auf einer Seite gedruckt werden sollen.

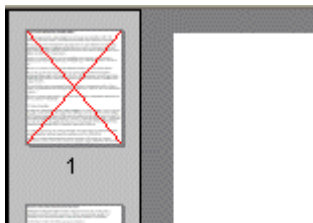


Sie müssen jedoch nicht alle Bilder drucken: im Dokumentpaneel (und mit den entsprechenden Befehlen im Menü „Bearbeiten“ und den entsprechenden Seitenbefehlen im „Kontext“-Menü der Seiten-Toolbar) können Sie Seiten (zeitweilig) ausschließen. Klicken Sie einfach auf die Seitenzahl im Dokumentpaneel, um die Seite vom Drucken (und vom Erkennungsvorgang) auszuschließen. Klicken Sie nochmals auf die Seitenzahl, um die Seite wieder einzubeziehen. Um eine größere Flexibilität zu erreichen, bietet das Menü „Bearbeiten“ entsprechende Befehle für *alle* Seiten.

Seite	Bildquelle	Scannen
<input type="checkbox"/> 1	C:\Program Files\Rea...	3.99
<input checked="" type="checkbox"/> 2	C:\Program Files\Rea...	0.75
<input checked="" type="checkbox"/> 3	C:\Program Files\Rea...	4.50



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.



Seite	Bildquelle	Scannen	OCR-Zeit	Auflösung	Breite	Höhe	Schräglage	Drehung
<input checked="" type="checkbox"/>	2 C:\Program Files...	0.91	0.22	300	2000	1888	-	-
<input checked="" type="checkbox"/>	3 C:\Program Files...	0.88	0.21	300	1912	2004	-	-
<input checked="" type="checkbox"/>	4 C:\Program Files...	0.93	0.21	300	2004	2000	-	-

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.

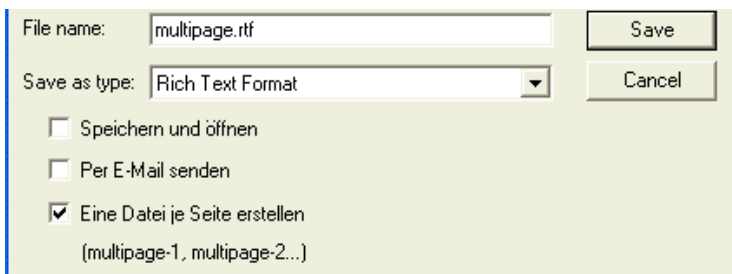
Wenn Sie auf die Schaltfläche „Ende“ klicken, werden alle vom System getroffenen Entscheidungen von nun an ohne benutzerseitige Bestätigung akzeptiert. Mit anderen Worten, das interaktive Lernen wird für *alle* Seiten



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 Zielanwendung gesendet wird, werden mehrere Seiten in einem einzigen Dokument erzeugt.)

Das ist zumindest dann der Fall, wenn die Option „Eine Datei je Seite Erstellen“ beim Speichern des erkannten Dokuments deaktiviert ist. Diese Option sorgt dafür, dass jede einzelne Seite eines mehrseitigen Dokuments in eine gesonderte Datei gespeichert wird. Wenn der Benutzer als Dateiname Text.doc wählt, werden die Dateien Text-1.doc, Text-2.doc usw. genannt. (Diese Option ist nicht verfügbar, wenn Sie die OCR-Ergebnisse an eine Zielanwendung weiterleiten, die Option ist nur beim Anlegen einer externen Datei verfügbar.)



## **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.

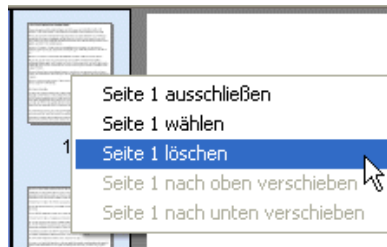
Zuerst das Navigieren. Um zu *einer Seite zu gelangen*, klicken Sie im Dokumentpaneel oder auf der Seiten-Toolbar auf diese Seite. (Oder halten Sie den Mauszeiger über das Symbol der Seite, klicken Sie mit der rechten Maustaste, um das „Kontext-Menü“ zu öffnen, und verwenden Sie den Befehl „Seite Wählen“.) Um zur vorigen Seite zu gehen, können Sie die Taste Bildschirm-



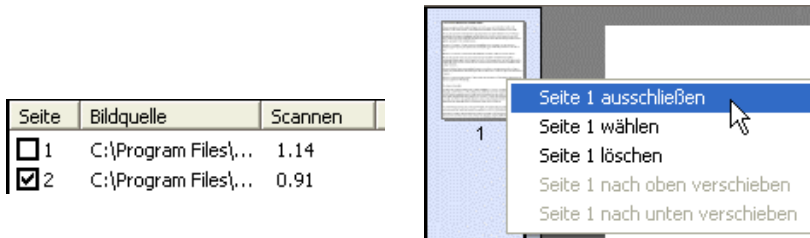
nach-oben drücken, um zur nächsten Seite zu gehen, können Sie die Taste Bildschirm-nach-unten drücken. Drücken Sie die Pos 1-Taste, um zur ersten Seite zu gelangen; drücken Sie die Ende-Taste, um zur letzten Seite zu gelangen. Sie können ebenfalls die entsprechenden Befehle im Menü „Ansicht“ verwenden.

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

Nun zur Bearbeitung des Dokuments: um *eine Seite zu löschen*, markieren Sie die Seite im Dokumentpaneel und drücken Sie die Taste Entf. Sie können auch den Mauszeiger auf die Miniatur der Seite platzieren, mit der rechten Maustaste klicken und im „Kontext“-Menü den Befehl „Seite Löschen“ auswählen.

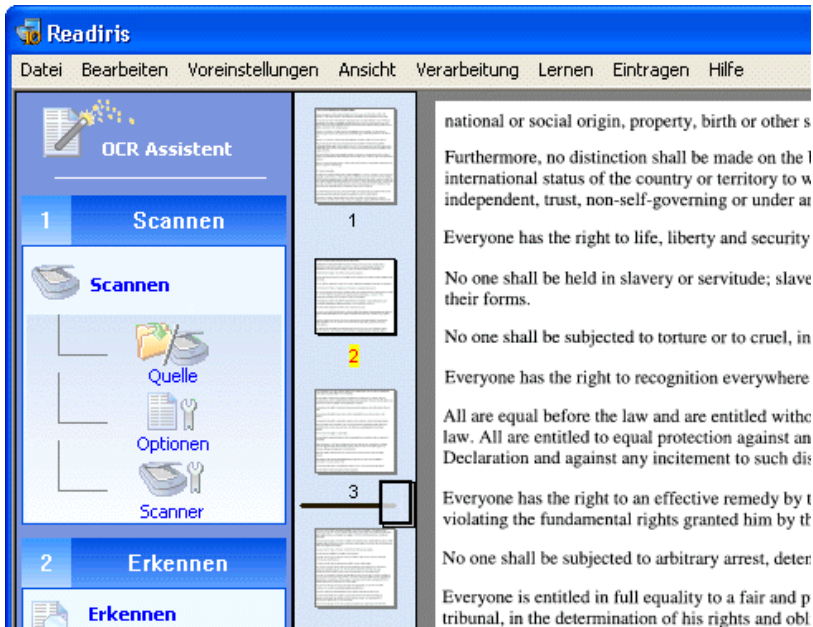


Und wir erwähnen nochmals, dass Sie Seiten zeitweilig vom Erkennungsvorgang (und vom Bildausdruck) ausschließen - nicht löschen - können: die dazu notwendigen Befehle finden Sie im Dokumentpaneel, auf der Seiten-Toolbar und im Menü „Bearbeiten“.



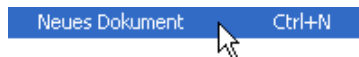
Um eine *Seite nach oben* zu *verschieben*, wählen Sie den Befehl „Seite nach oben Verschieben“, um eine *Seite nach unten* zu *verschieben*, wählen Sie den Befehl „Seite nach unten Verschieben“.

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



## EIN NEUES DOKUMENT STARTEN

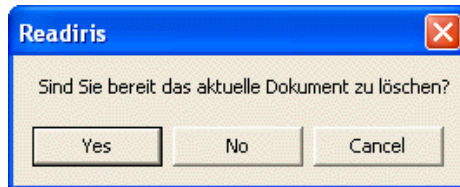
Sie können den Befehl „Neues Dokument“ im „Datei“ Menü wählen, um das aktuelle Dokument zu schließen.



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.

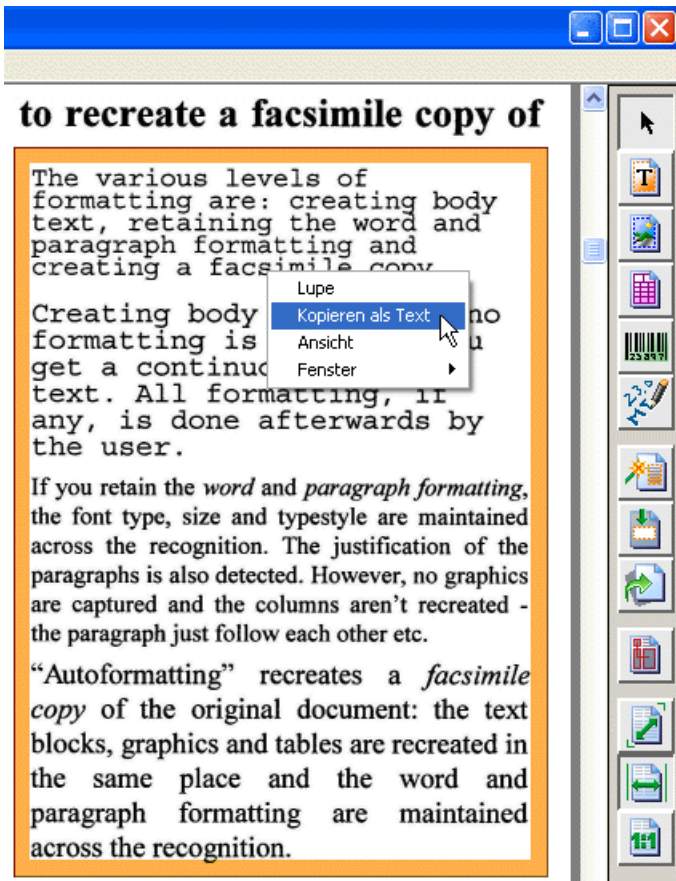
Wenn der OCR-Vorgang bereits ausgeführt wurde und Sie den Scanvorgang (oder das Laden von Bildern) erneut starten, fragt das Programm nach, ob ein neues Dokument angelegt oder das aktuelle Dokument ergänzt werden soll.



## **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.

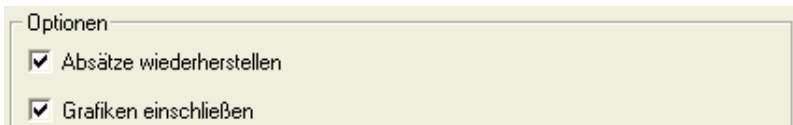
## DIE AUSSICHT DER TEXTAUSGABE

---

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 einen 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

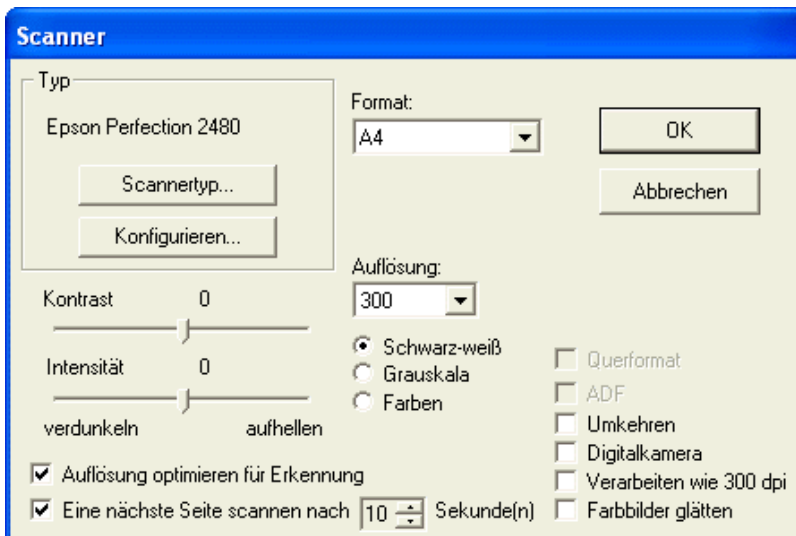
Jetzt werden wir den Scanner einstellen. Wichtige Voraussetzung ist, dass der Scanner und die erforderlichen Treiber richtig installiert wurden.

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.



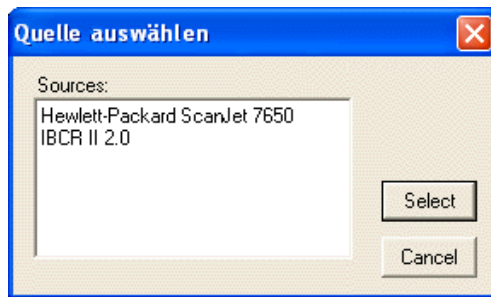
Klicken Sie auf das Schaltfläche „Scannertyp“, um das **Scannermodell** zu bestimmen.





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.

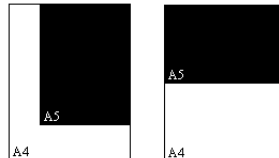
Die Schaltfläche „Konfigurieren“ ist nur aktiviert, wenn der Scanner es erlaubt. Es ermöglicht den Zugriff zu einigen fortschrittlichen Scanparametern: bei Twain Scannern können Sie die Twain Quelle selektieren, indem Sie auf den Knopf „Konfigurieren“ klicken. (Sie können ebenfalls den Befehl „Quelle 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.

Das Scannen mit Zwischenraum haben wir bereits erläutert. Dies ist eine hocheffiziente Methode zum Scannen von mehrseitigen Dokumenten mit einem Flachbettscanner, der nicht mit einem Dokumenteinzug ausgestattet ist.

## **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 Tumbleweed 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 Tumbleweed Communications and Phone.com. Shigeta is also developing new businesses that take advantage of the growth of the Internet and mobile communications.

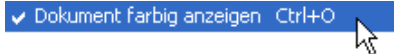
Readiris erzeugt eine Schwarzweiß-Fassung jedes Graustufen- und Farbbilds. Dank intelligenter ProgrammROUTINEN können selbst schwierige Fälle gelöst werden - und so wird unseres „schwieriges“ Bild digitalisiert!

**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 Tumbleweed 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.



Tatsächlich werden Sie auf Ihrem Computerbildschirm kein Schwarzweiß-Bild sehen - selbst dann nicht, wenn Sie Strichmodus-Bilder scannen! Das ist so, weil Readiris die Bilder für eine optimale Bildschirmdarstellung optimiert. Die spezielle Methode von I.R.I.S. zur Hochauflösungsdarstellung konvertiert Schwarzweißbilder zu Graustufenbilder.

### 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 15" Bildschirm. Die hohe Anzeigequalität ist standardmäßig aktiviert, kann jedoch auf hochauflösenden Computerbildschirmen überflüssig sein.

Zoomen Sie auf die tatsächliche Größe (oder größer), um das „Roh“-Bild wie gescannt anzuzeigen!

## 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!) Selbst invertierte Initiale werden erkannt...



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

**L**e Festival de Wallonie est dans sa phase ultime et nale avec l'ouverture de ses dernières sections.  
Cette semaine, c'est la branche

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. Unter anderem ist die Bildauflösung derartiger Bilder im Dokumentpaneel rot markiert.

Seite	Bildquelle	Scannen	OCR-Zeit	Auflösung	Breite	Höhe
<input checked="" type="checkbox"/> 1	C:\Documents and...	1.24	-	96	400	447

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.

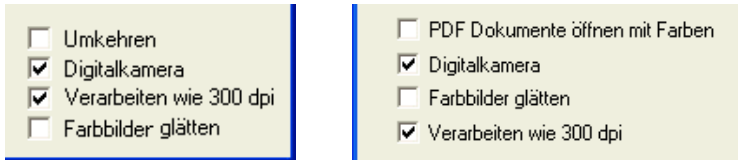
Auflösung optimieren für Erkennung  
 Eine nächste Seite scannen nach  Sekunde(n)

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  
 Farbbilder glätten

PDF Dokumente öffnen mit Farben  
 Digitalkamera  
 Farbbilder glätten  
 Verarbeiten wie 300 dpi

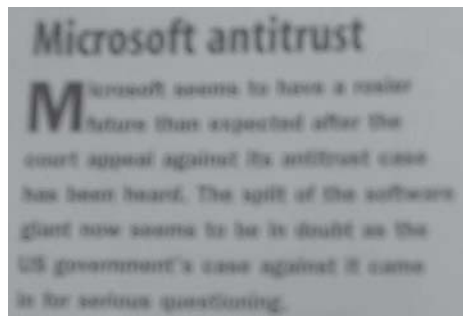
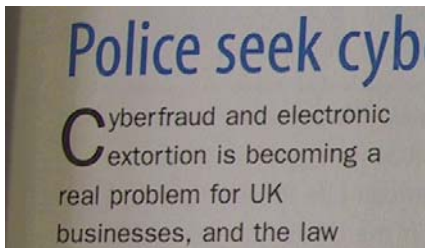
Wenn Ihre Bilder von einer **Digitalkamera** statt von einem Scanner stammen, müssen Sie unbedingt eine Sonderoption aktivieren. Dieser Parameter gilt wiederum für direkte Scanvorgänge wie auch für vorgescannte Bilder.



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.

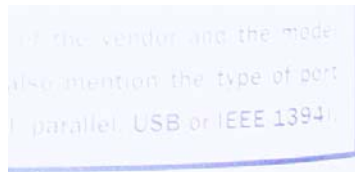




Wählen Sie keine oder eine möglichst geringe Kompression: eine umfangreiche Kompression reduziert die Schärfe des erfassten Texts. Zoomen Sie manuell, um Ihr Dokument zuzuschneiden - einige Kameras werden mit einer Software zum Aneinanderfügen von Bildern („photo stitching“) geliefert; Sie sollten diese Software jedoch nicht für das Dokument-Capturing verwenden.

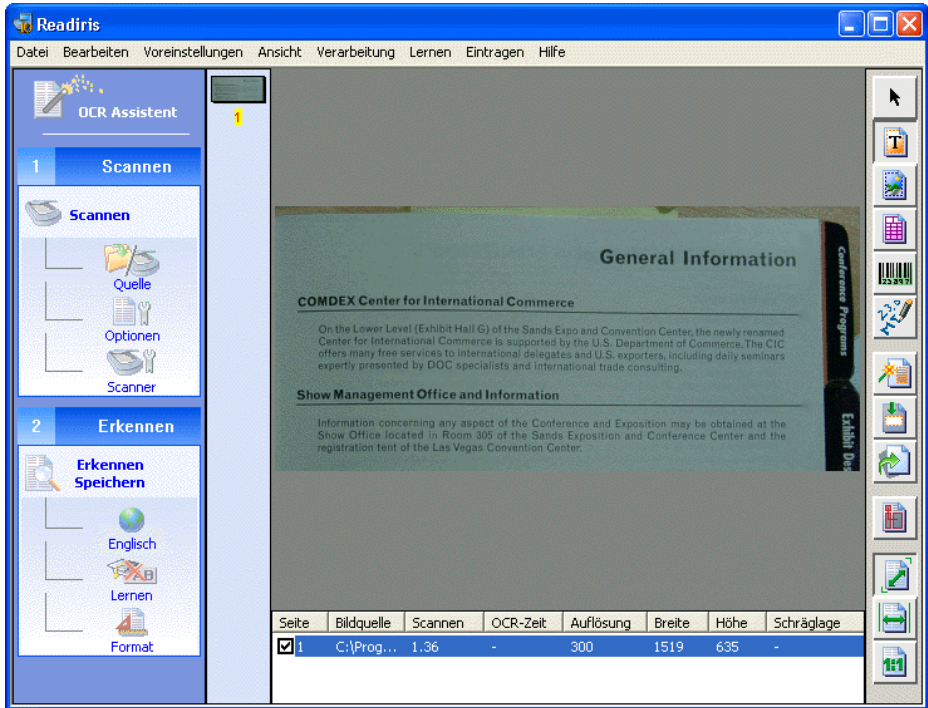
Halten Sie die Kamera direkt über das Dokument, um das Dokument nicht aus einem schrägen Winkel zu fotografieren. Sie sollten vermeiden, dass die Kamera oder Ihre Hände Schatten auf das Dokument werfen! Erzeugen Sie unverwackelte Bilder. Bei Bedarf montieren Sie die Kamera auf einen Standfuß.

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.



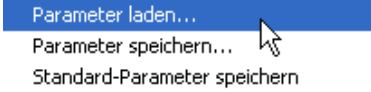
Standard-Parameter speichern

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

---

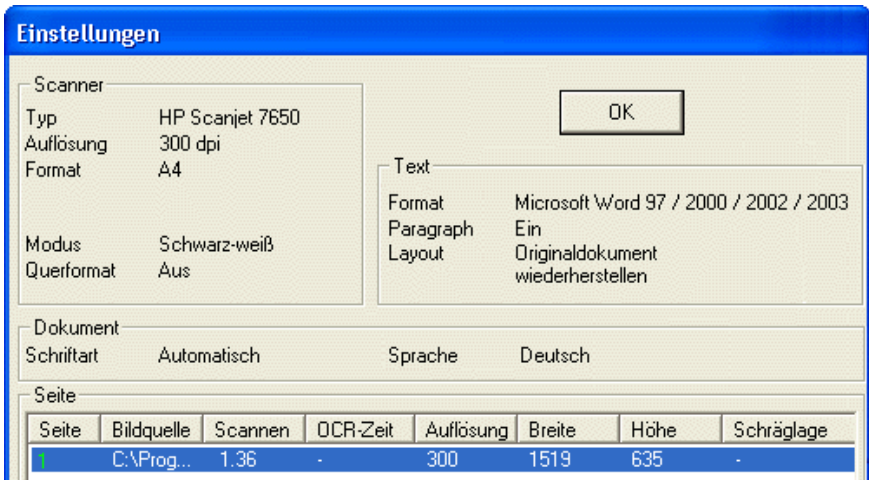
Bei jedem Start werden die Standard Einstellungen geladen. Um zu vermeiden, dass Sie die eingestellten Parameter ständig anpassen müssen, können Sie auch unterschiedliche Einstellungen speichern. Die Befehle sind „Parameter Speichern“ und „Parameter Laden“. Sie befinden sich im Menü „Datei“.



Parameter laden...  
Parameter speichern...  
Standard-Parameter speichern

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.



Dieser Befehl zeigt außerdem die Informationen an, die im Dokumentpanel angegeben sind, für alle Seiten.

## **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°) kann ignoriert werden, da die Genauigkeit des OCR-Vorgangs hierdurch nicht beeinträchtigt wird.

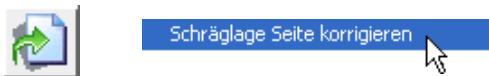
Die Option „Seitenschräglage Korrigieren“ (unter der Schaltfläche „Optionen“ oder im Menü „Voreinstellungen“) legt fest, ob die Schräglage von Seiten, die in



Schräglage eingescannt wurden, korrigiert wird - eine geringfügige Zeilenschräglage wird ignoriert. Diese Option ist standardmäßig deaktiviert.

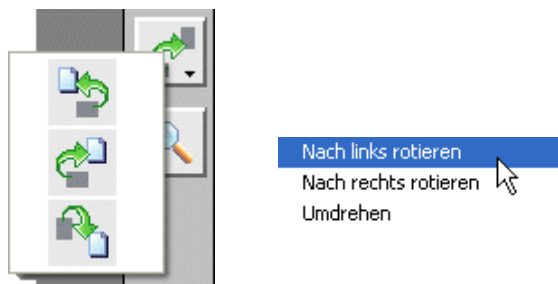


Wurde diese Option fälschlicherweise deaktiviert, verwenden Sie die Schaltfläche „Schräglage Seite Korrigieren“ in der Bild-Toolbar (oder die entsprechende Option im Menü „Verarbeitung“), um Seiten zu begradigen, die schräg eingescannt wurden.



Die Korrektur der Schräglage dauert einige Sekunden: das 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.

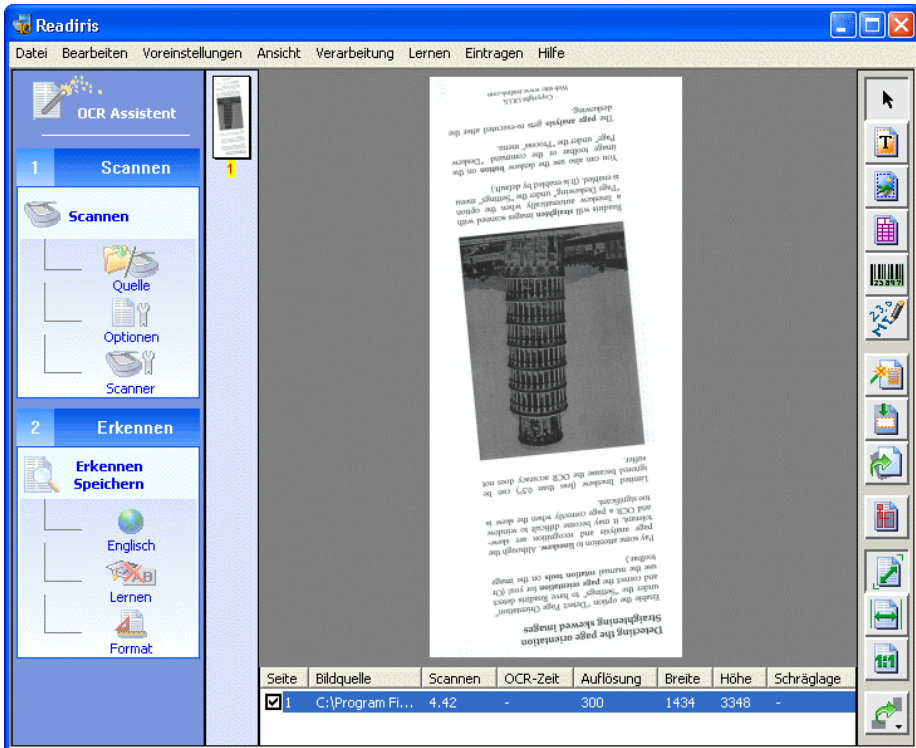


Readiris kann jedoch falsch ausgerichtete Seiten für Sie korrigieren. Aktivieren Sie die Option „Seitenorientierung Entdecken“ mit der Schaltfläche „Optionen“

(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!



Beachten Sie außerdem, dass im Dokumentpanel angezeigt wird, welcher Schrägwinkel korrigiert wurde und welche Drehung ausgeführt wurde!

Breite	Höhe	Schräglage	Drehung
1776	3506	5,69°	umgedreht

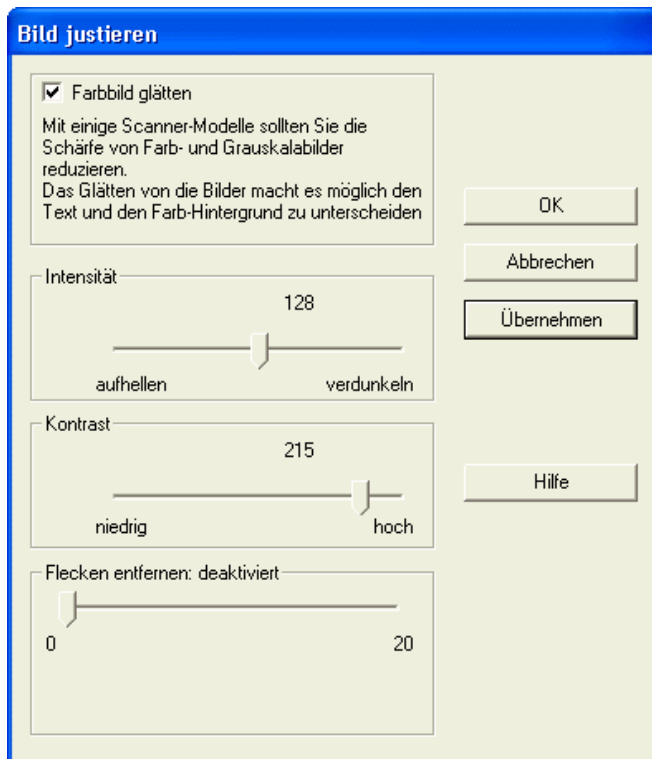
## **JUSTIEREN DER GESCANNTEN BILDER**

---

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“.



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.

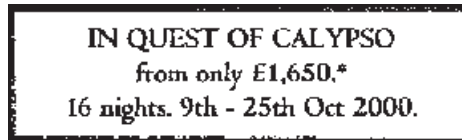
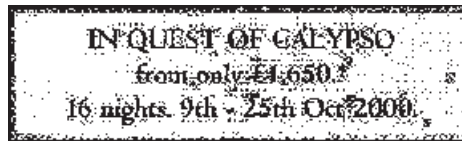
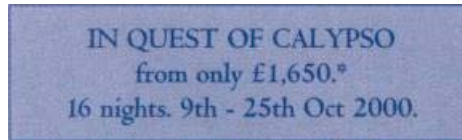


Die Option „Farbbild Glätten“ macht Graustufen- und Farbbilder homogener, indem relative Intensitätsunterschiede „abgeflacht“, „geglättet“ werden. Dies führt zu einem schärferen Kontrast zwischen dem Vordergrund - dem Text - und dem Hintergrund - Farbe, Verzerrungen usw.

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!

Files of type:

- PDF Dokumente öffnen mit Farben
- Digitalkamera
- Farbbilder glätten
- Verarbeiten wie 300 dpi

Kommen wir zur **Intensität**. Mit dieser Einstellung wird die allgemeine Intensität des Bildes bestimmt: jede Verdunkelung oder Aufhellung des Bildes wird für alle Pixel übernommen. Das Ziel dabei ist, dass Sie den Seitenhintergrund loswerden. Wir geben zwei Beispiele an. Im ersten Beispiel ist jede Zone des Bildes dunkel. Wir hellen das Bild daher auf, um den Seitenhintergrund zu eliminieren. Der Vordergrund - der Text - bleibt ausreichend dunkel, um von der



Binärkonvertierung erkannt zu werden. Beispiel 2: das Bild ist so hell, dass selbst der Text im Vordergrund nicht im binärkonvertierten Bild auftaucht! Wir dunkeln das Bild so ab, dass der Text erkennbar wird.

**Verenigde Staten,  
een antwoord te vi  
maar met name or**

*wyjścia każdego  
brawia, że nasze*



*wyjścia każdego  
brawia, że nasze*

**Verenigde Staten,  
een antwoord te vi  
maar met name or**

*wyjścia każdego  
brawia, że nasze*

Durch den **Kontrast** wird der örtliche Kontrast zwischen den dunkleren und den helleren Bereichen des Bildes festgelegt. (Der Text ist normalerweise dunkler als der Hintergrund; wenn Sie mit umgekehrtem Text arbeiten, gilt das Gegenteil.) Das Ziel besteht darin, die Zeichenformen deutlich vom (farbigen) Hintergrund abzuheben. Hier sehen Sie ein Beispiel, in dem wir den Kontrast erhöhen müssen, da die Standardeinstellungen zu gebrochenen Zeichen führen.

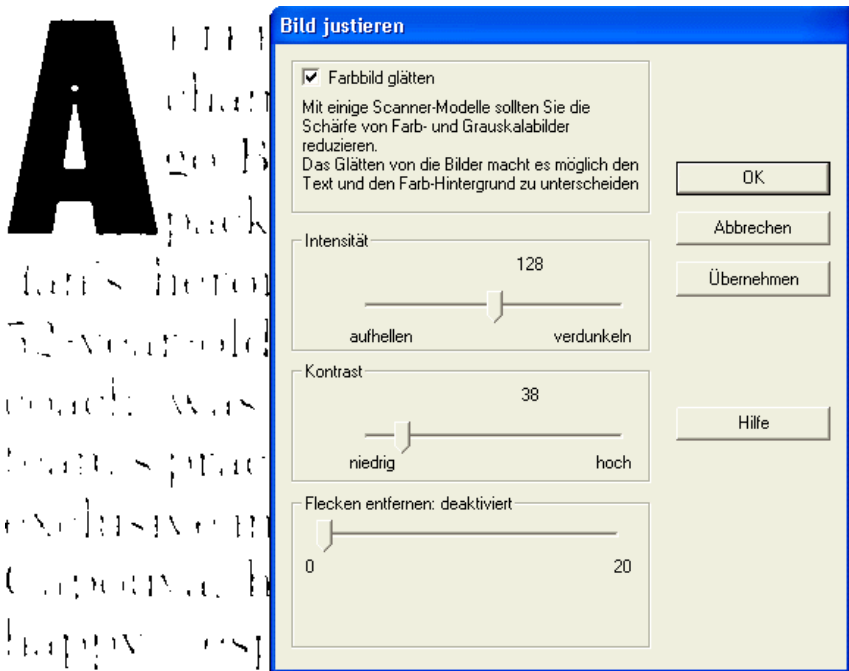
**A Look at International  
Planning the Future .....**

**A Look at International  
Planning the Future .....**

**A Look at International  
Planning the Future .....**

Bitte beachten Sie, dass keine Bildanpassung ausgeführt wird, bevor Sie auf „Übernehmen“ klicken! Wenn Sie auf „OK“ klicken, wird die Anpassung

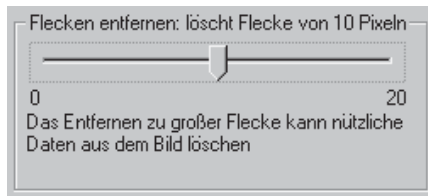
vorgenommen *und* das Fenster geschlossen. Hier ein Beispiel, in dem wir das Schwarzweißbild wesentlich aufgehellt haben - dabei stand die OCR-Präzision allerdings nicht im Vordergrund!



Diese Optionen gelten für Graustufen- und Farbbilder, die letztgenannte, „Flecken Entfernen“ (Despeckling) wird ausschließlich auf Schwarzweißbilder angewendet. Als „Flecken entfernen“ bezeichnet man die Entfernung von „Parasiten-Pixels“ aus Schwarzweißbildern.

# The Olympic Games The Olympic Games

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.

## **WENN DER OCR ASSISTENT FÜR SIE ARBEITET...**

---

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, benützen 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



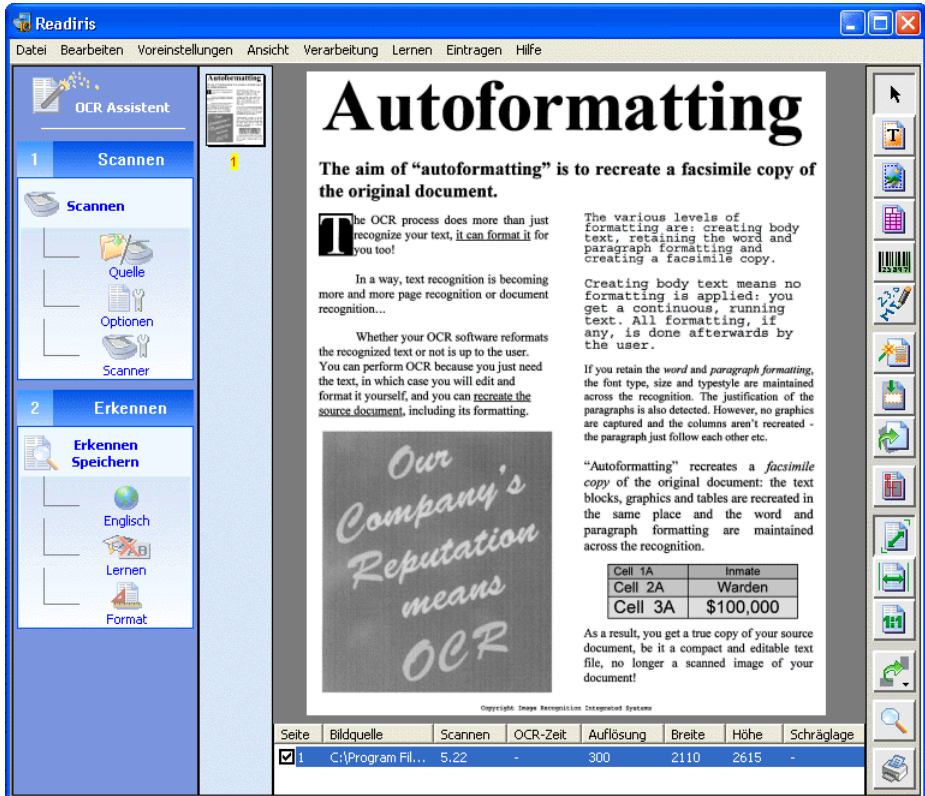
eine **Faxkopie** des gescannten Dokumentes erstellt. Die Wort-, Absatz-, und Seitenformatierung vom Originaldokument werden übernommen.

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!

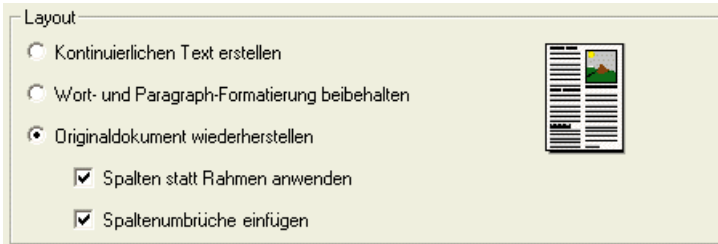
All das bedeutet, dass die Fenstersortierung nur *teilweise* arbeitet, wenn „Autoformatierung“ aktiviert wurde. Sie können Zonen löschen oder hinzufügen, aber eine neue Reihenfolge der Zonen wird einfach ignoriert!

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



Klicken Sie auf die Schaltfläche „Format“ auf der Haupttoolbar 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

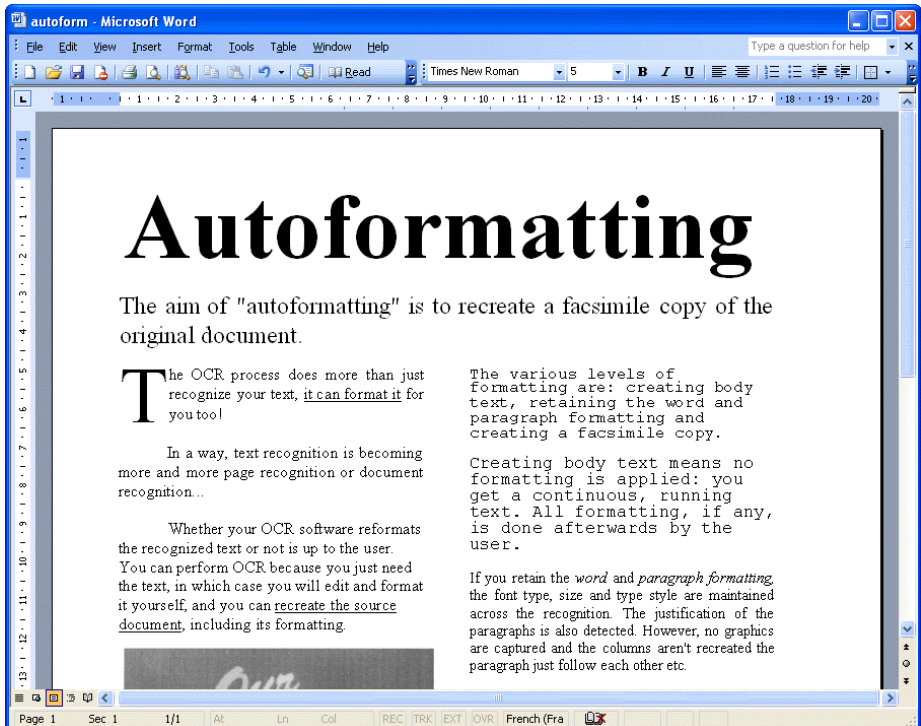
Wiederherstellen“ wird automatisch aktiviert, um den Zeilenumbruch zu aktivieren.)



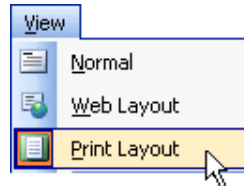
Ob die Layoutrekonstruktion verfügbar ist oder nicht, hängt vom gewählten Ausgabemodus ab. Einige „arme“ Formate, die „einfachen“ Text erzeugen, zum Beispiel Text (ANSI), MS-DOS Text (ASCII) usw., unterstützen *keine* fortgeschrittenen Formatierungs-codes und ermöglichen somit keine automatische Formatierung. 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.





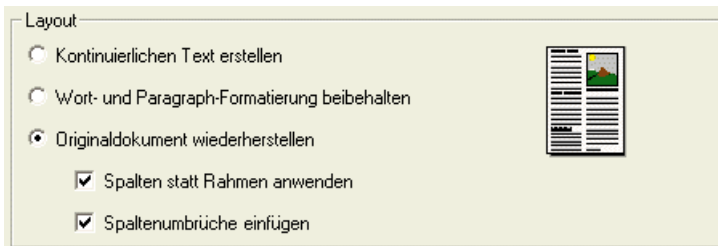
Um das Dokument in der richtigen Ansicht zu sehen, müssen Sie den „WYSIWIG“-Modus, meist „Seitenlayout“ genannt, aktivieren. Wenn Sie aber das erkannte Dokument direkt in Microsoft Word exportieren, wird die Seitenlayout- oder Seitenansicht-Ansicht automatisch aktiviert.



Also, Readiris kann Texte erkennen und formatieren. Bei OCR handelt es sich nicht mehr nur um Texterkennung, OCR wurde zur **Dokumenterkennung** weiterentwickelt!

## 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.

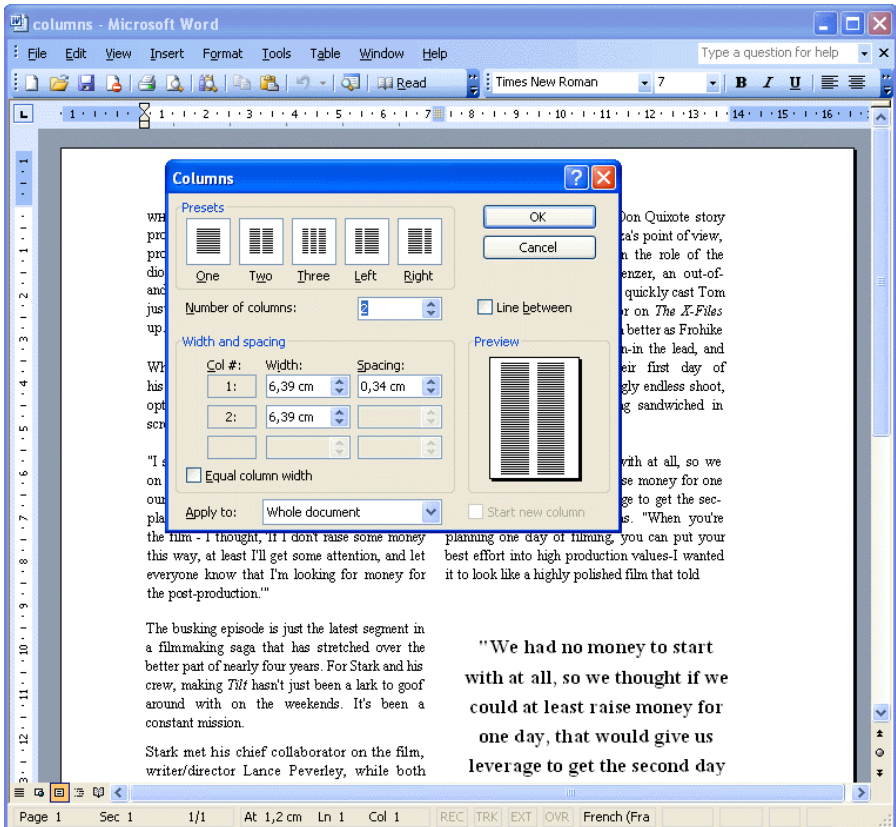


Rahmen („frames“) sind getrennte „Containers“, die zur Positionierung mehrerer Textblöcke, Grafiken und Tabellen auf einer Seite verwendet werden. In Spalten fließt der Text natürlich von einer Spalte zur nächsten. In Spalten gegliederte Texte sind wesentlich einfacher zu bearbeiten.

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.





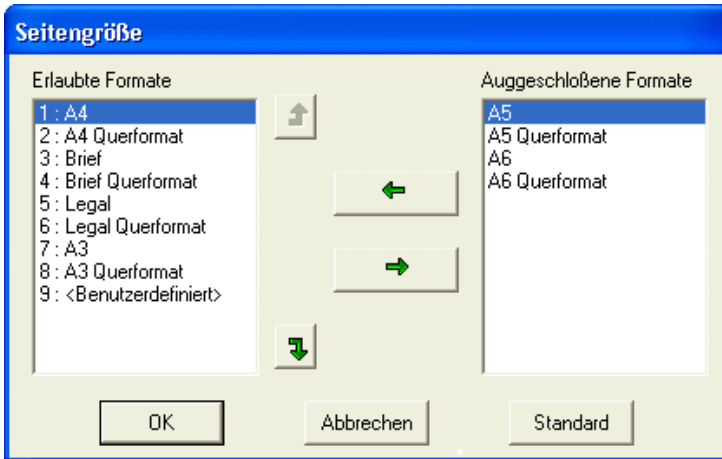
Die Option „Spaltenumbrüche Einfügen“ verfeinert die Wiederherstellung von Spalten: Sie legt fest, ob Sie „harte“ Spaltenumbrüche am Ende jeder Spalte einfügen möchten oder nicht. Mit Spaltenumbrüchen bleibt jeder Text, den Sie bearbeiten, hinzufügen oder entfernen, in der jeweiligen Spalte; kein Text 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.

Es gibt einen Gesichtspunkt, bei dem Sie sich möglicherweise dazu entschließen, das Originaldokument *nicht* wieder herzustellen: die Seitengröße des Ausgabedokuments. Was meinen wir damit? Schauen wir uns einige Beispiele an: Sie scannen zurzeit Letter-Seiten, erzeugen jedoch Ausgaben im A4-Format, da sich dieses Format besser für das Drucken eignet, Letter hingegen eine manuelle Papierzufuhr erforderlich macht. Oder Sie sind ein Rechtsanwalt und Sie scannen Dokumente im Letter-Format, die Sie im Legal-Format speichern.

Aus diesem Grund können Sie in Readiris bevorzugte Seitengrößen für das Ausgabedokument festlegen. Klicken Sie im Dialogfeld „Format“ auf die Schaltfläche „Seitengröße“.



Wählen Sie die zu verwendenden und die auszuschließenden Seitengrößen aus: die zu verwendenden Seitengrößen können zur Formatierung der erkannten Dokumente genutzt werden, die ausgeschlossenen Formate werden nicht genutzt. Sortieren Sie die zu verwendenden Seitengrößen: Readiris durchsucht die verschiedenen Seitengrößen in der vorgegebenen Reihenfolge und verwendet die erste Seitengröße, auf die das gescannte Dokument passt. Durch die Schaltfläche „Standard“ wird wieder die Standardeinstellung gültig. (Dabei werden die Einstellungen von Windows berücksichtigt!)

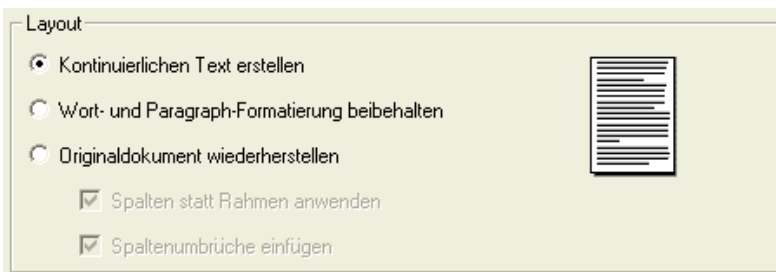
Sie müssen wissen, dass diese Option nicht für HTML-Dateien gilt - ein Textformat für das Internet, in dem keinerlei Seitenformate festgelegt sind! Diese Option gilt auch nicht für PDF-Dateien; für diese Dateien gelten die vom Ersteller vorgenommenen Einstellungen, um das ursprüngliche Dokument exakt zu reproduzieren.

## 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.

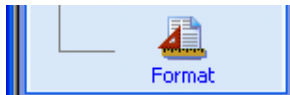
Die Option „Wort- und Paragraph-Formatierung Beibehalten“ ist ein Mittelding. Die **Wortformatierung** - Schriftart, Schriftgröße und Schriftauszeichnung - werden beibehalten. Das Gleiche gilt für die **Absatzformatierung** - der Randausgleich und die Einzüge werden ebenfalls beibehalten.

Verwechseln Sie diese Formatierungsoption nicht mit „kompletten“ Autoformatierung: hier werden die Absätze hintereinander gestellt, es werden keine Spalten rekreatiert 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-Speichern“. 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 ungeduldigen 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 wurde die Grafik in das erkannte Dokument einbezogen, ob dies der Fall ist, hängt von der Formatierungsoption „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!

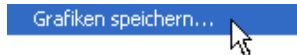


Sie können mit Readiris jedoch auch Grafiken speichern, ohne die Texterkennung durchzuführen. Da Readiris Schwarzweiß-, Graustufen- und Farbbilder erzeugt, können Sie Schwarzweiß-Grafiken und Fotos erfassen.

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“.



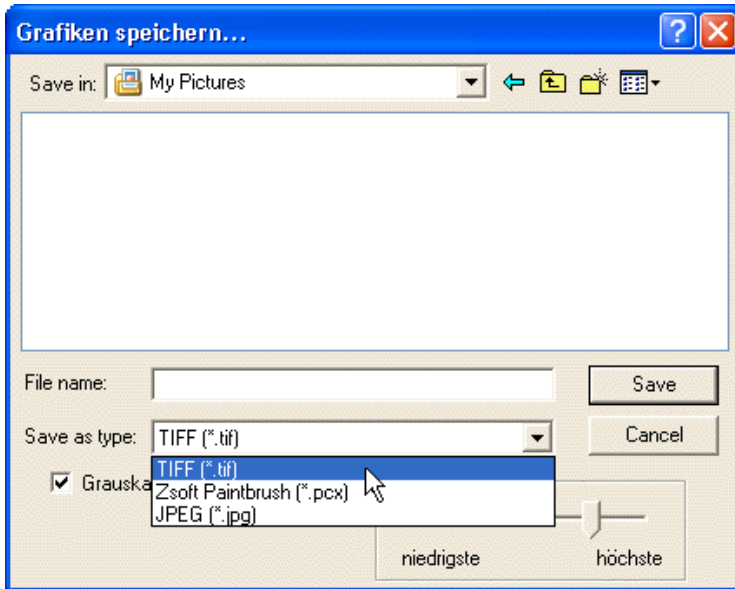
Selektieren Sie danach den Befehl „Grafiken Speichern“ im Menü „Datei“.



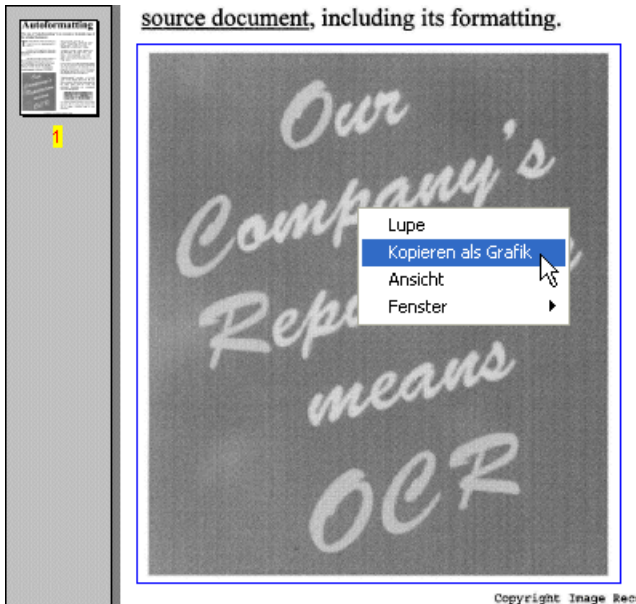
Legen Sie fest, welches Grafikdateiformat Sie benutzen möchten. Wählen Sie ein Format, das von Ihrer Fotoretuschiersoftware unterstützt wird. Verfügbar sind die Formate JPEG, TIFF und ZSoft Paintbrush (\*.pcx). Readiris Corporate unterstützt außerdem das kompakte Format JPEG 2000 (\*.j2c)!

Aktivieren Sie diese Option „Grauskala/Farben“, um die Grafiken als Farb- oder Graustufengrafik zu speichern. Wenn Sie Schwarzweißgrafiken im TIFF-Format speichern, erfolgt eine Komprimierung Gruppe 4. Wenn Sie Graustufen- und Farbgrafiken im TIFF-Format speichern, erfolgt eine JPEG-Komprimierung und Sie können die JPEG-Qualität festlegen.





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!



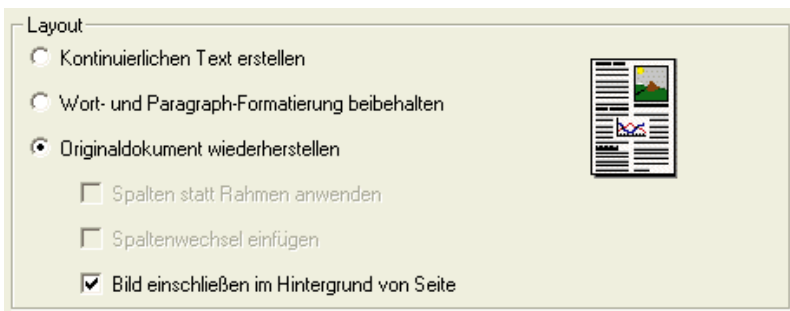
## SPEICHERN FARBIGER HINTERGRÜNDE

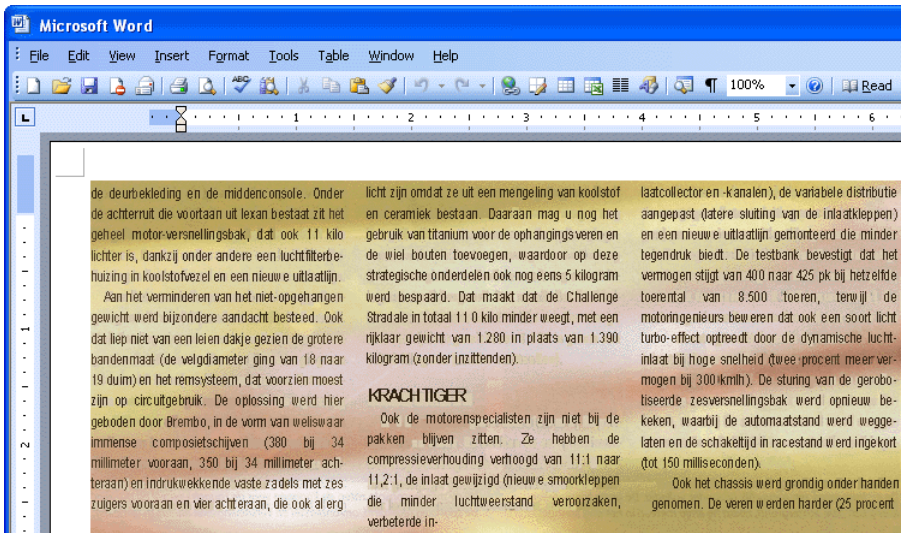
---

Mit Readiris Corporate können Sie auch den Hintergrund in den erkannten Dokumenten speichern! Die Option „Bild Einschließen im Hintergrund von Seite“ platziert das gescannte Bild als Seitenhintergrund unter den erkannten Text; diese Funktion wird für Ausgaben in HTML, RTF, Word (\*.doc) und WordML (\*.xml) unterstützt. (Wenn Sie PDF-Dateien erzeugen, können Sie diesen Effekt durch Auswählen des Formats „PDF Text-Bild“ erreichen.)

Es muss nicht extra gesagt werden, dass durch diese Option die Dateien der erkannten Dokumente wesentlich größer werden: diese Option fügt das gescannte Bild mit allen Details in den Hintergrund der Ausgabe des Dokuments ein. Mit

*allen* Details? Nicht wirklich: wie auch bei Ausgaben in PDF „Text-Bild“ werden die Pixel des erkannten Textes gelöscht, um ein lesbares Dokument zu bekommen. Die Darstellung von erkanntem Text in Schwarz über Bitmaps schwarzer Zeichen würde zu einem Text mit starker Schattierung führen... (Das Beispielbild BACKGROUND.JPG verdeutlicht, wie das funktioniert.)



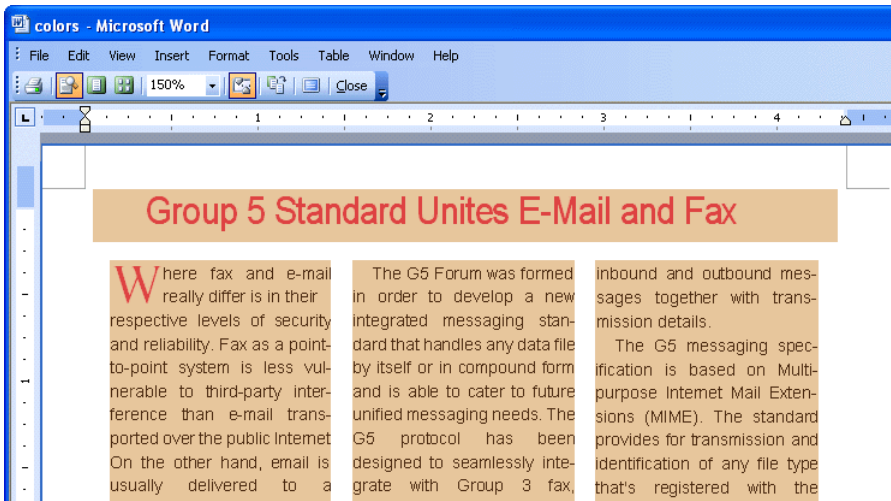


Readiris bietet auch eine weniger drastische und kompaktere Alternative: durch die Option „Farben von Hintergrund Beibehalten“ werden die Volltonfarben der Seite während der Erkennung beibehalten. (Diese Option bedeutet auch, dass Sie die Farben des Textes beibehalten.)



Das Ergebnis ist eine einheitliche Hintergrundfarbe - falls eine solche im Originaldokument enthalten ist - je Absatz. Auch dies gilt für Dokumente in HTML, Word und WordML. Die Details zum Beispiel einer Fotografie im Hintergrund

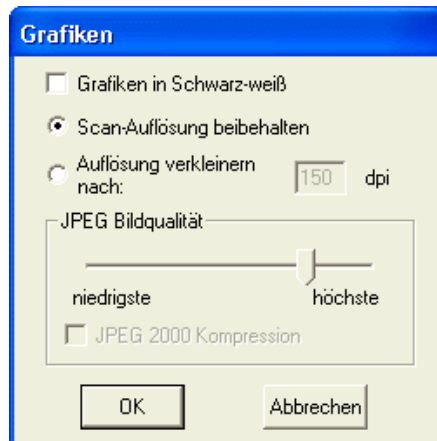
über die ganze Seite werden hier *nicht* beibehalten, die Vollfarbe eines Textrahmens jedoch schon. (Führen Sie probeweise eine Erkennung des Beispielbildes COLORS.JPG aus...)



## GRAFIKEN DURCH UND DURCH

Readiris Corporate bietet weitere fortgeschrittene Optionen für Grafiken. Sie finden diese Optionen unter den Knopf „Fortgeschritten“ des Befehles „Textformats“.

Diese Optionen gelten für *alle* Grafiken - die Grafikbereiche in erkannten Dokumenten und das Seitenbild, mit dem Sie in einer „Bild-Text“ Adobe Acrobat PDF Datei den Text überlagern!



Legen Sie den Farbmodus fest: speichern Sie Ihre Grafiken als Farb-/Graustufen- oder als Schwarzweißbilder. Wählen Sie die Auflösung der Bilder in den erkannten Dokumenten: behalten Sie die Scanauflösung oder reduzieren Sie sie. (Sie können die Auflösung nicht auf diese Weise *erhöhen!*) Wenn die erkannten Dokumente als HTML-Dateien auf einer Website veröffentlicht werden, möchten Sie die Auflösung sicherlich auf die Bildschirmauflösung verringern. Daher sind die Grafiken bei HTML-Seiten in der Standardeinstellung auf 72 dpi reduziert. (Sie können von Hand einen höheren Wert eingeben.) Und schließlich können Sie die JPEG Qualität wählen. (JPEG Bilder dienen zum Speichern von Farb- und Graustufenbildern in PDF-Dokumenten, Word- und RTF-Dokumenten usw.)

(Die Option „JPEG 2000 Kompression“ wurde bereits erläutert: diese Option bewirkt eine Komprimierung JPEG 2000 aller Grafiken und Bilder in PDF-Dokumenten.)

Mit diesen Optionen können Sie die **Dateigröße** der Ausgabedatei maßgeblich beeinflussen! Ein Beispiel: wenn Sie „Bild-Text“ PDF-Dateien mit Zweitongrafiken erzeugen, speichern Sie die Bilder als Gruppe 4 komprimierte TIFF-Dateien.

Speichern Sie dieselben Scans als Farbbilder, so werden die JPEDateien standardmäßig in hohe Qualität (0,8) gespeichert.

## **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 aufgefordert, die gesamte Seite als Datei im Format PDF, JPEG, JPEG 2000 (\*.j2c), TIFF oder ZSoft Paintbrush (\*.pcx) zu speichern.

A screenshot of a menu item "Ganze Seite als Bild speichern..." with a mouse cursor pointing at it. The text is white on a dark blue background.

Wählen Sie den Befehl „Alle Seiten als Bild Speichern“, um ein mehrseitiges Dokument zu speichern. Hier stehen zwei Dateiformate zur Verfügung: PDF und mehrseitiges 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 die Unterstützung für Farbbilder, Graustufenbilder und Schwarzweißbilder auf derselben Grundlage beruht, können Sie in Readiris Adobe Acrobat PDF-Dokumente, DCX-Faxnachrichten (eine Mehrseitenversion des Paintbrush-Formats), DjVu-Bilder (\*.djv, \*.djvu), JPEG-Bilder, JPEG 2000-Bilder (\*.j2c, \*.jp2), PNG-Bilder, TIFF-Bilder (unkomprimiert, LZW, PackBits, Gruppe 3, Gruppe 4 und JPEG-komprimiert), mehrseitige TIFF-Bilder, Windows Bitmaps (\*.bmp) und ZSoft Paintbrush-Bilder (\*.pcx) öffnen.

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.

## **DOKUMENTSTAPEL ERKENNEN**

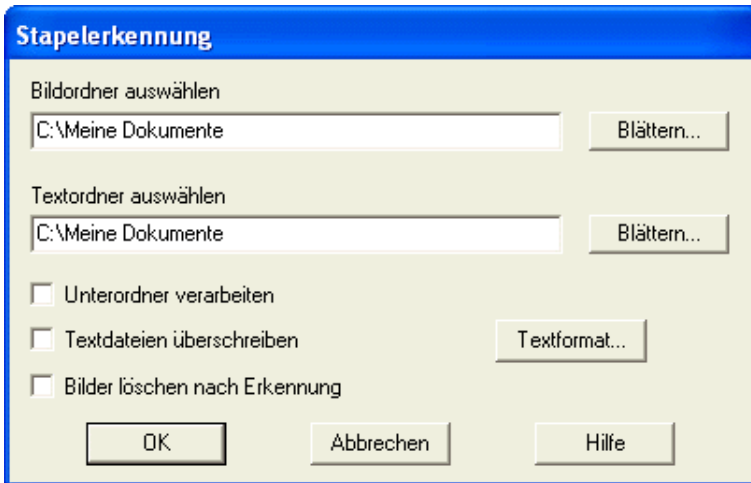
---

Readiris Corporate ist wesentlich leistungsfähiger bei der Erkennung vorgescannter Bilder: Sie können ganze Dokumentstapel automatisch erkennen und einen „überwachten“ Ordner festlegen. Werfen wir einen genaueren Blick auf diese Möglichkeiten!

**Stapelerkennung** führt die Erkennung aller vorgescannten Bilder in einem bestimmten Ordner durch. Sie können beispielsweise tagsüber scannen und diese Bilder nachts einlesen... Erfassen Sie einfach alle zu erkennenden Dokumente; wenn Sie damit fertig sind, starten Sie die Stapelerkennung und lassen Sie Ihren PC weiterarbeiten. Wenn Sie am nächsten Tag zurückkommen, sind alle Dokumente erkannt!

Die erkannten Dokumente erhalten dieselben Dateinamen wie die Bilddateien. Die Dateierweiterung hängt logischerweise vom gewählten Ausgabeformat ab: Bilddatei 001.TIF wird zum Beispiel zu 001.DOC konvertiert, wenn Word das gewählte Ausgabeformat ist.





Wählen Sie den Bild- und den Textordner. Der Textordner kann ein anderer Ordner als der Bildordner sein, muss aber nicht! Wenn der Bildordner mit dem Ausgabeordner übereinstimmt, finden Sie die Textdokumente zusammen mit den Scans (es sei denn, Sie haben die Option „Bilder Löschen nach Erkennung“ aktiviert)!

Die Option „Unterordner Verarbeiten“ legt fest, ob die Unterordner des Bildordners ebenfalls verarbeitet werden. Dies ist sinnvoll, wenn die vorgescannten Dokumente in bestimmten Unterordnern abgelegt wurden. In einem Ordner namens „25-2-2005“, der das Datum angibt, können sich beispielsweise die Unterordner „1“, „2“ usw. mit den eigentlichen Dokumenten befinden.

Wenn sie aktiviert ist, löst diese Option die Verarbeitung aller Unterordner aus; Sie können die Erkennung nicht auf einen oder mehrere spezifische Unterordner beschränken. Wenn der Textordner ein anderer Ordner als der Bildordner ist, werden die Unterordner ebenfalls im Ausgabeordner angelegt („gespiegelt“)!



Klicken Sie auf „OK“, um den Vorgang zu starten. Es werden keine weiteren Schritte benötigt, um die Dokumente umzuwandeln: die Erkennung erfolgt vollständig automatisch. (Interaktives Lernen gilt hier nicht.)

Außerdem sind Sie auf die „externen“ Textformate beschränkt. Die OCR-Ergebnisse werden auf Festplatte gespeichert. Die eingelesenen Ergebnisse direkt an eine Zielanwendung zu senden, sie per E-Mail zu versenden oder sie nach der Erkennung automatisch zu öffnen, macht einfach keinen Sinn, wenn die Erkennung an einem unüberwachten PC erfolgt. Aber alle anderen OCR-Optionen - Sprache, Typ der Schriftart usw. des aktuellen Dokuments - gelten weiterhin: Sie sollten sich vergewissern, dass die richtigen Optionen aktiviert sind, bevor Sie die Erkennung starten!

Die Stapelerkennung bietet einige fortschrittliche Merkmale. Die Option „Textdateien Überschreiben“ legt fest, ob der OCR-Vorgang zuvor vorhandene Erkennungsergebnisse überschreiben darf. Deaktivieren Sie diese Option, wenn Sie neue Bilddateien zu einem Ordner hinzufügen, der Bilder enthält, die bereits zuvor verarbeitet worden sind. (Andernfalls würden Ihre Bilddateien erneut erkannt, wenn Sie den Ordner ein zweites Mal verarbeiten!)

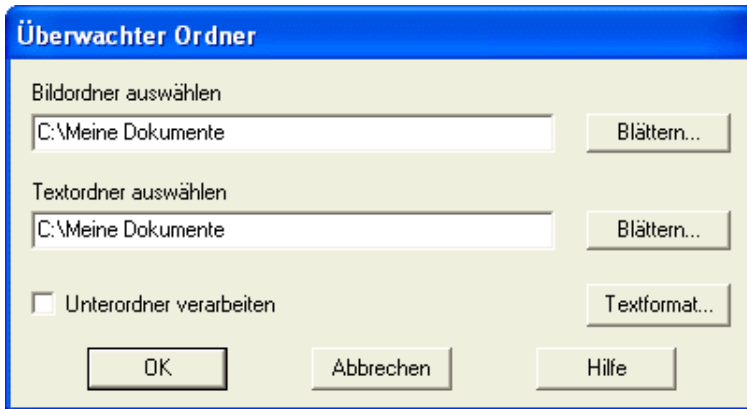
Die Option „Bilder Löschen nach Erkennung“ legt fest, ob die Bilddateien nach der Erkennung gelöscht werden. Mit anderen Worten: Readiris kann den Bildordner für Sie „aufräumen“!

Aktivieren Sie diese Option, wenn Sie ausschließlich die erkannten Dokumente speichern und archivieren möchten und die „temporären“ Bilddateien verwerfen möchten. Deaktivieren Sie die Option, wenn Sie sowohl die Scans als auch die erkannten Dokumente speichern möchten. (Es sei daran erinnert, dass Readiris PDF Dokumente vom Typ „Bild-Text“ erstellt: dieses Ausgabeformat speichert das gescannte Bild *und* das erkannte Dokument in einer einzigen Datei!)

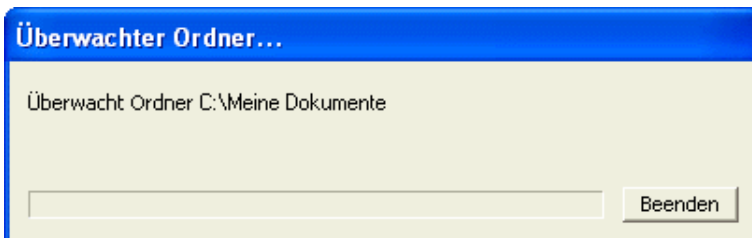
## **EINEN „ÜBERWACHTEN“ ORDNER FESTLEGEN**

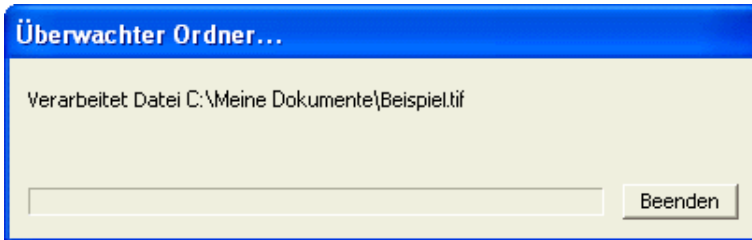
Die Nutzung eines „überwachten“ Ordners entspricht weitgehend der Stapelerkennung. Der wichtigste Unterschied ist, dass der Benutzer diesmal nicht die Erkennung starten muss. Ganz im Gegenteil: Readiris führt die Erkennung

systematisch für jede Bilddatei durch, die in diesem speziellen Ordner abgelegt wird. Sie können die OCR-Software tagelang laufen lassen... Erfassen Sie neue Dokumente, und schon werden sie erkannt.



Klicken Sie auf „OK“, um die „Überwachung“ des beobachteten Ordners zu starten. Readiris verarbeitet die Bilder, sobald sie erstellt und in dem beobachteten Ordner abgelegt werden. Klicken Sie auf „Beenden“, um die Überwachung zu beenden.



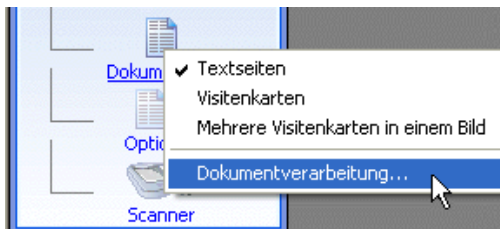


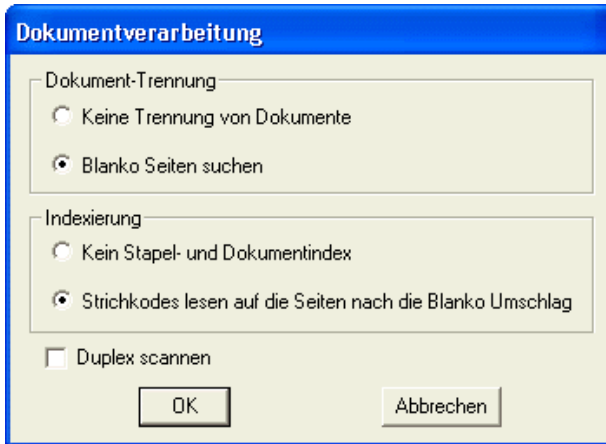
Wie bei der Stapelerkennung verarbeitet Readiris die Bilder aller unterstützten Dateiformate. Sie können die OCR nicht auf ein bestimmtes Grafikformat begrenzen: Readiris verarbeitet die Bilder *aller* unterstützten Dateiformate. (Dateien mit anderem Format werden einfach ignoriert.)

## ORGANISIEREN VON STAPELN

---

Das bedeutet nicht, dass die vorgescannten Seiten unbedingt blind, einfach nacheinander verarbeitet werden. Im Gegenteil, Readiris verfügt über ausgeklügelte Routinen, die ein intelligentes Verarbeiten der gescannten Stapel ermöglichen. Rufen Sie über die Schaltfläche „Dokumenttyp“ (oder im Menü „Einstellungen“) den Befehl „Dokumentverarbeitung“ auf, um sich mit der Arbeitsweise dieser Funktionen zu befassen.



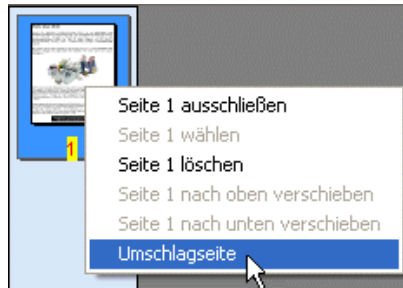


Zwischen den Seiten können leere Umschlagseiten eingefügt werden, um **Dokumente** voneinander zu **trennen**. Eine "leere" Seite ist eine Seite mit fast keinen schwarzen Pixel - schwarze Ränder ausgenommen. Auf der Seiten-Toolbar werden die Umschlagseiten hervorgehoben, im Dokumentpaneel werden sie markiert.



Seite	Umschlagseite	Strichcode	Bildquelle	Scannen
<input checked="" type="checkbox"/> 1	ja		C:\Program File...	0.54
<input checked="" type="checkbox"/> 2	-	978047135652	C:\Program File...	1.29

Sie können eine Folge gescannter Bilder auch manuell trennen: wählen Sie eine beliebige Seite - leer oder nicht - auf der Seiten-Toolbar aus, klicken Sie mit der rechten Maustaste und wählen Sie im „Kontext“-Menü den Befehl „Umschlagseite“ aus. (Beachten Sie, dass der Inhalt einer Umschlagseite verloren geht: die Umschlagseiten dienen als Trenner, werden jedoch *nicht* gelesen!)

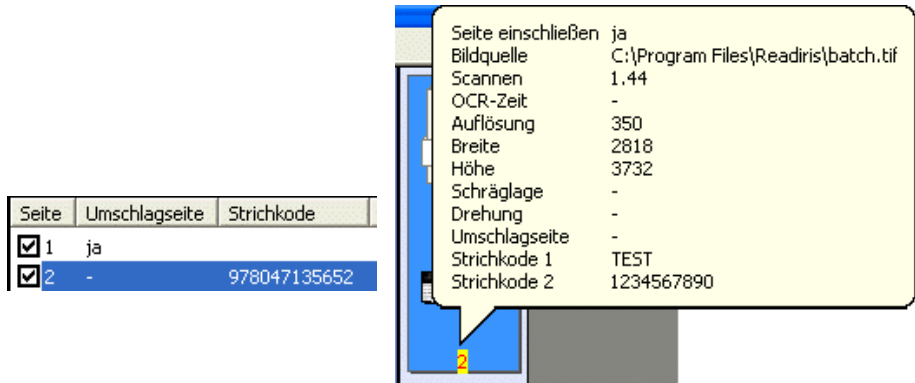


Das Aufteilen der verschiedenen Dokumente in einer Folge gescannter Bilder ist der erste Schritt. Im nächsten Schritt erfolgt die **Indexierung** der Dokumente. Zu diesem Zweck erfolgt das Lesen von Strichcodes. Alle Strichcodes, die auf einer Seite nach einer Umschlagseite auftreten, werden automatisch gelesen. (Führen Sie probierhalber eine Erkennung des Beispielbildes BATCH.TIF aus...)



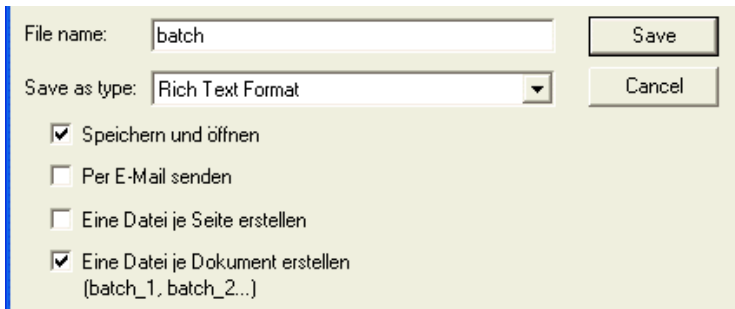
Die Option „Duplex Scannen“ bedeutet, dass Sie die Vorderseite und die Rückseite von Dokumenten scannen. Die Rückseiten werden beim Entdecken von Strichcodes auf der nächsten Seite offensichtlich ignoriert: die nächste Seite wird zum nächsten *Titebild*, nicht zum nächsten Bild!

Die Schnittstelle von Readiris ist auf andere Weise nützlich, wenn Sie mit der Verarbeitung der Dokumentstapel beginnen: im Dokumentpaneel und in den Tooltipp der Seitenminiaturen wird der Wert der erkannten Strichcodes angezeigt!



Die Ergebnisse des Auslesens von Strichcodes werden im XML-Index gespeichert, nicht im erkannten Dokument!

Wenn Sie Stapel manuell verarbeiten, speichert die Option „Eine Datei je Dokument Erstellen“ im Dialogfeld „Speichern“ separate Dokumente, nicht vollständige Stapel. (Dies ist nur möglich, wenn Sie externe Dateien anlegen, nicht wenn Sie die Ausgabe direkt an eine Zielanwendung weiterleiten.)





## **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.

Seite	Bildquelle	Scannen	OCR-Zeit	Auflösung	Breite	Höhe	Schräglage	Drehung
<input checked="" type="checkbox"/> 1	C:\Program ...	4.61	-	300	2092	2517	?	-

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.)



table - Microsoft Word

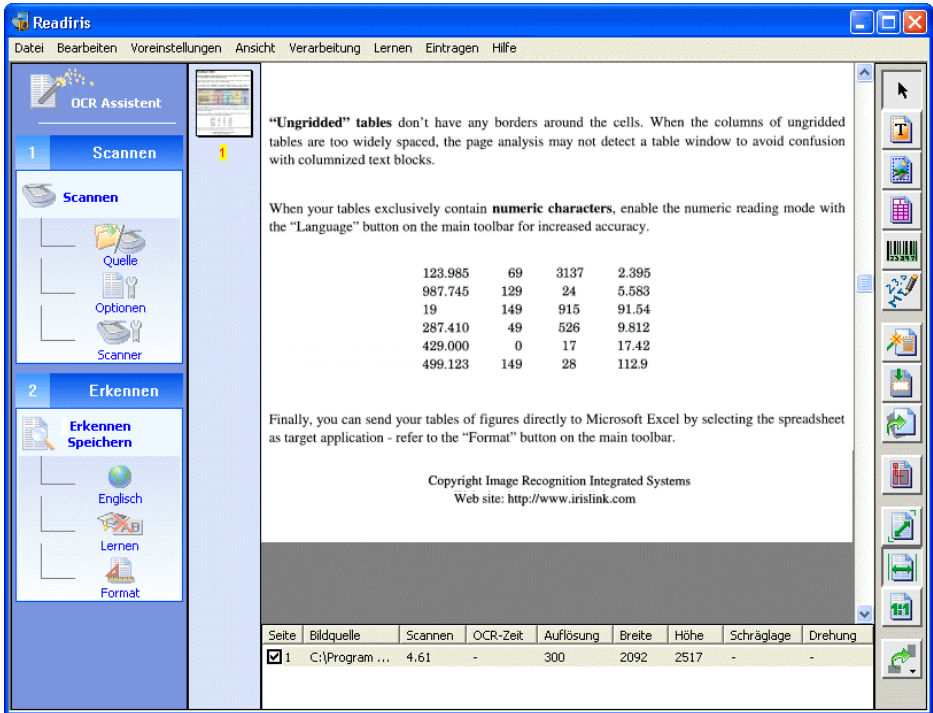
File Edit View Insert Format Tools Table Window Help

Type a question for help

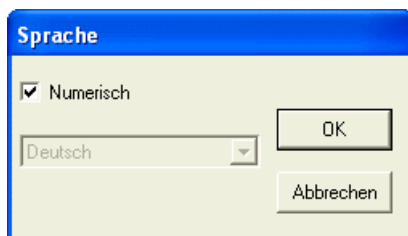
Times New Roman 12 B I U

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!

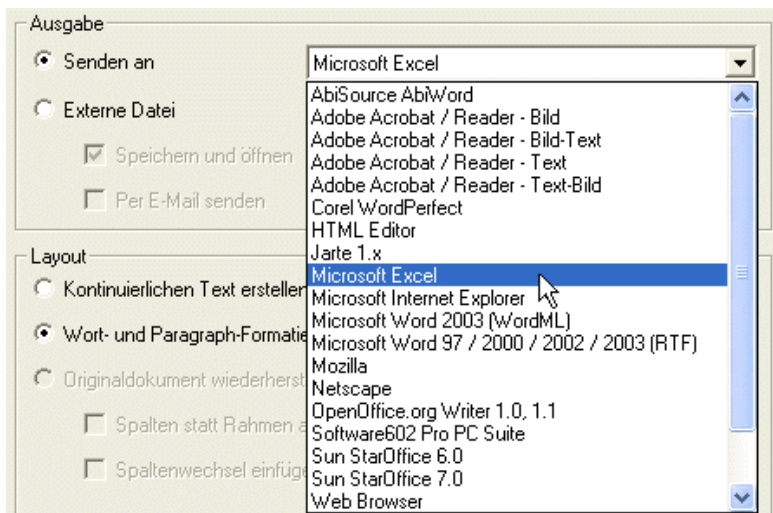


Um eine optimale Scanqualität zu bekommen, sollten Sie die Erkennung auf **numerischen Symbole** mit der Schaltfläche „Sprache“ begrenzen. (Der numerische Modus ist nicht ausschließlich numerisch, er enthält die Zeichen „0“ bis „9“, „+“, „\*“, „/“, „%“, „,“ (Komma), „.“ (Punkt), „(“ „)“, „-“, „=“, „\$“,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

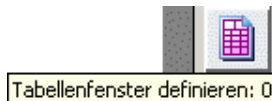
Diesmal werden wir das OCR-Ergebnis direkt in das Spreadsheet Microsoft Excel senden. Wir wählen Excel als gewünschte Anwendung mit der Schaltfläche „Format“.



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; der Tooltipp des Befehls gibt wie üblich die Anzahl Tabellenfenster an.



## HANDSCHRIFTLICHEN TEXT ERKENNEN

Wir haben gescannte Dokumente, Tabellen, Faxnachrichten, mit einer Digitalkamera aufgenommene Bilder erkannt, wir haben Grafiken gespeichert und wir haben PDF-Dokumente konvertiert. Readiris verfügt nun über eine weitere Lesefunktion: die Erkennung von handschriftlichen Texten.



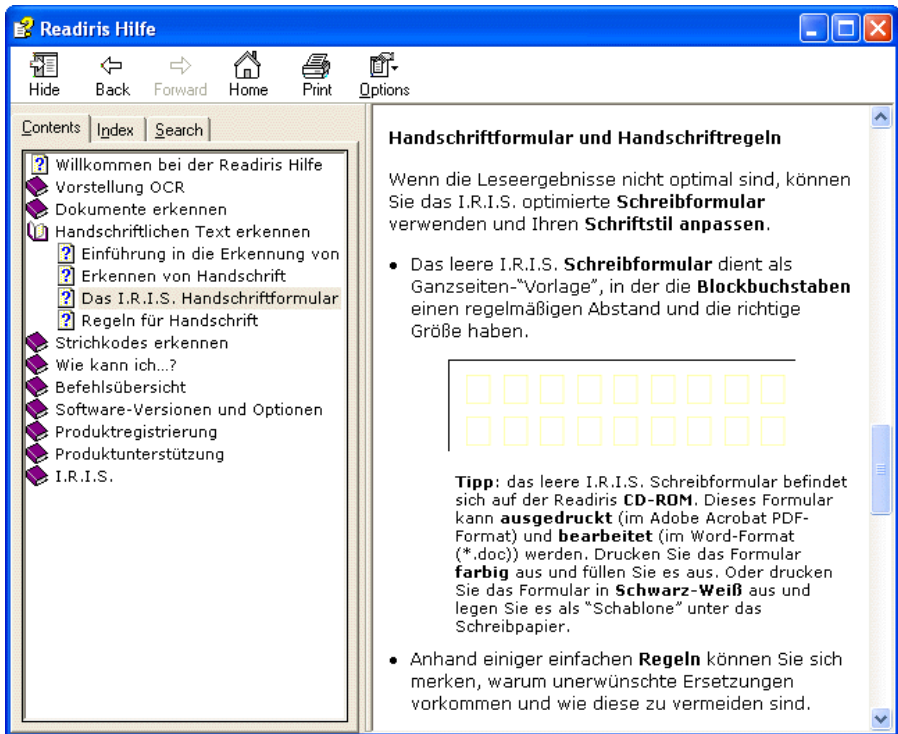
Genauer sollten wir handgedruckte Texte sagen, nicht handschriftliche Texte! *Handschrift* benennt einen kontinuierlichen, „kursiven“ von Hand geschriebenen Text. Die Zeichen innerhalb eines Wortes oder einer Zeichenfolge sind miteinander verbunden, es kann nicht gesagt werden, wo das eine Zeichen aufhört und das andere Zeichen anfängt. Beim *Handschriftdrucken* sind die „Blockbuchstaben“ voneinander getrennt und die Erkennungssoftware kann die einzelnen Zeichen besser isolieren.

*Italienscheie*  
*001-0405369-28*

Zum Erkennen der von Hand geschriebenen Zeichen wird eine hochspezialisierte Software - „ICR“-Software („Intelligent Character Recognition“) - benötigt. Die leistungsstarke ICR-Technologie von I.R.I.S. basiert auf mehr als einer Millionen Handschriftproben! Readiris unterstützt alle natürlichen Handschriftstile - Europäisch und Amerikanisch. Es ist kein bestimmter Schriftstil notwendig.

Die Erkennung von Handschriften ist auf die Ziffern (0-9), die Großbuchstaben (A-Z) und die Satzzeichen „,“ (Komma) „.“ (Punkt) und „-“ (Bindestrich) beschränkt.

Wenn die Leseergebnisse nicht optimal sind, können Sie das I.R.I.S. optimierte Schreibformular verwenden und Ihren Schriftstil anpassen. Die Schreibregeln finden Sie in der Online-Hilfe von Readiris. Anhand einiger einfachen Tipps können Sie sich merken, warum unerwünschte Ersetzungen vorkommen und wie diese zu vermeiden sind. Das leere I.R.I.S. Schreibformular dient als Ganzseiten-„Vorlage“, in der die Blockbuchstaben einen regelmäßigen Abstand und die richtige Größe haben. Das leere Formular befindet sich auf der Readiris CD-ROM und kann ausgedruckt und bearbeitet werden.



Jetzt wissen wir also, wie wir in einem Meeting maschinenlesbare handschriftliche Notizen anfertigen können. Wie können wir nun im Anschluss an das Meeting diese Notizen mit Readiris erkennen? Zeichnen Sie ein Handschriftfenster um den handgeschriebenen Text und führen Sie die Erkennung aus. (Sie können das anhand des Beispielbildes HANDPRINTING.TIF ausprobieren!)

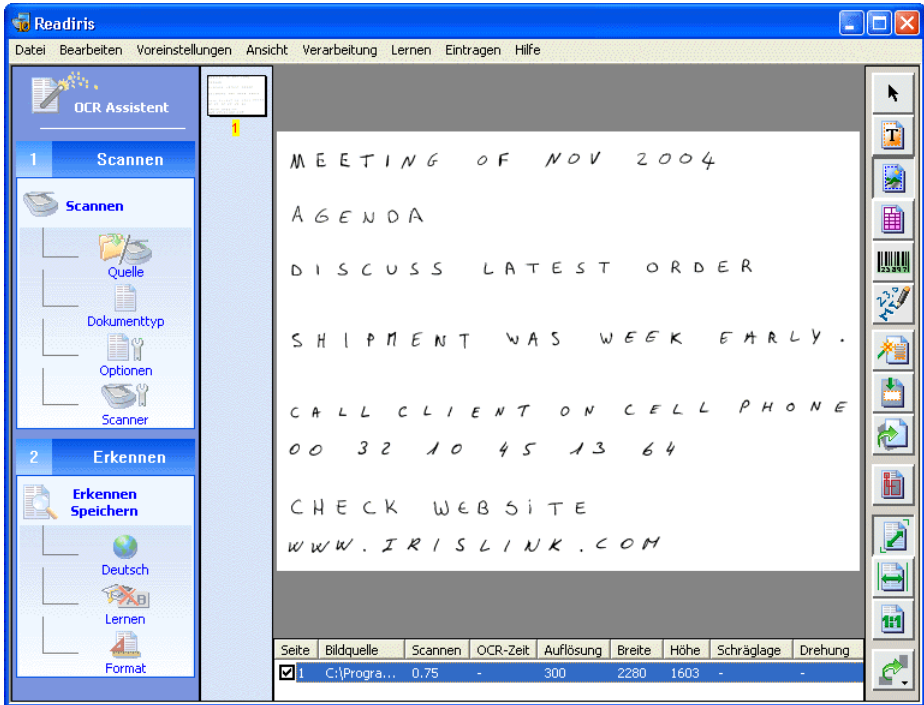


Handschriftenfenster zeichnen: 0



Die Dokumenteigenschaften - Sprache, Schriftart und Zeichenabstand - gelten nicht für Handschriften. Die Funktion ist auf einen einfachen englischen - oder besser gesagt „lateinischen“? - Zeichensatz in (Großbuchstaben)-Blockschrift beschränkt. Auch die Lernfunktion steht nicht zur Verfügung: die Lernfunktion macht keinen Sinn, da jede Person einen eigenen Handschriftstil hat. (Wie gesagt, diese leistungsstarke ICR-Technologie basiert auf mehr als einer Millionen Handschriftproben...)





## LESEN VON STRICHEN UND LÜCKEN

Und Readiris kann auch Strichcodes lesen...! Strichcodes in gescannten Bildern können gelesen und als erkannte Daten in die Ausgabedokumente eingefügt werden.

Strichcodes bestehen aus parallelen Strichen und Freiräumen zwischen diesen Strichen. Fest definierte Kombinationen aus Strichen und Freiräumen stellen



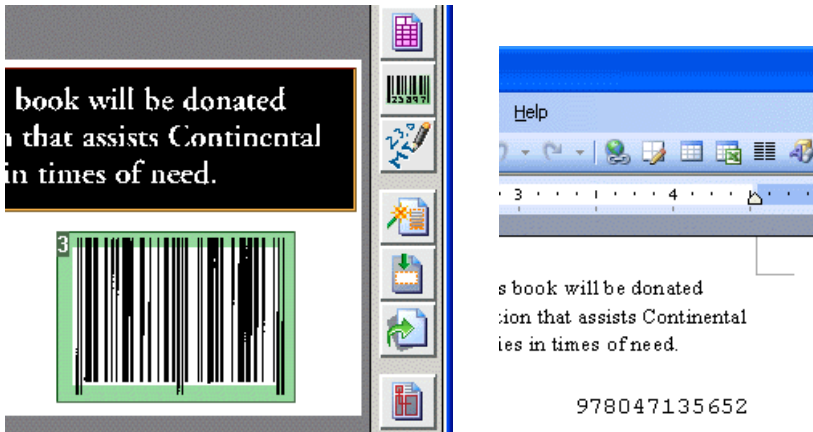
bestimmte Zeichen dar. Es gibt verschiedene Strichkodestandards oder „Symbolsätze“. Es werden alle gängigen Strichcodesymbolsätze unterstützt.



Lediglich die mit Laserdruckern oder Tintenstrahldruckern gedruckten Strichcodes erreichen eine ausreichende Qualität. Schließen Sie mit Nadeldruckern gedruckte Strichcodes aus: diese Codes erreichen keinen ausreichenden Kontrast und die Auflösung ist meist auf 60 dpi eingeschränkt. Readiris erkennt Strichcodes mit gutem Kontrast am besten; schwarze Striche auf weißem Hintergrund erzielen die besten Ergebnisse. Die meisten Strichcodes erfordern einen „störfreien“ Bereich um den eigentlichen Strichcode herum. Teilweise Ergebnisse können von Strichcodes nicht wiedergegeben werden; ein fehlendes Start- oder Stoppzeichen oder ein falsches Prüfsummenzeichen führt immer zu einem fehlerhaften Auslesen mit dem Ergebnis „null“!



Zeichnen Sie um jeden einzelnen Strichcode ein Strichcodefenster - die Seitenanalyse erkennt diese Strichcodes nicht - und führen Sie die Erkennung aus. Die Strichcodes werden gelesen und in den Ausgabertext eingefügt. Sie können auch mit der rechten Maustaste auf eine Strichcodezone klicken und im „Kontext“-Menü den Befehl „Kopieren als Daten“ ausführen; der Strichcode wird gelesen und in die Zwischenablage kopiert... (Die Prüfsummenzeichen einiger Strichkodestandards werden überprüft, jedoch aus dem Ausleseergebnis gestrichen.) Das Beispielbild BARCODE.TIF verdeutlicht, wie das funktioniert.



## VISITENKARTEN ERKENNEN

Wir haben gescannte Dokumente, Tabellen, Faxnachrichten, Digitalkamera-Schnappschüsse erkannt, wir haben Grafiken gespeichert und PDF-Dokumente konvertiert. Readiris Corporate weist jedoch noch eine zusätzliche Lesefunktion auf: die Erkennung von Visitenkarten („BCR“ - „Business Card Reading“)!

Mit Readiris können Sie Visitenkarten scannen, erkennen und in eine **Adressdatenbank** konvertieren. Hierbei ermöglicht Ihnen die Texterkennung mittels OCR, Visitenkarten zu digitalisieren, ohne sie zeitaufwändig neu eintippen zu müssen. Denken Sie an ihren letzten Messebesuch, von dem Sie einen ganzen Stapel Visitenkarten zurückbrachten, und Ihre Sekretärin brauchte mehrere Tage, um sie in den Computer einzugeben!

Die Kartendaten werden automatisch aus dem Bild herausgelesen, und die erkannten Daten werden spezifischen Datenbankfeldern zugeordnet. Readiris macht umfassenden Gebrauch von einer Wissensdatenbank, wodurch es die erforderliche Intelligenz erwirbt, um den Vor- und Nachnamen, eine Stadt und



ihr Bundesland, eine Telefon- und eine Telefaxnummer usw. zu unterscheiden. Jedes Land hat einen anderen Stil bei seinen Visitenkarten; die Amerikaner ordnen eine Anschrift anders an als die Franzosen usw.

Das funktioniert für bis zu **28 Länder!** Die größeren Länder aller Kontinente werden unterstützt: nord- und südamerikanische Visitenkarten sowie Visitenkarten aus den europäischen Staaten, einschließlich der osteuropäischen Nationen. (Optional können Sie ebenfalls asiatische Visitenkarten aus China, Japan, Korea und Taiwan einlesen.)

Die daraus resultierenden Daten stehen zum **Export** zur Verfügung. Sie können Ihre Kontaktinformationen in einer strukturierten Textdatei abspeichern, zum Beispiel in das durch Kommata begrenzten Format oder das vCard Format, um sie später in eine beliebige Adressen-Datenbank zu importieren, zum Beispiel Microsoft Access.

Alternativ können Sie die Kontaktinformationen direkt an Ihr Adressenverwaltungsprogramm Microsoft Outlook (Express) oder Ihre **PDA** Software Palm Desktop senden. Die Erkennung von Visitenkarten ist eine nahtlose Ergänzung für Anwendungen wie Kontaktmanager, Datenbanken oder sogar Textverarbeitungsprogramme, deren „Serienbrieffunktion“ das Drucken von Briefen, Umschlägen und Etiketten ermöglicht.

## **VISITENKARTEN SCANNEN**

---

Wie funktioniert das alles? Zunächst einmal muss der Scanner richtig eingerichtet sein.

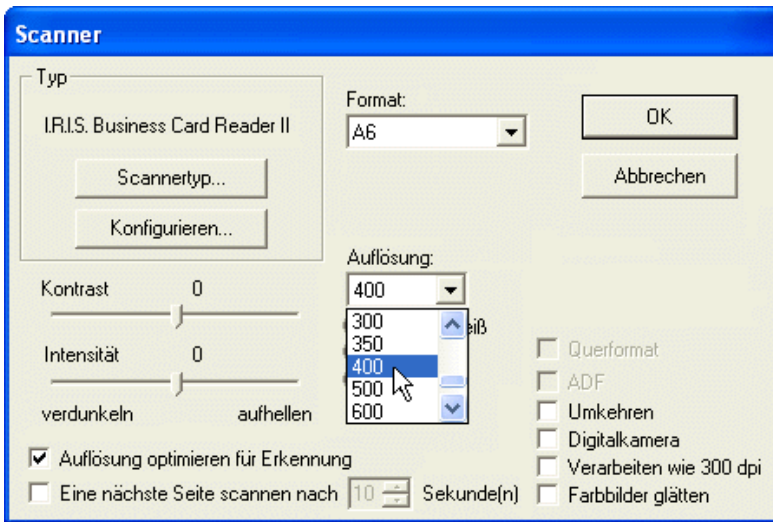
Wenn das Einlesen von Visitenkarten eine größere Anwendung ist - wenn Sie zum Beispiel eine große Anzahl von Visitenkarten auf einem Messestand einscannen -, können Sie dazu den speziellen Visitenkartenscanner **IBCR-II** von I.R.I.S. verwenden. Spezielle **Visitenkartenscanner**, die für das Scannen von Visitenkarten optimiert sind, weisen gegenüber anderen Scannertypen viele Vorteile auf: sie nehmen kaum Platz auf Ihrem Schreibtisch ein, können mit Visitenkarten von unterschiedlicher Papier- und Druckqualität umgehen, und Sie brauchen sich niemals über die Wahl des Formats Gedanken zu machen.

Verwenden Sie die Twain-Oberfläche des IBCR-II, um den **automatischen Scan-Modus** zu aktivieren. Geben Sie einmal den Scan-Befehl und aktivieren Sie die Option „Automatisch Scannen“ in der Twain-Benutzeroberfläche.



Sie können nun eine Visitenkarte nach der anderen einlegen: sobald eine Visitenkarte in den Scanner eingelegt wird, beginnt der Scanvorgang...

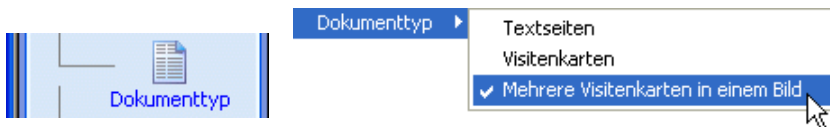
Um Visitenkarten erfolgreich einzuscannen, empfehlen wir eine Scan-**Auflösung** von 400 dpi.



Wenn Sie einen **Flachbettscanner** verwenden, können Sie mehrere Visitenkarten gleichzeitig auf der Vorlagefläche des Scanners einscannen und von der Software aufteilen lassen. Der Hintergrund muss schwarz sein, damit Readiris die verschiedenen Visitenkarten unterscheiden kann. Es gibt eine einfache Möglichkeit, das zu erreichen: lassen Sie beim Scannen der Visitenkarten die Dokumentabdeckung geöffnet!

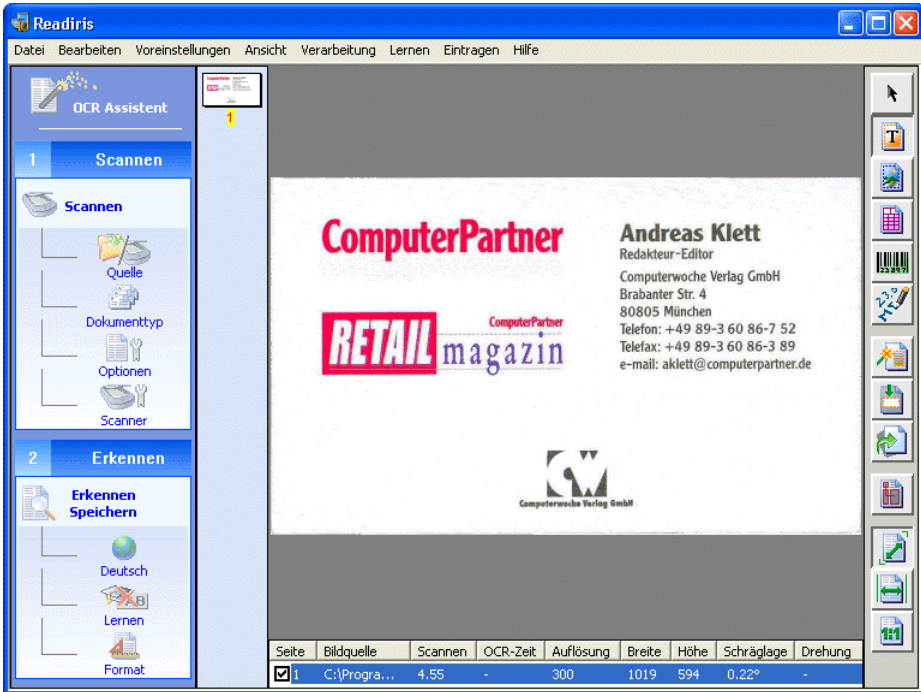


Wählen Sie die Option „Mehrere Visitenkarten in Einem Bild“ mit der Schaltfläche „Dokumenttyp“ in der Haupttoolbar (oder mit dem Befehl „Dokumenttyp“ im Menü „Voreinstellungen“) und scannen Sie Ihre Visitenkarten.



Mit Flachbettscannern braucht man nur auf die Schaltfläche „Scannen“ zu klicken, um die Bilderfassung zu starten. (Alternativ können Sie auch den Befehl „Scannen“ im Menü „Verarbeitung“ wählen, um ein Bild zu scannen.)

Die gesamte von Ihnen gescannte Seite wird niemals als solche am Bildschirm angezeigt; lediglich die aufgesplitteten Visitenkarten werden angezeigt.



Falls Sie vergessen haben, die Aufspaltung des Bildes im Modus für mehrere Karten zu aktivieren, klicken Sie auf den Befehl „Visitenkarten Ausziehen“ im Menü „Verarbeitung“, um das Gesamtbild in die einzelnen Visitenkarten aufzusplitten und die überflüssigen schwarzen Ränder zu entfernen.

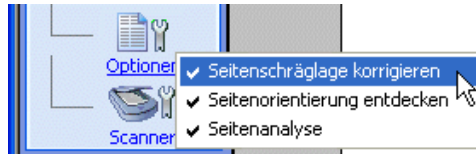
Visitenkarten ausziehen



Unabhängig von Ihrem Scannertyp werden einige Optionen wie Schräglagenkorrektur und Seitenausrichtung unsichtbar aktiviert, um den



Lesevorgang zu optimieren. Klicken Sie auf die Schaltfläche „Optionen“ in der Haupttoolbar, um diese Optionen zu sehen.



Es ist höchst unwahrscheinlich, dass Visitenkarten, die Sie auf einen Flachbettscanner legen, (alle) perfekt gerade liegen; Sie sollten die Software sich darum kümmern lassen... (Sollten Sie diese Option irrtümlich deaktiviert haben, können Sie die Visitenkarten immer noch nachträglich mit der Schaltfläche „Seitenschräglage Korrigieren“ in der Bild-Toolbar gerade ausrichten, aber es kann sehr schnell lästig werden, jede gescannte Visitenkarte manuell ausrichten zu müssen!)

Gleiches gilt für die Erkennung der Ausrichtung: lassen Sie die Software dies für Sie erledigen. Andernfalls müssen Sie die Visitenkarten, die um 90° gedreht oder auf dem Kopf auf dem Scanner lagen, manuell drehen. (Verwenden Sie die Drehungsschaltflächen in der Bild-Toolbar, falls Sie diese Option deaktiviert haben...)



Seitenschräglage korrigieren

## DEN LESEMODUS FÜR VISITENKARTEN AKTIVIEREN!

Wenn die einzelnen Visitenkartenbilder im Bildfenster richtig angeordnet sind, müssen Sie noch den **Visitenkarten-Lesemodus** mit der Schaltfläche „Dokumenttyp“ (oder mit der Befehl im Menü „Voreinstellungen“) aktivieren.





Wählen Sie die Option „Visitenkarten“, wenn Sie die Visitenkarten einzeln einlesen, wählen Sie „Mehrere Visitenkarten in Einem Bild“, um gleichzeitig mehrere Visitenkarten mit Ihrem Flachbettscanner einzulesen. Diese Option betrifft nicht nur die „Aufsplittung“ des gescannten Bilds: eine dieser beiden Optionen muss aktiviert sein, um eine *Erkennung* in Visitenkarten zu ermöglichen! Wählen Sie „Textseiten“, um den Visitenkarten-Lesemodus wieder zu deaktivieren.

Die Schaltfläche „Dokumenttyp“ bestätigt, dass der Lesemodus für Visitenkarten aktiviert ist.



Das Aktivieren dieses speziellen Modus ist notwendig, weil spezielle Texterkennungsrouitinen genutzt werden, mit denen die Software die erkannten Daten spezifischen Datenbank-Feldern zuzuordnen kann - Firmenname, E-Mail-Adressen und Websites usw. Wenn Readiris die Zeichenfolge „Johann Westkampf“ erkennt, „weiß“ es, dass Johann ein Vorname und Westkampf ein Nachname ist. Auf dieselbe Weise erkennt das System Titel, wie zum Beispiel „Vizepräsident“, „Diplomingenieur“, Städte wie Dortmund und Aachen, Bundesstaaten wie Hessen und Bayern usw.

Das funktioniert für bis zu 28 Länder! Die größeren Länder aller Kontinente werden unterstützt: nord- und südamerikanische Visitenkarten sowie Visitenkarten aus den europäischen Staaten, einschließlich der osteuropäischen Nationen. (Zum Einlesen von chinesischen, japanischen, koreanischen und taiwanischen Visitenkarten wird die Software-Option „Asiatisches BCR Modul“ benötigt. Englischsprachige Visitenkarten aus diesen asiatischen Ländern werden von der Standardversion der Readiris-Software unterstützt!)

Wenn dieser Lesemodus aktiviert ist, ändert sich der Inhalt des Dropdown-Menüs „Sprache“. Wählen Sie immer das richtige Land aus.



Durch die Auswahl des Landes der Visitenkarte in der Haupttoolbar geben Sie nicht nur die Sprache des zu erkennenden Textes an, sondern auch das allgemeine Layout, den Stil der Visitenkarte. Sieht sie eher amerikanisch oder deutsch aus? Jedes Land hat einen anderen „Stil“ bei seinen Visitenkarten: Die Amerikaner ordnen eine Anschrift anders an als die Franzosen, niederländische Postleitzahlen und Telefonnummern verwenden eine andere Syntax als ihre britischen Gegenstücke usw. (Manche Kartenstile entsprechen mehreren



Sprachen: Belgien und Kanada haben zwei Landessprachen, die Schweiz hat deren drei! Readiris erkennt die ausgewählte Sprache automatisch...)

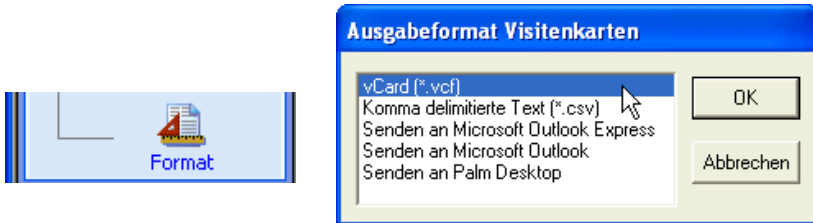
 <p><b>Recognition Systems</b></p> <p>Recognition Systems, Inc. 1520 Dell Avenue Campbell, CA 95008 USA www.handreader.com</p>	<p><b>Helen Thomas</b> International Sales</p> <p>(408) 341-4100 (408) 341-4101 Fax (408) 341-4135 Direct</p> <p><a href="mailto:helen_thomas@irco.com">helen_thomas@irco.com</a></p>	<p><b>APACABAR</b> La Référence Logicielle</p> <p><b>Stéphan Mallavergne</b> Coordinateur Technique</p> <p>10 rue Vercingétorix 75680 Paris cedex 14 Tel 01.56.54.80.22 Fax 01.56.54.80.01 E-mail : <a href="mailto:smallavergne@apacabar.fr">smallavergne@apacabar.fr</a> Web : <a href="http://www.apacabar.fr">http://www.apacabar.fr</a></p> <p></p>
---	---	---

Die Feldanalyse umfasst noch eine dritte Aufgabe: Readiris analysiert den erkannten Text nicht nur, sondern *formatiert* ihn auch. Das System filtert alle irrelevanten Daten aus einer Visitenkarte heraus, selbst wenn sie eine aktive Rolle spielen! Wenn Ihre Visitenkarte die Angabe „Telefon: (508) 898-42 89“ enthält, ergibt sich daraus als Ausgabewert „5088984289“ im Telefon-Feld. Das Wort „Telefon“ wird ausgelassen, obwohl Readiris es benutzt hat, um die Telefonnummer zu erkennen. Klammern, Bindestriche und Leerstellen innerhalb der Telefonnummer werden ebenfalls gelöscht.

## VISITENKARTEN ERKENNEN

Verarbeiten Sie Ihre Scans wie gewohnt, indem Sie auf die Schaltfläche „Erkennen-Speichern“ klicken.

Es ist jedoch ratsam, das Ausgabeformat bereits vorher festzulegen. Klicken Sie hierzu auf die Schaltfläche „Format“.



Eine Reihe weit verbreiteter „Personal Information Manager“ („PIMs“) werden direkt unterstützt: Microsoft Outlook (Express) und Ihre PDA Software Palm Desktop. Um eine Verbindung zu anderen Anwendungen zu ermöglichen, können Sie die „universellen“ Formate vCard und durch Kommata getrennte Daten verwenden. (vCard Dateien werden problemlos in jede vCard kompatible Anwendung importiert: doppelklicken Sie auf eine vCard Datei, um die Kontaktinformationen in Ihr Windows Adressbuch aufzunehmen!)

Das Textergebnis sieht beispielsweise wie folgt aus, wenn Sie das Ergebnis direkt an Microsoft Outlook (Express) senden.



**Kirk Y. Matsuo - Contact**

File Edit View Insert Format Tools Actions Help

Save and Close

General Details Activities Certificates All Fields

Full Name... Kirk Y. Matsuo

Job title: Product Manager Business & Home/Office Proc

Company: Seiko Instruments USA Inc.

File as: Matsuo, Kirk Y.

E-mail... kirk.matsuo@seikosi.com

Display as: Kirk Y. Matsuo (kirk.matsuo@seikosi.com)

Web page address: <http://www.seikosmart.com>

IM address:

Phone numbers

Business... 40843210

Home...

Business Fax... 408 4333261

Mobile...

Addresses

Business... 1130 Ringwood Court  
95131 San Jose California

This is the mailing address

Contacts... Categories... Private

## 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.

