

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-Taste im Submenü "I.R.I.S. Anwendungen - Readiris" oder klicken Sie die Verknöpfung in Ihrem Desktop.



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 den URL, um die Website von



I.R.I.S. zu öffnen. Mit einem Mausklick außerhalb dieses URL verschwindet dieses Dialogfenster.



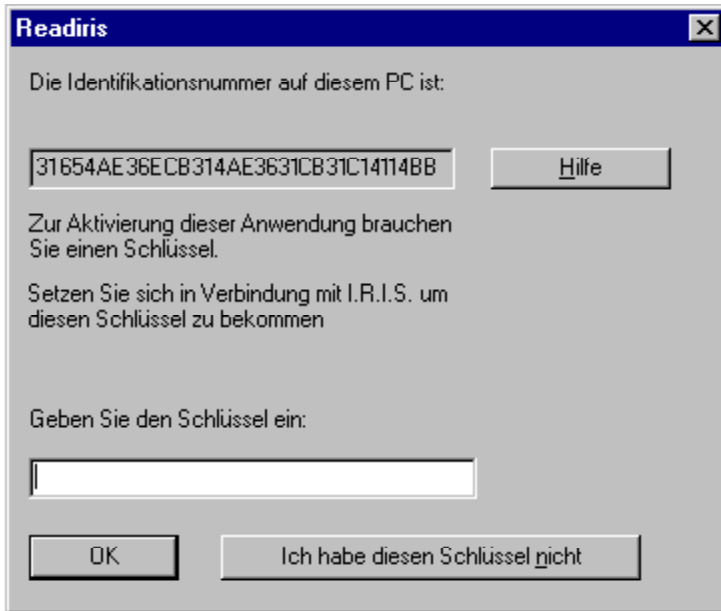
Das nächste Fenster bezieht sich auf den OCR Assistenten. Klicken Sie in dem Augenblick auf "Annullieren".

BEIM ERSTEN START

Je nach Software, werden Sie beim ersten Start gefragt, Ihre Lizenz zu registrieren.

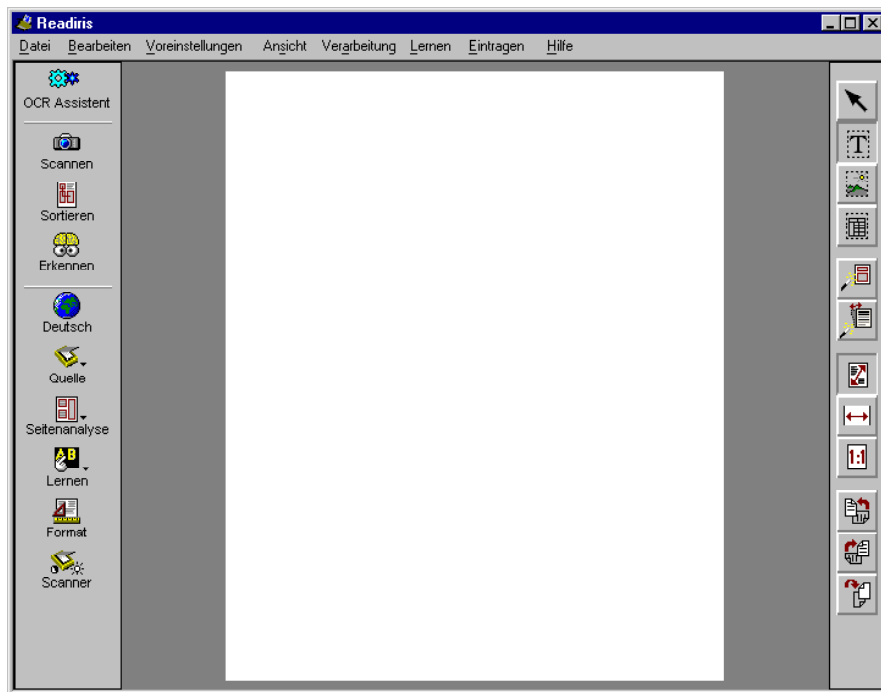
In diesem Fall ist der Gebrauch von Readiris auf einem Monat beschränkt. Bei Registrierung bekommen Sie einen **Softkey** gratis um die Software nach 30 Tagen weiter zu benutzen.

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



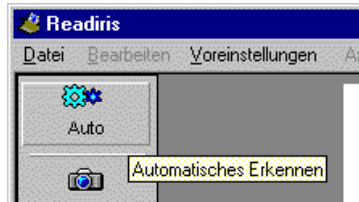
ENTDECKEN SIE DAS READIRIS INTERFACE

Das Readiris Anwendungsfenster enthält nicht nur **Befehlsmenüs**, sondern auch zwei Funktionsleisten, die Ihnen den schnellen Zugriff zu den meisten Befehlen ermöglichen. Anfänglich sind die Menüs m.B.a. der Vorschau ausgeblendet: solange noch keine Bilddateien geöffnet wurden, bleiben sie ausgeblendet.



Das gleiche gilt für die **Bildfunktionsleiste**, die sich rechts in Ihrem Anwendungsfenster befindet. Die Bildfunktionsleiste enthält alle für die Bildvorschau notwendigen Befehle. Die **Hauptfunktionsleiste** links ermöglicht Ihnen den schnellen Zugriff zu fast allen Befehlen.

Damit Sie wissen, welcher Befehl mit welcher Funktionstaste übereinstimmt, brauchen Sie nur den Mausschalter kurz über die Taste zu bewegen: das **QuickInfo** teilt Ihnen die Funktion der Taste mit.



Es gibt auch das **Fenster**, in dem die gescannten Bilder angezeigt werden. Zur Erkennung der Bilddateien können diese in das Anzeigefenster gezogen werden ("Drag & Drop").

Wenn Seiten verarbeitet werden, erscheint eine weitere Toolbar, die **Seiten-Toolbar**, unter dem Bildbereich. In dieser Toolbar sind die verschiedenen Seiten des Dokuments aufgeführt. Sie können auf die Seiten-Befehle zugreifen, indem Sie rechts klicken ("Kontext" Menü).



BEGINNEN MIT EINER ERSTEN EINFÜHRUNG

Die beste und einfachste Art, sich mit Readiris anzuvertrauen ist es zu benutzen. Zu Ihrer Software gehört eine Zahl von **vorgescannten Bildern**, die Ihnen den Start auch ohne Scanner ermöglichen.

Die Funktion "Quelle" auf der Funktionsleiste bestimmt, ob Sie den Scanner oder das vorgescannte Bild als Bildquelle benutzen werden.

Farb-, Graustufen- und Schwarz-weißbilder werden gleichermaßen unterstützt. Mit Readiris können Sie JPEG-Bilder, Paintbrush-Bilder (PCX), mehrseitige Paintbrush-Bilder (PCX), TIFF-Bilder (nicht-komprimiert, PackBits, Gruppe 3 und Gruppe 4 komprimiert), mehrseitige TIFF-Bilder und Windows Bitmap-Dateien (BMP) öffnen.



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

Bevor Sie ein vorgescanntes Bild als Bildquelle öffnen, selektieren Sie die Diskette und nicht den Scanner mit der Schaltfläche "Quelle" auf der Hauptfunktionsleiste.



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



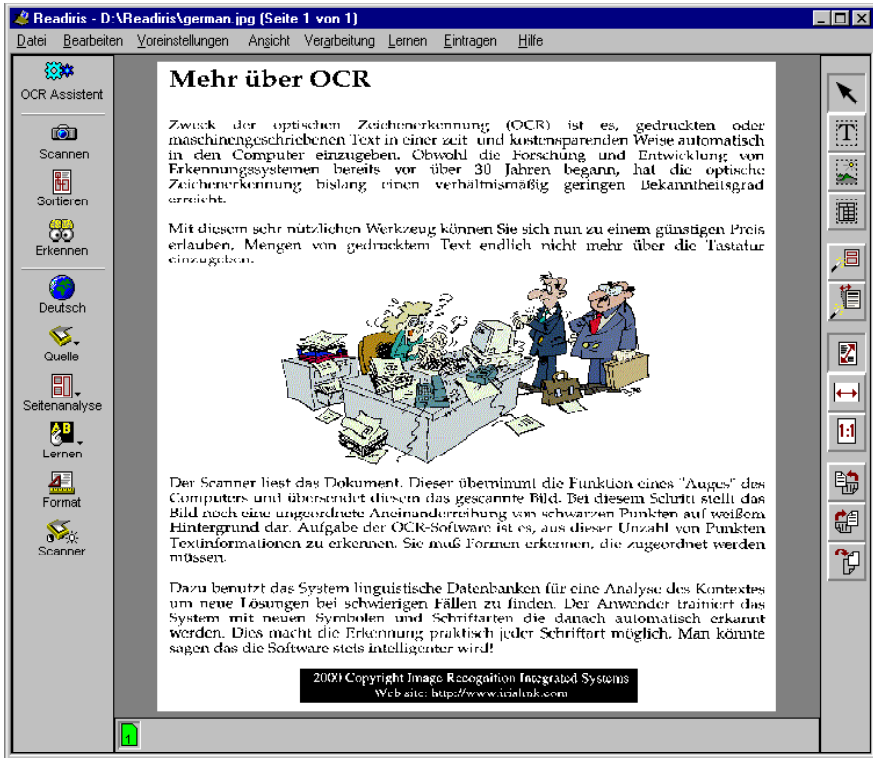
Um das vorgescannte Bild direkt zu öffnen, können Sie auch den Befehl "Öffnen" im "Datei" Menü selektieren. Diese Funktion arbeitet sogar wenn Ihr Scanner als Bildquelle läuft.



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

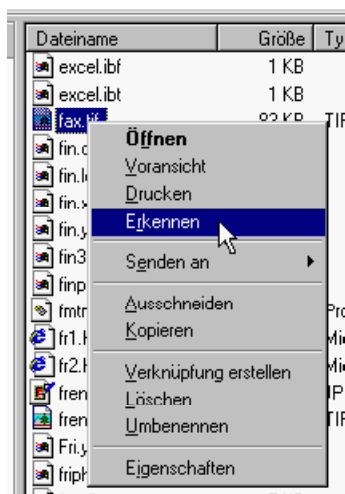


Dieses Bild wird schließlich im Bildbereich angezeigt. In der Seiten-Toolbar sieht man, dass eine einzige Seite in Readiris eingelesen wurde.



Eine dritte Möglichkeit ist, vorgescannte Bilder mit Hilfe der "Drag & Drop"-Funktion zu öffnen. Ziehen Sie Bilder vom Windows Explorer zum Readiris Bildfenster. Die Bilder werden sofort 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 Bildfunktionsleiste rechts im Readiris Anwendungsfenster enthält alle notwendigen Befehle für die Bildvorschau: Funktionstasten um die zu bearbeitenden Blöcke zu bestimmen, um das Bild zu drehen, zum Ein- und Auszoomen, usw.

AUF BILDER ZOOMEN

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

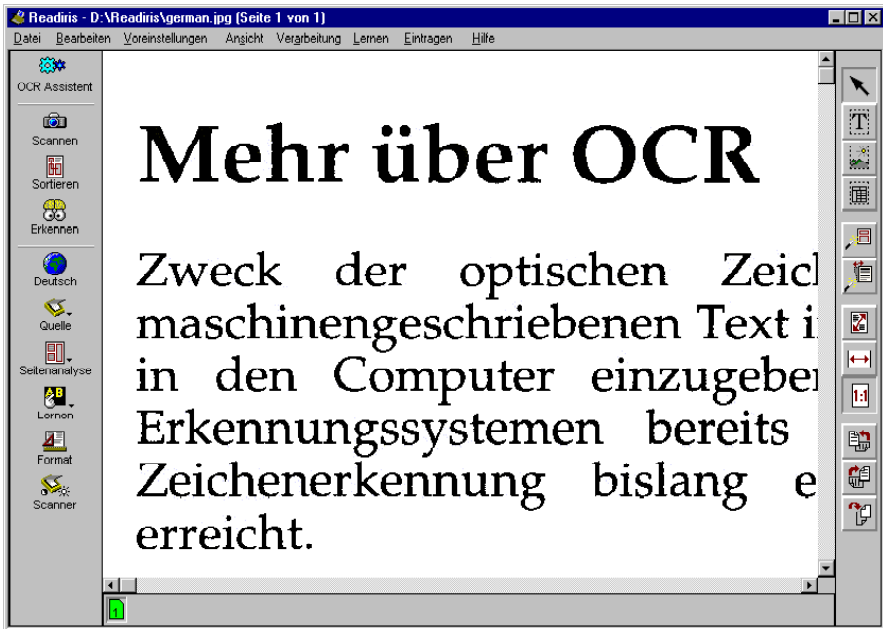
Auf der Bildfunktionsleiste befinden sich Funktionstasten, mit denen Sie das gescannte Bild auf die wahre Größe bzw. auf Seitenbreite oder auf Bildschirmgröße einzoomen können. Das Menü "Ansicht" enthält noch zwei zusätzliche Zoomstufen:

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!)

**Ansicht**

<u>G</u> anze Bild	Ctrl+F
Ganze <u>B</u> reite	Ctrl+W
<u>5</u> 0% Verkleinerung	Ctrl+5
<u>R</u> eelle Größe	Ctrl+1
<u>2</u> 00% Vergrößerung	Ctrl+2

Schließlich können Sie durch Anklicken und Festhalten der *rechten* Taste, die Maus über den zu vergrößernden Teil des gescannten Bildes ziehen. Mit einem zweiten Klick der rechten Maustaste wird wieder ausgezoomt.



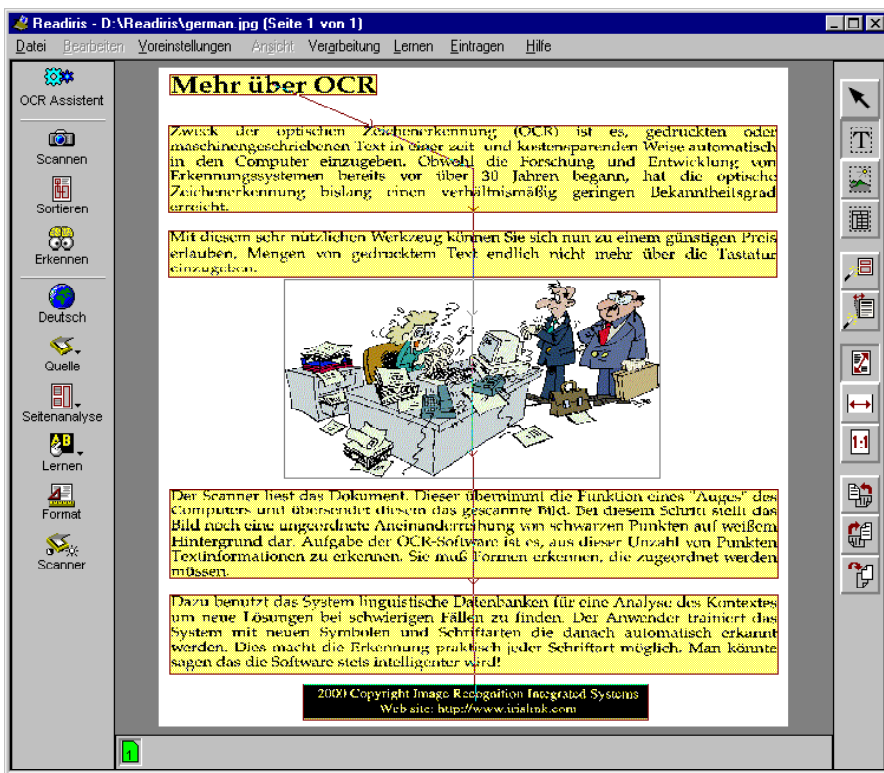
EIN, ZERLEGUNG EINES GESCANNTEN BILDES

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

Wenn die Option "Seitenanalyse" auf der Hauptfunktionsleiste aktiviert wurde, wird diese Handlung automatisch von Readiris ausgeführt.



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



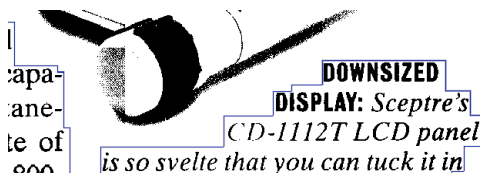
Grafiken und erkannte Tabellen speichert, wird weiter unten noch lang und breit erklärt.)

Ein **Farbcode** gibt den Fenstertyp an: Textzonen sind gelb, Tabellen sind violett. Grafikfenster haben einen blauen Rand, aber keine "Füllfarbe" - es wäre nicht zweckmäßig, die tatsächlichen Farbtöne der Grafik zu überlagern.

Die Zahl der Fenster wird jederzeit in dem QuickInfo der "Textfenster", "Grafikfenster" und "Tabellenfenster"-Symbolen 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: da 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.)

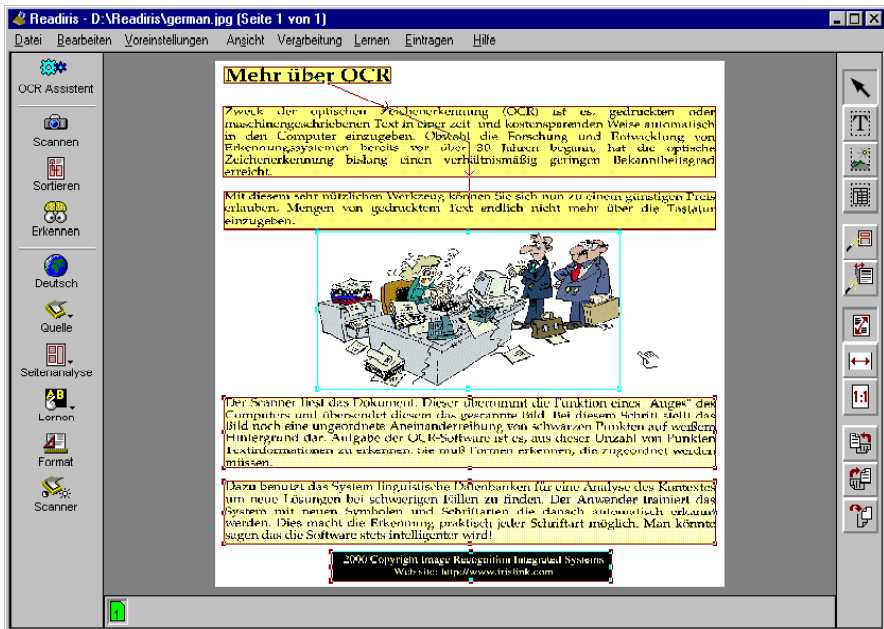
ANDERTHALB, FENSTER SORTIEREN

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

Sie können die **Sortierreihenfolge** natürlich auch selbst bestimmen - hierzu klicken Sie die Taste "Sortieren" in der Hauptfunktionsleiste an. Der Mauscursor wird zum Zeigefinger, wenn die Sortierfunktion aktiviert ist.



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





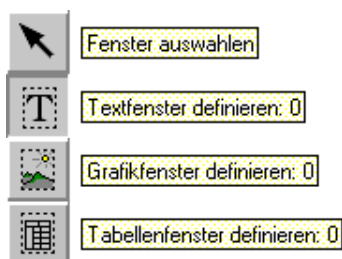
Die Seitenanalyse wird als Standardeinstellung eingestellt. Wenn die Seitenanalyse inaktiviert wurde weil Sie z.B. versehentlich einige Fenster gelöscht haben, können Sie wieder aktiviert werden, indem Sie die Taste "Seite Analysieren" in der Bildfunktionsleiste anklicken.



(Wenn Sie asiatische Dokumente bearbeiten, wählen Sie die Dokumentensprache, *bevor* Sie die Seitenanalyse ausführen. Für diese Sprachen gelten ganz spezifische Handlungsweisen: im allgemeinen ist der Zeilenabstand in asiatischen Dokumenten größer als in westlichen Dokumenten, der Text besteht aus kleinen Ikonen ("Ideogrammen") die in westlichen Dokumenten mit Grafiken verwechselt werden können, und der Text kann von oben nach unten und von rechts nach links gelesen werden.)

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 **Fenstersymbole** von Readiris.

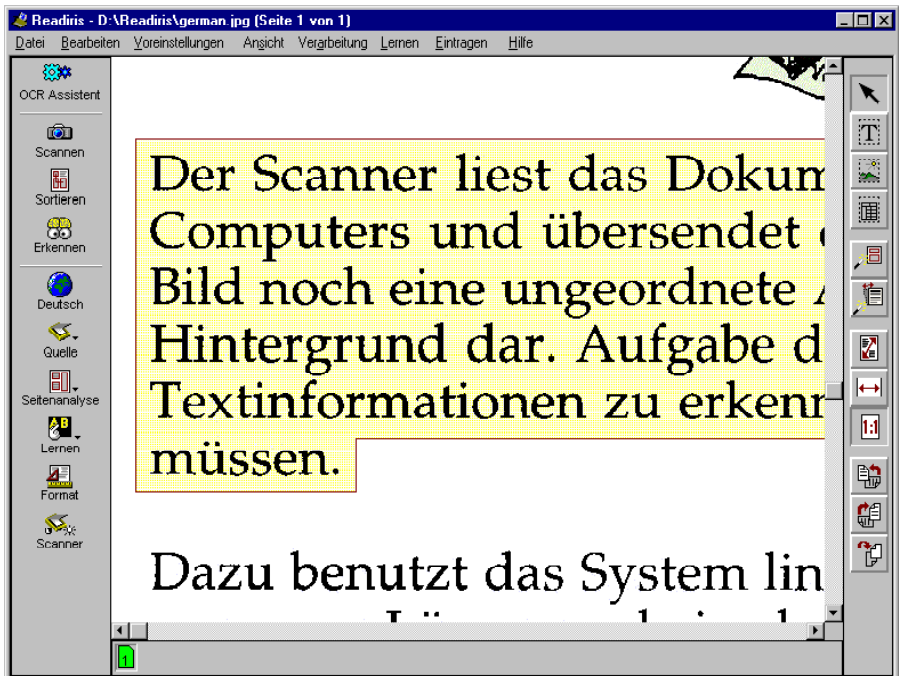


Um ein rechteckiges Fenster um eine gewünschte Zone zu **erstellen**, wählen Sie das entsprechende Symbol in der Bildfunktionsleiste, bewegen Sie den Cursor in die rechte Ecke oben am Fenster und ziehen Sie das Fenster durch bewegen

des Mauszeigers in die untere rechte Ecke und klicken Sie nochmals. (Seiten von weniger als 1 mm. sind nicht erlaubt Sie würden ohnehin kein Zeichen enthalten.)

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

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





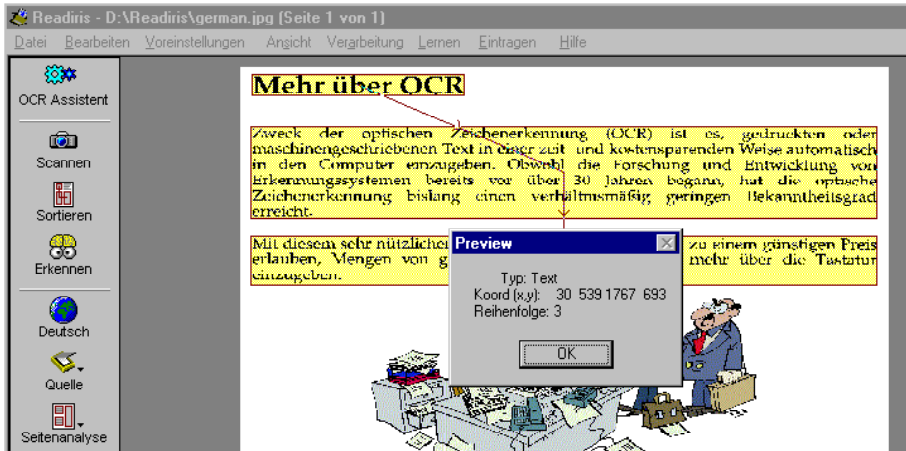
Weiters kann manuelle Fensterbestimmung mit Fenster Sortierung kombiniert werden: selbst wenn die Sortierungsfunktion aktiviert ist, können Sie neue Fenster zeichnen. Sie verwenden dann die Sortierungsfunktion, um einige der erkannten Fenster zu behalten und bestimmen dann manuell neue Fenster, dort, 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, gelöscht!

Um Fenster zu verändern, zu verschieben und zu löschen müssen sie zuerst **bestimmt** werden. Selektieren Sie hierzu "Fenster Selektieren" (das "Pfeil"-Symbol) in der Bildfunktionsleiste und klicken Sie in das Fenster. Es erscheinen jetzt rechteckige Aktivpunkte in jeder Ecke und in der Mitte des Fensters.



Um die **Selektion** eines Fensters zu **annulieren**, klicken Sie mit der linken Maustaste 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 (des anderen Typs) befindlichen Fenster zu aktivieren, halten Sie die Strg-Taste gedrückt und klicken Sie dabei das Hauptfenster an.

Beachten Sie, daß die Sortierreihenfolge, den Typ und die Koordinaten angezeigt werden können indem die Alt-Taste gedrückt bleibt während Sie selektieren.

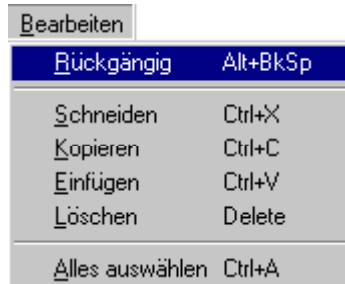


Soweit die Fensterauswahl. Um ein Textfenster zu **verändern**, selektieren Sie zuerst das Fenster, bewegen Sie anschließend den Mauszeiger über einen der Aktivpunkte und bewegen Sie die Maus um die Fenstergröße zu verändern.

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 und selektieren Sie den Befehl "Schneiden" oder "Löschen" aus dem Menü "Bearbeiten". Der Befehl "Schneiden" schneidet das Fenster aus und fügt es in ein interner Puffer ein. Der Befehl "Löschen" löscht das Fenster.

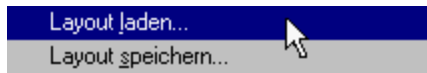
Es können also *alle* bekannten Befehle aus dem Menü "Bearbeiten" bei den Fenstern angewendet werden: Sie können löschen, ausschneiden, kopieren und einfügen. Der Befehl "Rückgängig" kann ebenfalls angewendet werden. Wenn Sie versehentlich einige Fenster gelöscht, bewegt, verändert, usw. haben, wird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



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" benutzen oder die Taste Delete. Jetzt kann das Layout wiederhergestellt werden. Um das vorherige Layout wiederherzustellen, selektieren Sie "Rückgängig" oder benutzen Sie die Tastenkombination Alt+Backspace.

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, z.B. ein 50-seitiges Protokoll, wo der Briefkopf und die Fußnote 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 das Menü "Bearbeiten" automatisch inaktiviert. Die Layoutvorlage bleibt solange aktiviert bis die Funktion Seitenanalyse auf der Hauptfunktionsleiste wieder aktiviert wird.

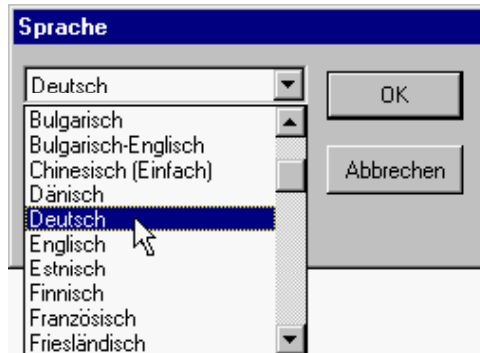
MIT READIRIS UM DIE WELT

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 Spracheinstellungen finden Sie auf der Hauptfunktionsleiste.



Um die Dokumentensprache zu verändern klicken Sie "Sprache" an.



Sie können auch direkt eine Taste auf der Tastatur drücken: wenn Griechisch oft gewählt wird und Sie möchten jetzt Indonesisch selektieren, drücken Sie dann die "G"-Taste auf Ihrer Tastatur um direkt Zugang zur griechischen 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 Polnisch und Portugiesisch lesen. Wenn Sie das "P" einmal drücken, selektieren Sie Polnisch, wenn Sie das "P" ein zweites Mal drücken, selektieren Sie Portugiesisch und beim dritten Mal sind Sie wieder bei Polnisch.



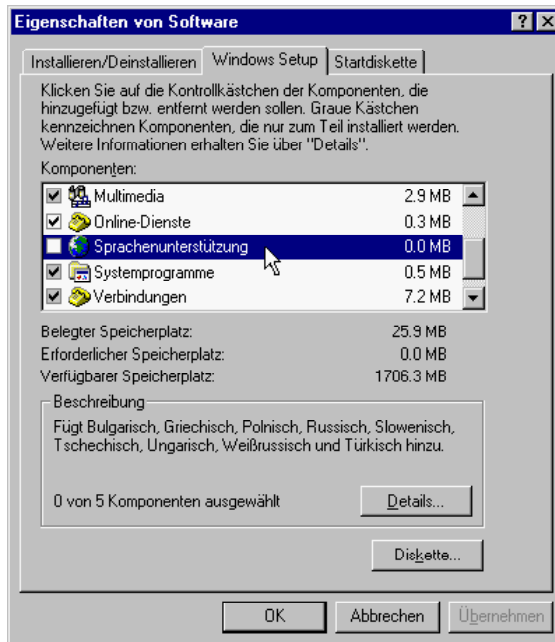
(Um einen *anderen* Buchstaben, sagen wir ein T, einzugeben, gehen Sie zuerst auf Zurück, bevor Sie das T eingeben.)

Readiris beschränkt sich nicht nur auf Deutsch: es werden bis zu 56 **Sprachen** unterstützt! Readiris unterstützt alle amerikanischen und europäischen Sprachen sowie die zentraleuropäischen, griechischen, türkischen und kyrillischen (“russischen”) und baltischen.

(Es ist optional möglich, **asiatische Dokumente** zu lesen: die Software-Option “Asiatisches OCR-Modul” ermöglicht die Erkennung der japanischen und vereinfachten chinesischen Sprache. Die traditionelle chinesische Sprache wird nicht unterstützt. (Die vereinfachte chinesische Sprache wird auf dem chinesischen Kontinent angewendet, die traditionelle chinesische Sprache wird von den “5 Großen” Gemeinschaften - Hong Kong, Taiwan, Singapur, usw. angewendet.)

Merken Sie sich auch, daß die britischen und amerikanischen - oder sagen wir “internationalen”? - Varianten der englischen Sprache ebenfalls differenziert werden.

Merken Sie sich ebenfalls, daß Sie die angepaßte Windows-Konfiguration installieren müssen, damit zentraleuropäische, griechische, türkische und baltische Zeichen angezeigt werden. Vielleicht werden Sie das Windows-Modul "Sprachenunterstützung" installieren müssen, bevor Ihr Windows-System in der Lage ist mit diesen Sprachen zu arbeiten. Um zu erfahren, ob diese Windows-Modul in Ihrem PC installiert wurde, selektieren Sie das Symbol "Software" im "Systemsteuerung". Abhängig von Ihrer Software-Version enthält die Readiris CD-ROM detaillierte Informationen über die Installation dieser Windows-Modul.



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

Es ist sehr wichtig, die dokumenteneigenen Sprache zu selektieren. Die Software weiß genau welche Zeichen sie zu erkennen hat. Dieses ist natürlich



abhängig von Ihrer Sprachenauswahl. Eine multilinguistische Datenbank hilft bei der Erkennung von “exotischen” Zeichen so 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 einen Tintenfleck, so daß das "r" wie ein "f" aussieht. Bei der Kontrolle im deutschen Wörterbuch, wird Readiris automatisch feststellen, daß das Wort "Präsident" gelesen wurden und daß ein "f" in diesem Kontext keinen Sinn ergeben würde. Diese “**autodidaktische**” **Technik** ist natürlich sehr abhängig vom linguistischen Kontext.

Die linguistische Unterstützung kann sehr hilfreich sein bei der Erkennung von ”**Zweifelsfällen**”, so wie das "O", das als "()" gelesen werden könnte. Ein anderes typisches Beispiel ist "l" und 'l', 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 'l'.

Folgendes Beispiel illustriert die Schattenvariationen von 'l' und "l". Die Schatten der ersten Linien sind deutlich abgegrenzt, die der zweiten Linien sind zweifelhaft abgegrenzt. Dieses Problem wird gelöst durch die Sprachwissenschaft.

193 1950s. 1hr
Well, Rossellini

BEI BEDARF VERÄNDERT READIRIS DIE SPRACHEN

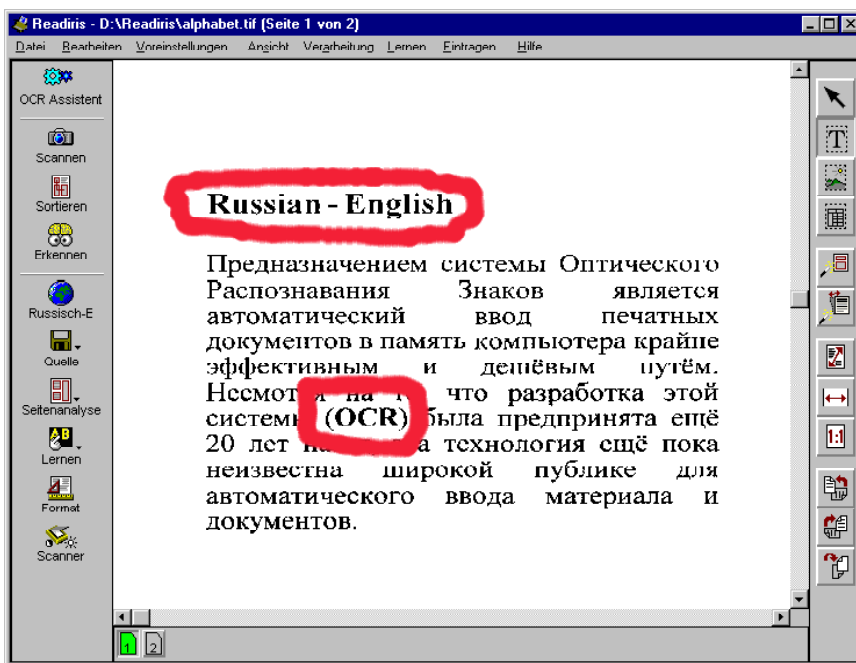
Aber das ist noch lange nicht alles. Readiris ist in der Lage, ohne Eingriff des Anwenders, in der Mitte eines Textes von einer Sprache zu einer anderen zu wechseln. Wenn “Westliche” Worte auftauchen in griechischen, kyrillischen oder asiatischen Dokumenten - viele Eigennamen, Markennamen usw. werden nicht umgesetzt aber mit “normalen”, lateinischen Zeichen geschrieben. Readiris kann sofort zum richtigen Alphabet wechseln. Mit anderen Worte, es ist möglich,

”**gemischte**” **Alphabete** von griechischen, kyrillischen oder asiatischen und westlichen Zeichen gleichzeitig zu aktivieren.

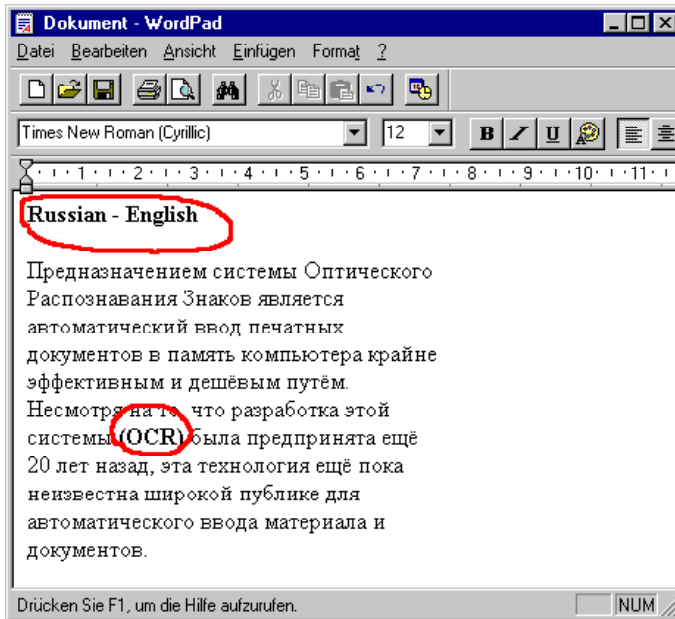
Vergewissern Sie sich, daß die "Griechisch-Englische" oder eine übereinstimmende kyrillische Einstellung gewählt wurde - z.B. "Weiß-Russisch-Englisch". Mit anderen Worten: selektieren Sie nicht nur "Griechisch" oder "Weiß-Russisch" als Dokumentensprache in der Hoffnung, daß die westlichen Symbole perfekt erscheinen werden!



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



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



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

BESTIMMEN DER DOKUMENTEIGENSCHAFTEN

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



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

Zuerst sollten Sie angeben, ob "normal" gedruckte oder **Matrix** gedruckte Dokumente erkannt werden sollen. Selektieren Sie den Befehl "Typ der Schriftart" im Menü "Voreinstellungen".

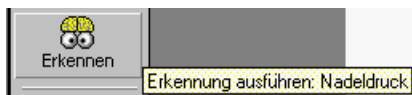


Dieser Parameter muß unbedingt richtig eingestellt werden: "9 Pin" Dot Matrix Symbole sind aus vielen einzelnen Punkten zusammengesetzt. Hier ist ein hoch spezialisiertes Erkennungssystem erforderlich.

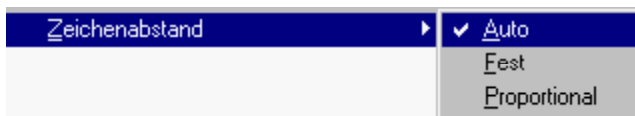
ape-descended life

"Briefqualität" Dot Matrix-Ausdrucke, auch "25 Pin" genannt, erfordern normale Einstellungen, so wie die **Druckeinstellungen** Druck, Maschinenschrift, Laserdruck und Inkjet-Druck. "Normal" ist die Standardeinstellung.

Der Fonttyp ist in dem Quickinfo der Schaltfläche "Erkennen" angegeben: wenn dem Quickinfo keine Angabe hinzugefügt wird, bleibt die normale Druckqualität, wenn die Angabe "Nadeldruck" angezeigt wird, ist die Nadeldruck Lesart 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 wurde.

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

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

READIRIS WIRD IMMER INTELLIGENTER!

Wenn die Dokumentensprache und die Dokumenteigenschaften eingestellt sind, können Sie die Schaltfläche "Erkennen" anklicken.



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 Hauptfunktionsleiste. Interaktives Lernen ist standardmäßig aktiviert.

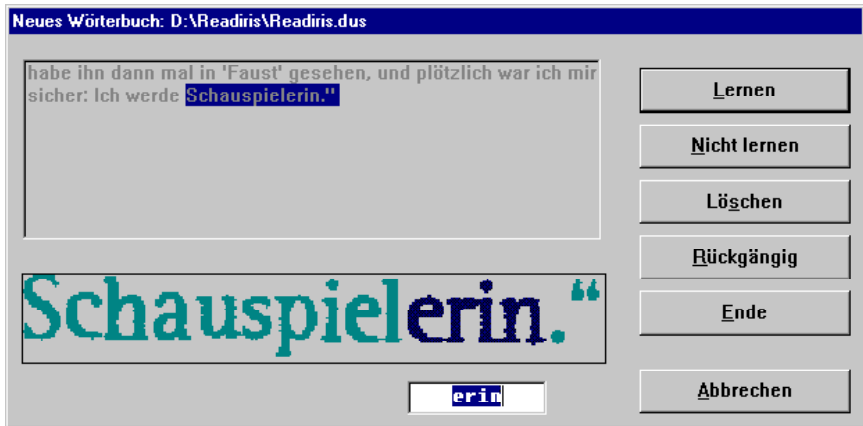
(Das interaktive Lernen findet keine Anwendung auf asiatische Dokumente: Lernen in diesen Sprachen ist zwecklos, weil es sich hier um tausende verschiedener Symbole handelt - und Sie müßten in der Lage sein, die Ideogramme mit einem westlichen Keyboard einzugeben!)



Zeichentraining ist 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. Mathematische und wissenschaftliche Zeichen und Dingbats können von Readiris nicht erkannt werden. Einige Beispiele: Readiris kann lernen, das π -Symbol als "pi" oder das Dingbat "☎" als "Tel" zu erkennen. Die Liste mit den erkannten Zeichen kann aber nicht mit den Symbolen " π " und "☎" aufgestockt werde.)

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 ein Tilde (das "~"-Symbol) dargestellt.



Kontrollieren Sie zuerst, ob das richtige Zeichenwörterbuch und den Wörterbuch-Modus aktiviert wurden. Beide werden immer im Fenster "Lernen" 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 starten Sie OCR neu. (Die Funktion "Zeichenwörterbuch" wird bald besprochen werden.)

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

Lernen

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

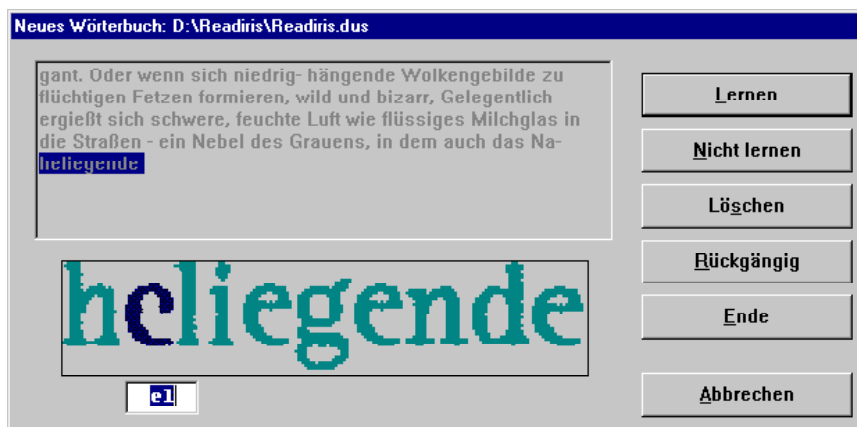
Im oben erwähnten Beispiel hält das System bei einige zusammenlaufenden Zeichen an. Wir klicken "Lernen", um die Zeichen, die nicht mit andere kunnen verwechselt werden, zu akzeptieren.



Nicht Lernen

Sie sind mit der vorgeschlagenen Lösung einverstanden, oder Sie haben die Lösung korrigiert. Der Unterschied mit der Funktion "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 "e" ist stark beschädigt - eigentlich sieht es mehr aus wie das "c"-Symbol. Jetzt sollten die Schaltfläche "Nicht Lernen" anklicken, damit es nicht mit dem "c"-Symbol verwechselt wird.

Löschen

Das angezeigte Zeichen verschwindet von der Ausgabedatei. Diese Schaltfläche wird benutzt, um bestimmte "Verunreinigungen" so wie z.B. Flecke,

Kaffeeeflecke, 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 neun Befehle rückgängig machen.

Ende

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

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

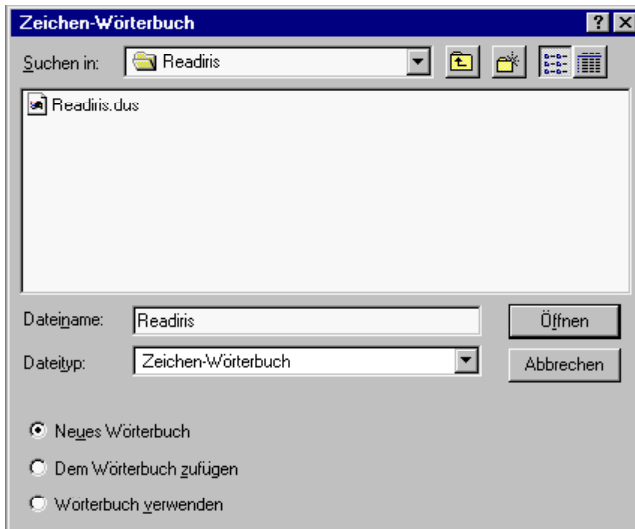
Verwechseln Sie "Ende" nicht mit "Abbrechen". Bei "Abbrechen" gibt es keine Textausgabe und müssen Sie neu anfangen. Bei "Ende" wurde der Text erstellt, aber noch nicht richtig korrigiert!

DIE ROLLE DER ZEICHENWÖRTERBÜCHER

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

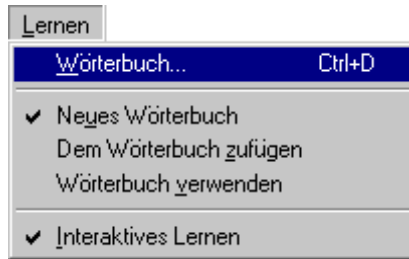
Wenn Sie ähnliche Dokumente erkennen möchten und die Intelligenz Zeichenwörterbücher nutzen wollen, ist es ratsam diese zu speichern. So wird Readiris immer intelligenter!

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, z.B. nach Typ oder nach Dokument. Es ist von großer Bedeutung die Wörterbücher deutlich zu benennen, z.B. BERICHT.DUS, PALATINO.DUS usw. Wörterbücher haben die Standarderweiterung *.DUS. Wenn ein Wörterbuch voll ist, ist es zwecklos noch weiter zu lesen. Die Ergebnisse können dann weder im Speicher, noch im Wörterbuch festgehalten werden.

Der Wörterbuch-Modus kann im Befehl "Wörterbuch" oder direkt im Menü "Lernen" eingestellt werden. Es gibt drei Wörterbuch-Modi: 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.)

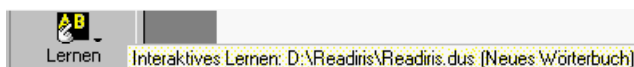
Der Modus "Zufügen" gibt an, daß die Ergebnisse in einem bereits existierenden Wörterbuch gespeichert werden: das Erkennungssystem nutzt bereits die zusätzliche Intelligenz des Wörterbücher und Sie fügen neue Zeichen hinzu. Mit anderen Worten: mit dieser Option wird ein Zeichenwörterbuch schrittweise aufgebaut.

(Wenn Sie für ein neues Wörterbuch einen neuen Dateinamen eingeben und den "Zufügen-Modus" aktivieren, wird ein leeres Zeichenwörterbuch 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 können es aber verwenden.

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 Hinzufügen für einige darauffolgende Seiten und den Modus Verwenden für den Rest des Dokumentes.

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





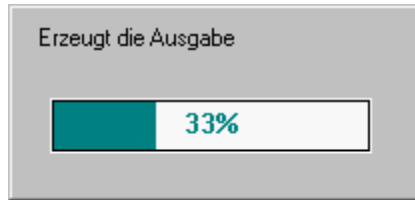
Wenn Sie im interaktiven Lernbereich sind, werden die Wörterbücher und ihre Modi im Fenstertitel angezeigt. Wenn es sich um einen Fehler handelt, sollten Sie "Abbrechen" klicken.



DAS ERGEBNIS DIREKT IN IHRE ANWENDUNG SENDEN

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

Am Bildschirm erscheint ggf. ein Fortschrittsbalken, der den Fortschritt der Dokumentformatierung anzeigt. (Ob dieser Balken erscheint oder nicht, ist von der Größe des Dokuments und der Komplexität der auszuführenden Formatierung abhängig.)



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

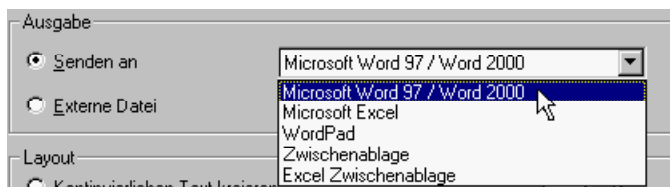
Sie haben jetzt ein Papierdokument in eine editierbare Textdatei verwandelt. Einen Vorgang der 30 Mal schneller ist, als das manuelle Abschreiben! Sie können es jetzt mit dem Bild, das Sie in Ihrem Readiris Fenster haben, vergleichen.

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 Hauptfunktionsleiste (oder des Befehl "Textformat" unter dem Menü "Voreinstellungen") gewählt.



Die Funktion “Senden nach” bietet eine direkte OCR Verbindung zwischen Ihrem Scanner und Ihrer Windows Anwendung: Sie **senden** die gescannten Dokumente direkt in Ihre Textverarbeitung oder in Ihre Spreadsheet! Readiris exportiert die erkannten Dokumente direkt in Microsoft Word 2000 und 97, Microsoft Excel und in das Windows Programm WordPad.



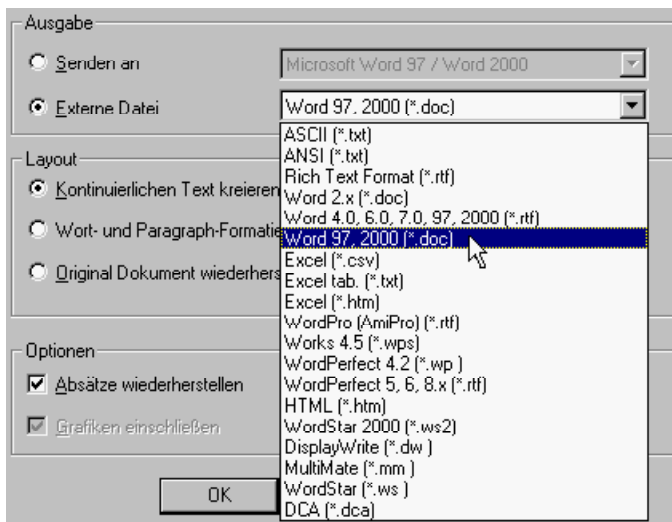
Am Ende der Erkennung, wird die gewünschte Anwendung geöffnet und das erkannte Dokument wird in einer neuen Textdatei oder Worksheet geöffnet.

Bitte warten, während Microsoft
Word 97 / Word 2000 geladen wird

Vergessen Sie nicht, dass Sie mit der Option "Senden nach" 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

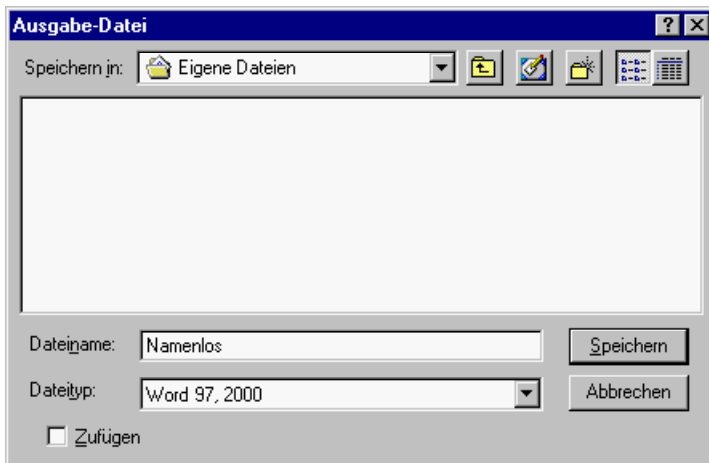
Sie können das OCR Ergebnis tatsächlich in eine "externe" Datei einschreiben. Readiris unterstützt eine grosse Palette von Datei Formaten, wie unter anderem alle bekannten Textverarbeitungen und Spreadsheets - Microsoft Word (DOC), RTF und HTML usw.



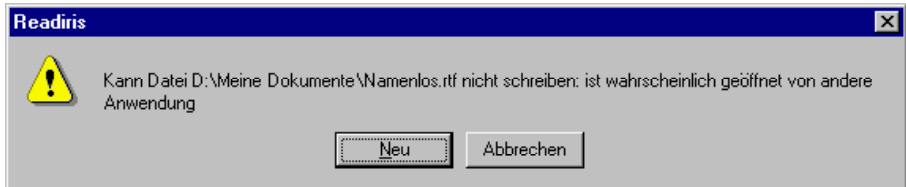


Wenn Sie die Option "Externe Datei" wählen, werden Sie am Ende der Erkennungsphase gebeten, den erkannten Text zu speichern.

Sie können eine neue Datei erstellen und in eine existierenden Textdatei zufügen - deswegen haben Sie eine Option "Zufügen". Sie können z.B. zwei gescannte Seiten mit Text, Ihrer Studie, die Sie selber geschrieben haben, zufügen. Wenn Sie eine existierende Datei hinzufügen, seien Sie sicher, dass sie nicht geöffnet ist, denn das wird Sie daran hindern hineinzuschreiben.

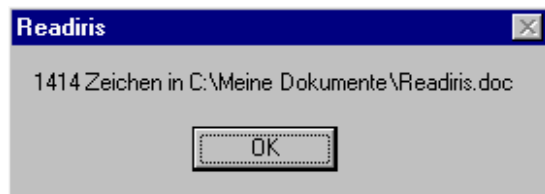


Sie können entweder eine neue Datei erstellen und eine existierende Textdatei hinzufügen - hierzu gibt es die Funktion "Zufügen". Sie könnten z.B. zwei Seiten eines gescannten Textes einem Dokument, das Sie selbst geschrieben haben, zufügen. Sie sollten sich vergewissern, daß diese Datei nicht geöffnet ist, sonst können Sie nichts zufügen!



MEHRERE SEITEN ERKENNEN

Wenn Sie auf "Speichern" klicken, um den Text zu speichern, oder den Text in die Zwischenablage setzen, werden einige statistische Angaben zur Anzahl eingelesener Zeichen angezeigt. (Wenn Sie den erkannten Text an eine Zielanwendung gesendet haben, wird keine Statistik angezeigt.)



Nach dem OCR-Vorgang wird das gescannte Bild mit den definierten Bereichen zur weiteren Verarbeitung erneut so lange angezeigt, bis Sie eine weitere Seite scannen.

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.

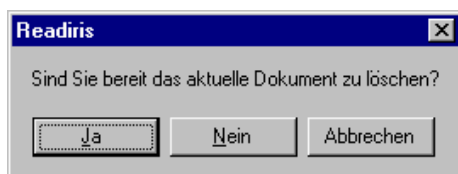
Wie aber speichern Sie den Text von zusätzlichen Seiten? Mit anderen Worten: Wie verarbeiten Sie mehrseitige Dokumente? Eigentlich ist es ganz einfach: Sie erkennen die Seiten nacheinander, aktivieren jedoch die Option "Zufügen", wenn Sie das Scanresultat in derselben Datei speichern möchten! Vergessen Sie nicht,



auch das Font-Wörterbuch in den Zufügen-Modus zu setzen, um das Schriftarten-Training mühelos fortzusetzen.



Sobald Sie Seiten innerhalb eines Dokuments scannen (oder Bilddateien öffnen), 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 "Datei"-Menü.



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.

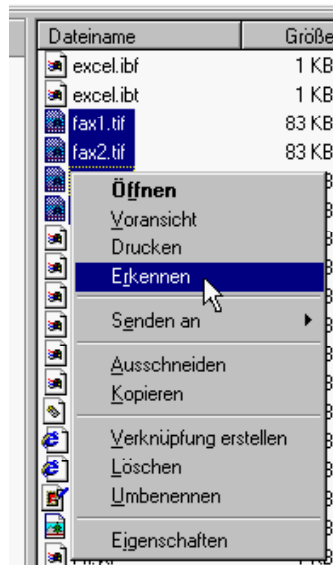


Legen Sie die Seiten Ihres Dokuments in den automatischen Dokumenteneinzug und beginnen Sie den Scanvorgang: alle Seiten werden gescannt, bis der Dokumenteneinzug leer ist.

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 Umschalttaste 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.



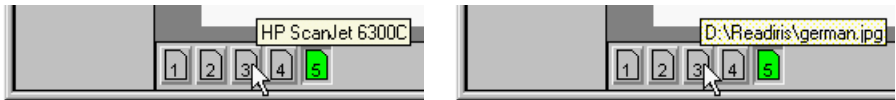
Sie können sogar mehrere vorgescannte Bilder aus dem Windows Explorer in das Readiris-Fenster *ziehen*! Auch hier gilt: alle Bilder, die Sie in das Readiris-Fenster ziehen, werden an das aktuelle Dokument angehängt, bis Sie den Befehl "Neues Dokument" wählen.

Sobald Seiten verarbeitet werden, erscheint unter dem Bild die **Seiten-Toolbar**. In dieser Toolbar sind die einzelnen Seiten des Dokuments zu sehen. Klicken Sie rechts, um auf die Seiten-Befehle zuzugreifen.



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

Die Seiten-Toolbar bietet einen Tooltip: Wenn der Benutzer den Mauszeiger auf einem Seitensymbol ruhen lässt, wird angezeigt, welches Bild in den Speicher geladen worden ist. (Wenn ein mehrseitiges Bild geöffnet worden ist, gibt es natürlich nur eine Datei für alle Bilder.) Wenn der Benutzer mehrseitige Dokumente *scannt*, gibt der Tooltip lediglich das Scanner-Modell an.



Laden Sie das Beispielfeld MULTIPAG.TIF und beginnen Sie den Erkennungsvorgang. Die einzelnen Seiten werden nacheinander angezeigt, die Readiris-Titelleiste und der OCR-Fortschrittsbalken zeigt die Seitennummer an.

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.

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.

Navigation: Um *zu einer Seite zu gehen*, klicken Sie auf das Symbol in der Page-Toolbar oder führen Sie den Mauszeiger über das Symbol, klicken Sie rechts, um das Kontextmenü zu öffnen, und wählen Sie den Befehl "Seite Wählen". Um zur vorigen Seite zu gehen, können Sie die Taste PageUp drücken, um zur nächsten Seite zu gehen, können Sie die Taste PageDn drücken. Oder gebrauchen Sie die entsprechende Befehle aus dem Menü "Ansicht".



Bearbeitung des Dokuments: Um eine Seite aus dem Dokument zu *löschen*, führen Sie den Mauszeiger über das Symbol der Seite, klicken Sie rechts und wählen Sie den Befehl "Seite Löschen". Um eine Seite im Dokument *nach oben zu verschieben*, verwenden Sie den Befehl "Seite nach oben Verschieben", um eine Seite *nach unten zu verschieben*, wählen Sie den Befehl "Seite nach unten Verschieben".





EIN NEUES DOKUMENT STARTEN

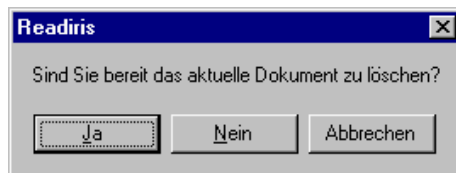
Sie können den Befehl "Neues Dokument" im "Datei"-Menü wählen, um das aktuelle Dokument zu schließen.



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 Bilddateien hinzufügen, beispielsweise Faxe. 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.



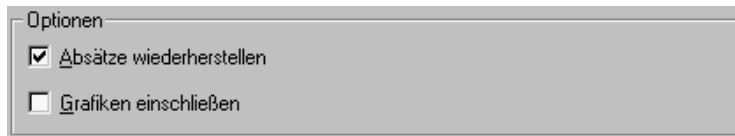
DIE AUSSICHT DER TEXTAUSGABE

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.

Sie können das Erscheinungsbild 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 und nicht am Ende einer Zeile. Die Funktion "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 einen Leerraum am Ende des Satzes zugefügt. Somit wurde das Aussehen des Textes beeinflusst und ein natürlicher Textfluß erstellt.

Ansonsten hätte Readiris einen feste Zeilenwechselmarke am Ende der Zeile und nicht am Ende des Absatzes gesetzt.

(Die Funktion "Format" enthält noch mehr Optionen. Mehr darüber später.)

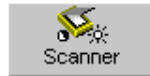
SCANNEREINSTELLUNG

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

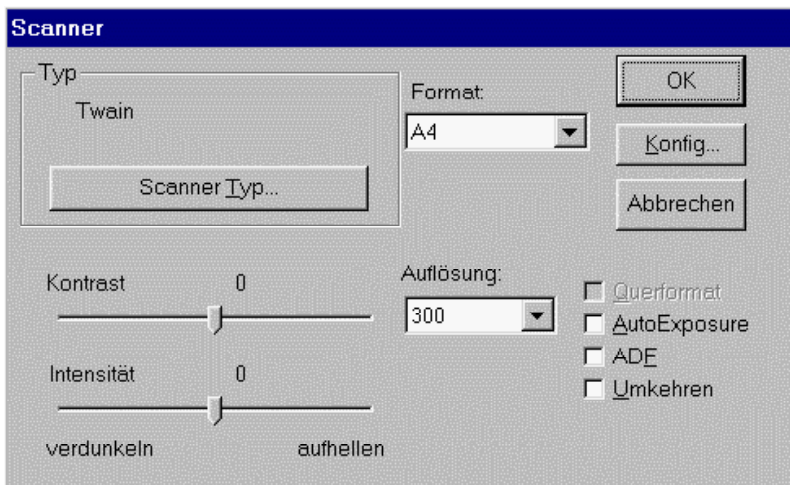


Wenn Ihr Scanner zusammen mit Ihrer Readiris Software lizenziert wurde brauchen Sie hier wahrscheinlich nichts mehr zu machen. Ihr Scanner ist dann oft bereits unter Readiris installiert.

Klicken Sie auf das Symbol "Scanner" in der Hauptfunktionsleiste.



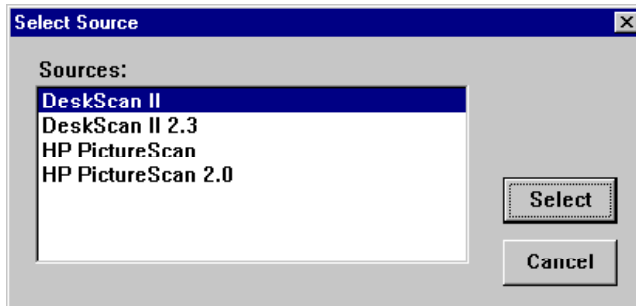
Klicken Sie auf das Symbol "Scannertyp" um das **Scannermodell** zu bestimmen.



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

Das Symbol "Konfig." 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 das Symbol

"Konfig." klicken. (Sie können ebenfalls den Befehl "Quelle Auswählen" im Menü "Datei" anklicken.)



Ihr Scanner kann ebenfalls mit einer **Scan-Plattform**, wie HP JetSuite oder HP Document Assistant arbeiten. Abhängig von Ihrer Software-Version gibt es hierzu spezielle Handbücher auf der Readiris CD-ROM.

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. Die Kontrasteinstellungen sind nur bei bestimmten Scannern vorhanden. Die ausgewählten Kontrastwerte werden numerisch vorgestellt.

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

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 zu Farbtonunterschieden kommen, die in Schwarz-weißbildern 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 Schwarz-weißmodus 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.

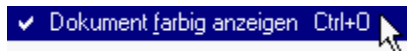
YASUMITSU SHIGETA, 35, has invested in more than 70 Web or mobile Net-based ventures in Japan and the U.S., including Tumbleweed Communications and Phone.com. Shigeta is also developing new businesses that take advantage of the growth of the Internet and mobile communications.

Readiris erzeugt eine Schwarzweiß-Fassung jedes Graustufen- und Farbbilds. Dank intelligenter ProgrammROUTINEN können selbst schwierige Fälle gelöst werden - und so wird ein “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 schwarz-weiß zu sehen, deaktivieren Sie die Option "Dokument Farbig Anzeigen" im "Ansicht"-Menü.



Bei Schwarz-weißbildern sind Textfenster gelb, Grafiken blau und Tabellenfenster violett. Bei Graustufen- und Farbbildern sind Textfenster gelb und Tabellenfenster violett. Grafikfenster haben einen blauen Rand, aber keine "Füllfarbe" - es wäre nicht zweckmäßig, die tatsächlichen Farbtöne der Grafik zu überlagern.

Unabhängig vom Scanmodus wählen Sie für normale Anwendungen eine **Auflösung** von 300 dpi, für eine kleinere Schriftart (unter 10 Punkte) 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 Punkten (0,21 bis 2,54 cm.) zu lesen.

6 Punkte

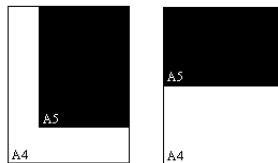
72 Punkte



Readiris kann auch **Initialen** ("Drop Letters") erkennen. "Initialen" sind große Buchstaben, die über mehreren Zeilen gehen - sie können natürlich nie größer sein als 72 Punkte!

Readiris erkennt
"Initialen" - das
sind große
Buchstaben die über
mehreren Zeilen gehen.

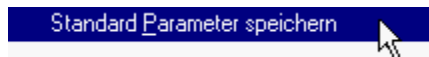
Durch Aktivierung der Option "Querformat" geben Sie an, daß die gewählte Seitenorientierung breit ist anstatt schmal. Die Seitenorientierung kann nur auf reduzierten Seitenformaten angewendet werden: mit einem A4 Flachbettscanner können A5 Texte (halb so groß) gescannt werden im Hoch- und Querformat.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



Schließlich ermöglicht es die Option "Umkehren", im Schwarz-weiß-Scanmodus **Negativbilder** zu erzeugen. Aktivieren Sie diese Option zur Verarbeitung von ganzen Seiten mit weißem Text auf einem schwarzem Hintergrund.

STANDARDEINSTELLUNGEN SPEICHERN

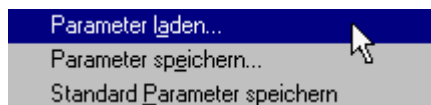
Wenn alle Scanparameter richtig eingestellt wurden, klicken Sie auf den Befehl "Standard Parameter Speichern" im Menü "Datei". Die laufenden Einstellungen werden als Standard **Einstellungen** gespeichert. Bei künftigen Aufrufen arbeitet Readiris mit den Einstellungen weiter.



Die Standardeinstellungen enthalten mehr als nur die Scanparameter: Sie können auch bestimmen, ob Sie die interaktive Lernfunktion benutzen werden, welches Benutzerwörterbuch (und Modus) benutzt wird, welche Sprache und welche Schriftart - z.B. Normalschrift, Proportionalschrift - aktiv sind, welche Art der Ausgabe, welche Layout-Optionen verwendet werden - z.B. kontinuierlichen Text nach WordPad senden - usw. Kurz, *alle* operative **Einstellungen** werden in den Readiris Standardaufstellungen gespeichert.

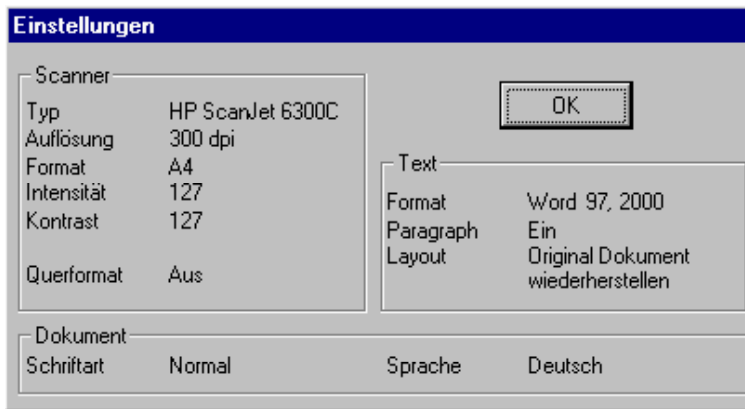
DAS SPEICHERN VON VORDEFINIERTEN EINSTELLUNGEN

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



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

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

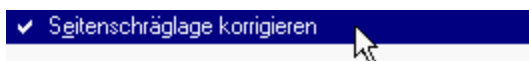


DOKUMENTE SCANNEN

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

Die Option "Seitenschräglage Korrigieren" im Menü "Voreinstellungen"-Menü legt fest, ob die Schräglage von Seiten, die in Schräglage eingescannt wurden, korrigiert wird - eine geringfügige Zeilenschräglage wird ignoriert. Diese Option ist standardmäßig aktiviert.

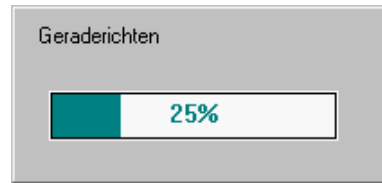
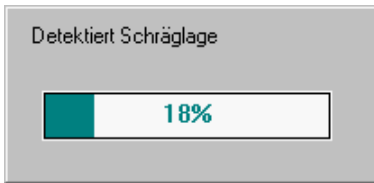


Wurde diese Option fälschlicherweise deaktiviert, verwenden Sie die Schaltfläche "Schräglage Seite Korrigieren" in der Toolbar (oder den Befehl

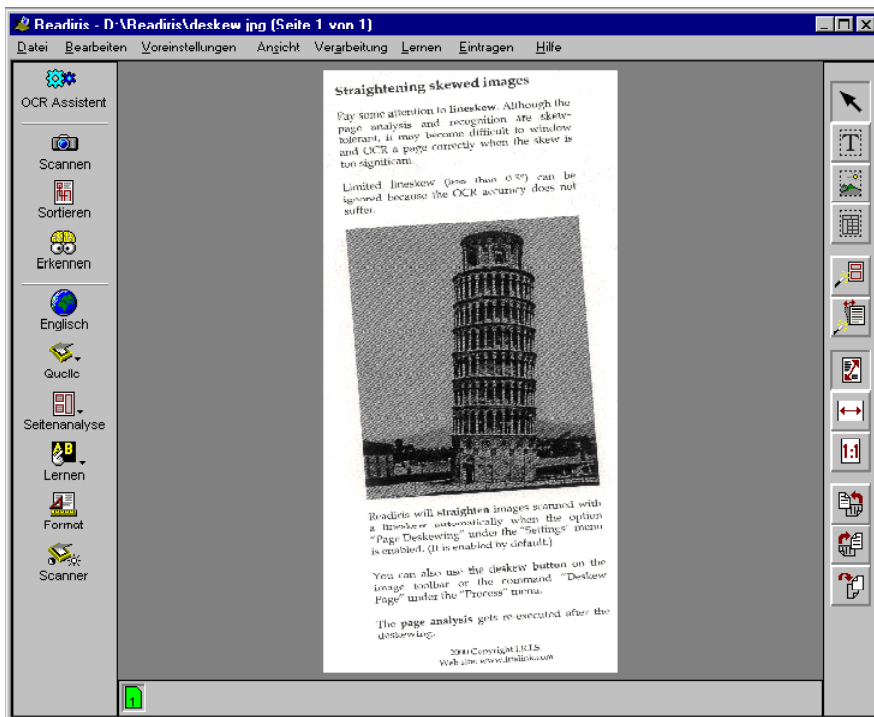
"Schräglage Seite Korrigieren" im Menü "Verarbeitung"), um Seiten zu begradigen, die schräg eingescannt wurden.



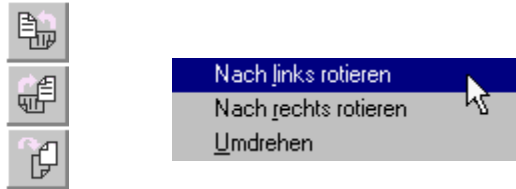
Die Korrektur der Schräglage dauert einige Sekunden. Das Bild wird analysiert, um den Winkel festzustellen, das Farb- oder Graustufenbild und dessen Schwarz-weiß-Fassung werden korrigiert und die Seitenanalyse wird erneut ausgeführt.



Sie können das Bild DESKEW.JPG im Readiris-Ordner verwenden, um diese Funktion auszuprobieren. Deaktivieren Sie zunächst die Option "Seitenschräglage Korrigieren", andernfalls wird die Schräglage automatisch korrigiert, bevor das Bild angezeigt wird!



Gegebenenfalls müssen Sie auch die **Seitenorientierung** korrigieren. Verwenden Sie hierzu die Drehung-Schaltflächen der Toolbar. (Die entsprechenden Befehle stehen auch im Menü "Ansicht".)



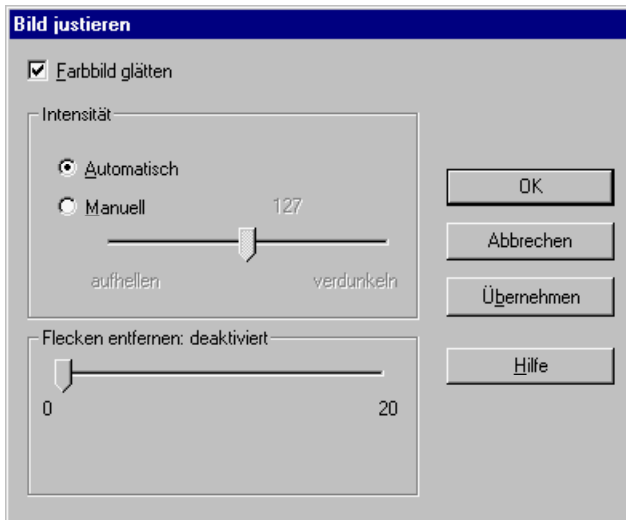
Es gibt drei Drehrichtungen: nach links, nach rechts und Drehung um 180 %.

JUSTIEREN DER GESCANNTEN BILDER

Wie bereits gesagt, konvertieren leistungsfähige Routinen Farb- und Graustufenbilder automatisch in Schwarz-weißbilder. 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.

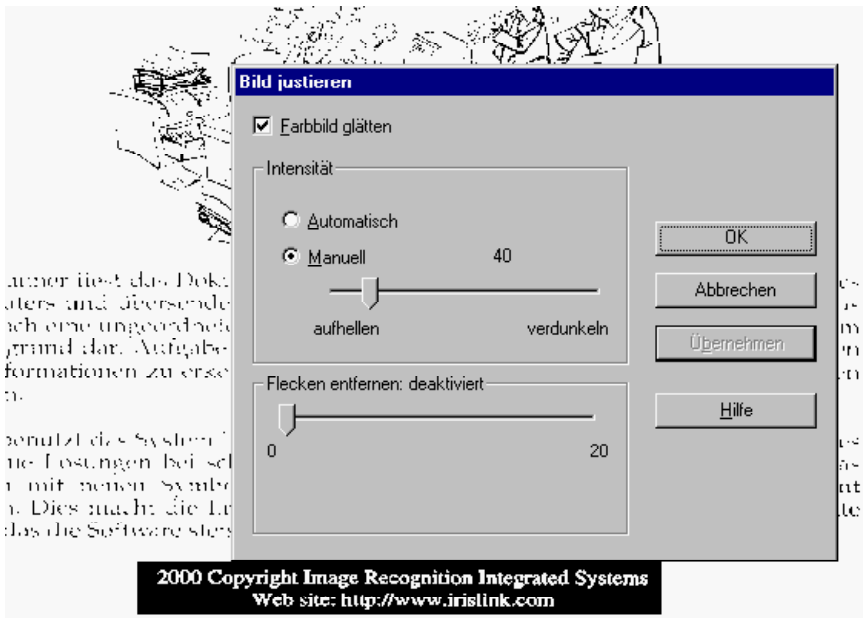


Die Option "Farbbild glätten" macht Graustufen- und Farbbilder homogener, indem relative Intensitätsunterschiede "geglättet" werden. Dies führt zu einem schärferen Kontrast zwischen dem Vordergrund (dem Text) und dem Hintergrund (Farbe, Verzierungen, usw.). Wir empfehlen, diese Option immer aktiviert zu lassen.

Helligkeit. Unter "Intensität" versteht man eigentlich den Schwarz-Weiß-Schwellenwert. Die Einstellung "Automatisch" legt den Schwarz-Weiß-Schwellenwert automatisch fest. Legen Sie bei Bedarf einen anderen Schwellenwert fest, indem Sie das Schwarz-weißbild dunkler oder heller machen. Wenn Sie das Bild dunkler machen, werden in der Schwarz-weiß-Fassung mehr Pixel schwarz gefärbt; wenn Sie das Bild aufhellen, werden in der Schwarz-weiß-Fassung mehr Pixel weiß gefärbt.

Bitte beachten Sie, dass keine Bildanpassung ausgeführt wird, bevor Sie auf "Übernehmen" klicken! Wenn Sie auf "OK" klicken, wird die Anpassung vorgenommen *und* das Fenster geschlossen. Hier ein Beispiel, in dem wir das

Schwarz-weißbild wesentlich aufgehellt haben - dabei stand die OCR-Präzision allerdings nicht im Vordergrund!



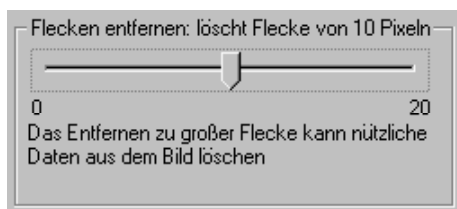
Die ersten beiden Optionen betreffen Farb- und Graustufenbilder, die letzte Option, "Flecken Entfernen", ausschließlich Schwarz-weißbilder. Als "Flecken entfernen" bezeichnet man die Entfernung von "Parasiten-Pixels" aus Schwarz-weißbildern.



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

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

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



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

LASSEN SIE DEN OCR ASSISTENTEN FÜR SIE ARBEITEN

Beginnen wir nun, Dokumente zu erfassen. Nachdem der Scanner eingestellt wurde, wollen wir damit anfangen Dokumente zu erfassen. Damit Sie nicht durch alle Parameter müssen, benutzen wir den **OCR Assistenten**, die bequemste Art, Seiten zu erkennen.

Klicken Sie auf "OCR Assistent" in der Hauptfunktionsleiste (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").

SEITEN AUTOMATISCH ERKENNEN

Der OCR Assistent ist eine halb-automatische Weise für die Seitenerkennung. Readiris hat auch eine *voll*-automatische Erkennung: Sie müssen nicht mehr Schritt

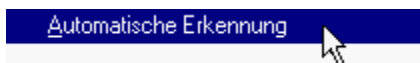


für Schritt durch die Parameter gehen, wir werden die Seite jetzt sofort - ohne jegliche Verzögerung - erkennen.

Um den Befehl "OCR Assistent" auf der Hauptfunktionsleiste durch den Befehl "Auto" zu ersetzen, deaktivieren Sie die Option "Assistent auf Funktionsleiste Setzen" im Menü "Voreinstellungen".



Wählen Sie den Scanner mit "Quelle" als Bildquelle und klicken Sie auf "Auto" oder wählen Sie den Befehl "Automatische OCR" im Menü "Verarbeitung".



Automatisches OCR bedeutet, daß die Seite nacheinander gescannt wird, durch die Seitenanalyse oder durch eine Fenstervorlage definiert und erkannt wird ohne interaktives Lernen. Sie brauchen nur die Befehle zu initiieren, den Rest übernimmt Readiris.

READIRIS ERSTELLT DAS LAYOUT

Automatisches Erkennen darf 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 Schriftzeichen (serif und sans-serif, proportional und fest, normal und verdichtet) werden wie im Originaldokument wiedererkannt. Die Schriftgröße und die Schriftauszeichnung (fett, kursiv und unterstrichen) werden ebenfalls wiedererkannt im Speicher. Der Randausgleich und die Einzüge (links, zentriert, rechts und justiert) von jedem Textblock werden wiederhergestellt. Die Position der Spalten, Text- und Grafikblöcke erscheinen wie im 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 Fenster AUTOFORM.JPG im Readiris Verzeichnis.



Readiris - D:\Readiris\autoform.jpg (Seite 1 von 1)

Datei Bearbeiten Einstellungen Ansicht Vergrößerung Lernen Eintragen Hilfe

Autoformatting

The aim of "autoformatting" is to recreate a facsimile copy of the original document.

The OCR process does more than just recognize your text, it can format it for you too!

The various levels of formatting are: creating body text, retaining the word and paragraph formatting and creating a facsimile copy.

In a way, text recognition is becoming more and more page recognition or document recognition...

Creating body text means no formatting is applied; you get a continuous, running text. All formatting, if any, is done afterwards by the user.

Whether your OCR software reformats the recognized text or not is up to the user. You can perform OCR because you just need the text, in which case you will edit and format it yourself, and you can recreate the source document, including its formatting.

If you retain the word and paragraph formatting, the font type, size and typestyle are maintained across the recognition. The justification of the paragraphs is also detected. However, no graphics are captured and the columns aren't recreated - the paragraph just follows each other etc.

"Autoformatting" recreates a facsimile copy of the original document: the text blocks, graphics and tables are recreated in the same place and the word and paragraph formatting are maintained across the recognition.

Cell 1A	Inmate
Cell 2A	Warden
Cell 3A	\$100,000

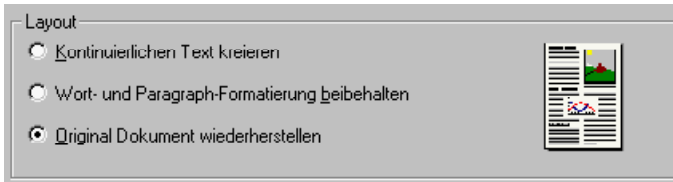
As a result, you get a text copy of your source document, be it a compact and editable text file, no longer a scanned image of your document.

File Copy all Copy as Rich Text Format Integration Options

OCR Assistent

- Scannen
- Sortieren
- Erkennen
- Englisch
- Quelle
- Seitenanalyse
- Lernen
- Format
- Scanner

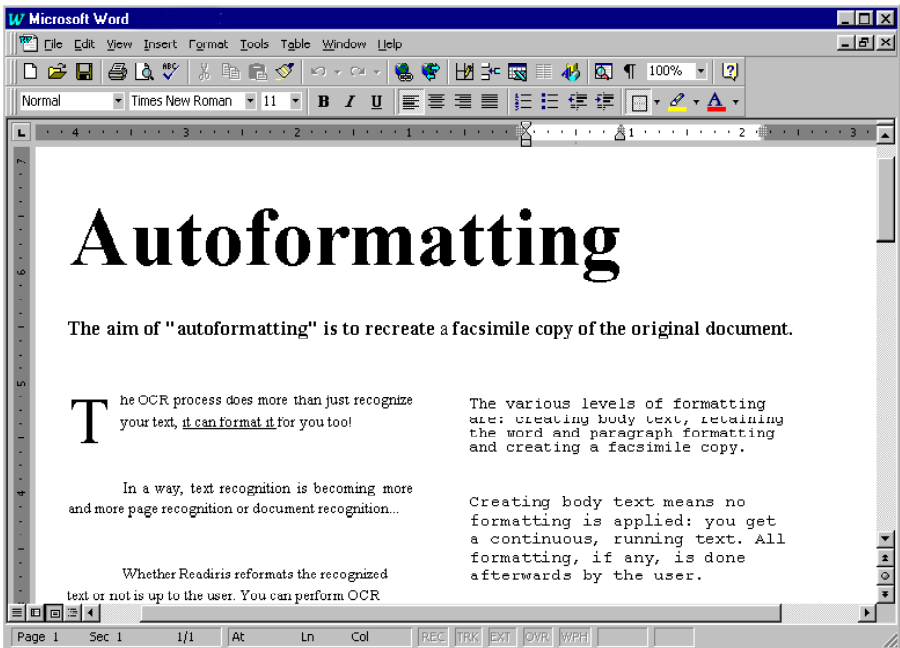
Klicken Sie auf das Symbol "Format" auf der Hauptfunktionsleiste und wählen Sie, das OCR Ergebnis in Microsoft Word zu senden, oder wählen Sie "RTF" (Rich Text Format) oder das Word (DOC) Format. Wählen Sie weiter "Original Dokument Wiederherstellen" als Layoutoption. (Die Option "Absätze Wiederherstellen" wird automatisch aktiviert, um den Zeilenumbruch zu aktivieren.)



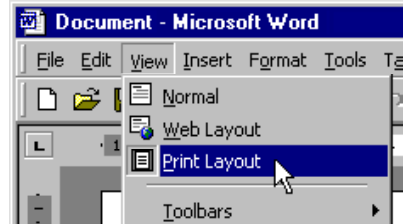
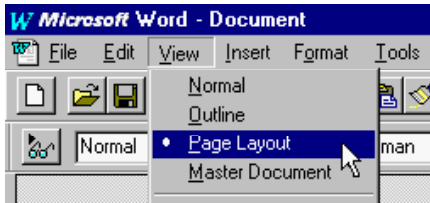
Die Rekonstruktion des Layouts ist nur für Word (DOC), RTF Format und für Dokumente, die direkt in Microsoft Word gesendet wurden, möglich. Dabei sind Word (DOC) und RTF Formate häufig benutzte Textformate, die von jeder beliebigen Textverarbeitung geöffnet werden können.

Andere, "begrenzte" Formate so wie ASCII, ANSI, usw. unterstützen *keine* fortgesetzte Formatierung und sollen deshalb nicht im Autoformat-Modus gewählt werden. Ebenso ist die Anwendung WordPad ein beschränkt einsetzbarer Textverarbeiter; WordPad ist in der Lage die Word (DOC) und RTF-Dateien zu öffnen, aber die Formatierungs-codes wie Textränder, Einzüge usw. werden ignoriert.

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



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

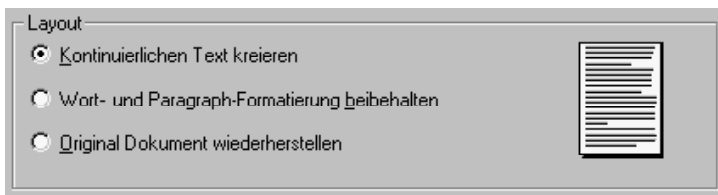


Also, Readiris kann Texte erkennen und formatieren. OCR ist nicht nur eine Texterkennung. OCR wird mehr und mehr eine **Seiten-** oder **Dokumentenerkennung!**

TEXTFORMATIERUNG, TEIL 2

Es gibt auch zwei andere Layoutoptionen: "Kontinuierlichen Text Kreieren" und "Wort- und Paragraph-Formatierung Beibehalten".

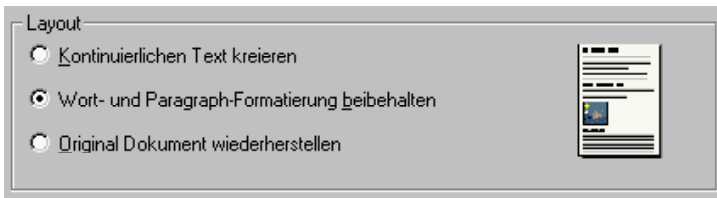
Sowie das Symbol rechts zeigt, bedeutet "Kontinuierlichen Text Kreieren" daß ein nicht formatierter, **kontinuierlicher Text** erstellt wird. Der Text wird ohne Formatierung gespeichert. Diese Option ist nur dann sinnvoll, wenn Sie nur den Text brauchen, aber nicht die Formatierung.



(Um asiatische Texte erstellen zu können muß ein kontinuierlicher Text erstellt werden - die Funktion "AutoFormatierung" kann nicht angewendet werden, die Wort- und Absatzformatierung kann nicht zurückgesetzt werden. Die Zahl der Formatdateien wird eingeschränkt.)



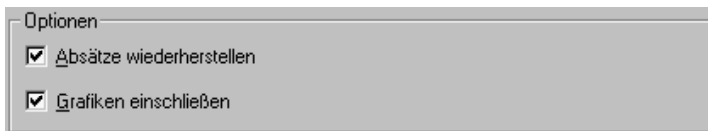
Die Option "Wort- und Paragraph-Formatierung Beibehalten" ist ein Mittelding. Die **Wortformatierung** - Schriftart (serif - sans serif, proportional - fest, normal - verdichtet), Schriftgröße und Schriftauszeichnung (fett, kursiv und unterstrichen) - werden beibehalten. Das Gleiche gilt für die **Absatzformatierung** - der Randausgleich und die Einzüge (links, zentriert, rechts und justiert) werden ebenfalls beibehalten.



Verwechseln Sie diese Format-Option nicht mit "komplettem" Autoformatierung: hier werden die Absätze hintereinander gestellt, es werden keine Spalten rekreiert oder verschiedene Zonen in bestimmten Positionen kopiert.

GRAFIKEN GETRENNT SPEICHERN

In unserem Beispiel war die Grafik im erkannten Text enthalten; Das hängt aber wiederum von der Option "Grafiken Einschliessen" ab. Grafiken innerhalb eines Textes zu speichern ist nur mit "voller" Autoformatierung möglich-, nicht mit "armen" Textformaten wie ASCII, ANSI usw.



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

Wie? Zeichnen Sie eine Grafikzone um die Illustrationen, Cartoons usw., die Sie brauchen herum. Grafikfenster manuell zu zeichnen, funktioniert genauso

wie mit Text- oder Tabellenfenstern, wählen Sie jetzt einfach nur das "Grafikfenster" Symbol.

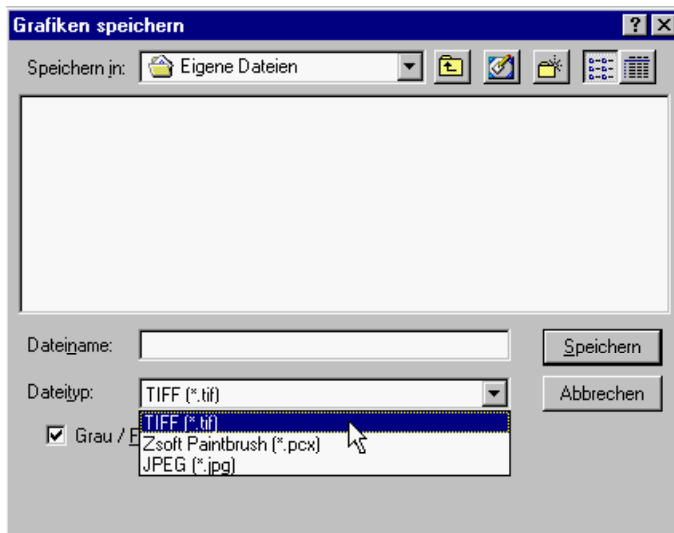


(Wie bei anderen Fensterarten gibt Ihnen das Quickinfo des Grafikfenster Symbols an, wieviele Grafikfenster vorhanden sind. Und Sie können auch Alt eingeben während Sie ein Fenster anwählen, um seinen Typ und seine Koordinaten zu sehen.)

Selektieren Sie danach den Befehl "Grafiken Speichern" im Menü "Datei".



Geben Sie einen Dateinamen ein.



Legen Sie fest, welches Grafikdateiformat Sie benutzen möchten. Wählen Sie ein Format, das von Ihrer Fotoretuschiersoftware unterstützt wird. Verfügbar

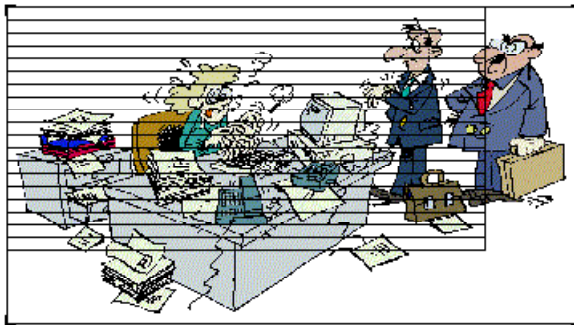


sind: JPEG, TIFF und Paintbrush (PCX). (Paintbrush-Dateien können immer mit den Windows-Programmen Paint und Imaging geöffnet werden.) Aktivieren Sie diese Option "Grau/Farbe", um die Grafiken als Farb- oder Graustufengrafik zu speichern. Diese Option ist standardmäßig aktiviert.



Die Grafiken werden in einer separaten Datei gespeichert. Sie brauchen sich nicht auf einer einzelnen Grafik zu beschränken. Bei der Erstellung von mehreren Grafikfenstern werden diese "gestapelt" in einer Datei. (Der Befehl "Abschneiden" im Fotoretuschierprogramm trennt die Dateien wieder von einander.)

Seiten von weniger als 1 mm. sind nicht erlaubt. Bitmap-Formate in der Größe enthalten ohnehin kaum Information. Fenster mit einer unregelmäßigen, nicht-rechteckigen Form bekommen eine rechteckige Form. Readiris bedeckt die größte Fläche, basiert auf den größten Koordinaten. Das Beispiel hier unten zeigt ein Grafikfenster (hat schwarzen Rand), daß vervollständigt wurde von einem schattierten Bereich.



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 wird mit den vorgescannten Bildern gemacht!

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





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

Jetzt können Sie die Diskette als Quelle selektieren und die Bilddatei mit dem Symbol "Öffnen" (oder mit dem entsprechenden Befehl im Menü "Verarbeitung") öffnen. (Wenn Sie den Befehl "Öffnen" im Menü "Datei" benutzen, brauchen Sie die Quelldatei nicht auf dem neuesten Stand zu bringen.)

Da Readiris Farb-, Graustufen- und Schwarz-weißbilder gleichermaßen unterstützt, öffnet Readiris JPEG-Bilder, Paintbrush (PCX)-Bilder, TIFF-Bilder (nicht-komprimiert, Packbits, Gruppe 3 und Gruppe 4 komprimiert), mehrseitige TIFF-Bilder und Windows Bitmaps (BMP).

Diese Funktion ist sehr praktisch, wenn Sie Ihre **Faxe** zur weiteren Bearbeitung in Textdateien konvertieren möchten! Um bessere OCR-Ergebnisse zu erzielen, überzeugen Sie den Faxabsender davon, Faxe in "feiner" Qualität mit einer Resolution von 200 dpi zu schicken.

Vergessen Sie nicht, dass Sie auf Bilder im Windows Explorer rechts klicken können 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, und schon werden die Dateien geöffnet.

TABELLENERKENNUNG

Bis jetzt haben wir uns mit der Erkennung von Faxen und Texten, 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 TABLE.JPG im Readirisverzeichnis.

Tatsächlich enthält das Bild TABLE.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 Ihre 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				
	Average access time (msec)	CPU utilization (%)	Video clip playbacks (frames dropped)	Sequential read 16 KB (K bps)
CD-ROM				
Digital Versatile Disk				
CD-ROM 4x speed	442	4.2	10 8	419
CD-ROM 12x speed	137	20.9	3 4	1,586
CD-ROM 24x speed	80	58.2	3 2	2,258
CD-ROM 32x speed	60	72.1	3 2	2,987
DVD	58	78.9	3 2	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 contains numeric characters, enable the numeric reading mode with the "Language" button on the main toolbar for increased accuracy.

Führen Sie die Erkennung durch, nachdem Sie die Layout Option "Wort- und Paragraph Formatierung Beibehalten" oder "Original Dokument Wiederherstellen" aktiviert haben, und die Tabelle wird wiederhergestellt. Öffnen Sie Ihre Textverarbeitung (falls Sie die erkannte Tabelle nicht direkt in Word gesendet haben), um das Ergebnis zu sehen. (Sie konnten natürlich auch Text Paragraphen in Ihre Textdatei miteinschliessen.)



Microsoft Word - table

File Edit View Insert Format Tools Table Window Help

Normal Times New Roman 11 B I U

Tip of the Day: To hyphenate a document as you type, click Hyphenation on the Tools menu.

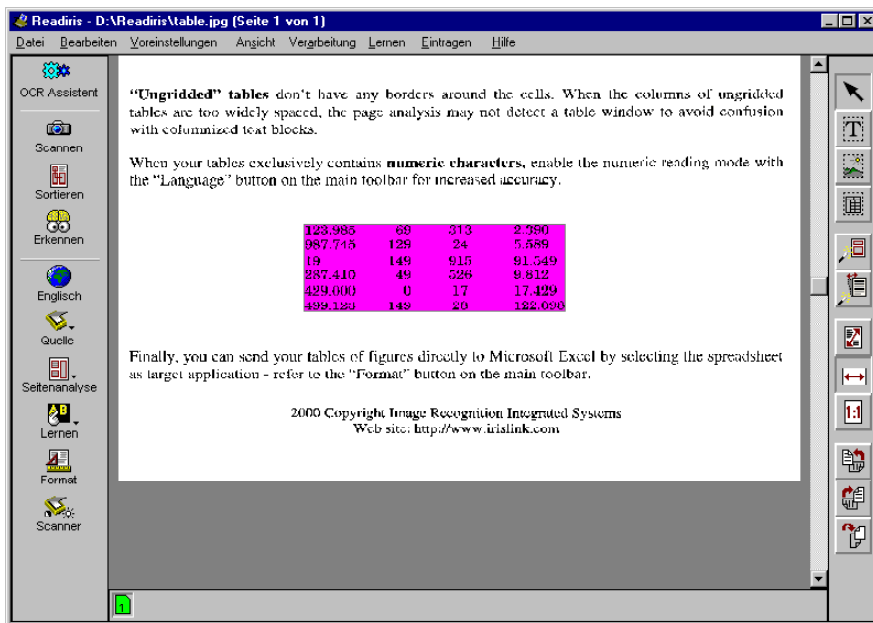
¾ pt Clear

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

Page 1 Sec 1 1/1 At 8.5cm Ln 13 Col 1 REC MRK EXT DVR WPH

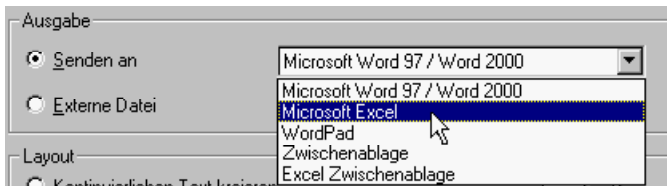
Und jetzt für “nicht umrahmte” Tabellen - sie haben keine Ränder um die Zellen herum. Beachten Sie, dass die Seitenanalyse die tabelle trotzdem erkennt!



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



Diesmal werden wir das OCR Ergebnis direkt in die Spreadsheet Microsoft Excel senden. Wir wählen Excel als gewünschte Anwendung im Befehl "Format".



Die Spreadsheet startet automatisch und die Ergebnisse sehen so aus: Die typische Tabellen Struktur mit Reihen und Spalten ist rekreiert, und sie ist zur direkten Weiterverarbeitung bereit.

	A	B	C	D	E
1	123.985	69	313	2.390	
2	987.745	129	24	5.589	
3	19	149	915	91.549	
4	287.410	49	526	9.812	
5	429.000	0	17	17.429	
6	499.123	149	28	122.098	

Beachten Sie, dass Sie Tabellen auch in die Zwischenablage senden können - also die Option "Senden an - Excel Zwischenablage" - und sie in einer externen Datei speichern können - also das Dateiformat Excel. (CSV bedeutet "Komma delimitierte" Daten.)

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 das Tabellenfenster Symbol in der Bildfunktionsleiste und gehen Sie wie üblich vor; das Quickinfo des Befehls gibt wie üblich die Anzahl bestehender Tabellenfenster an.

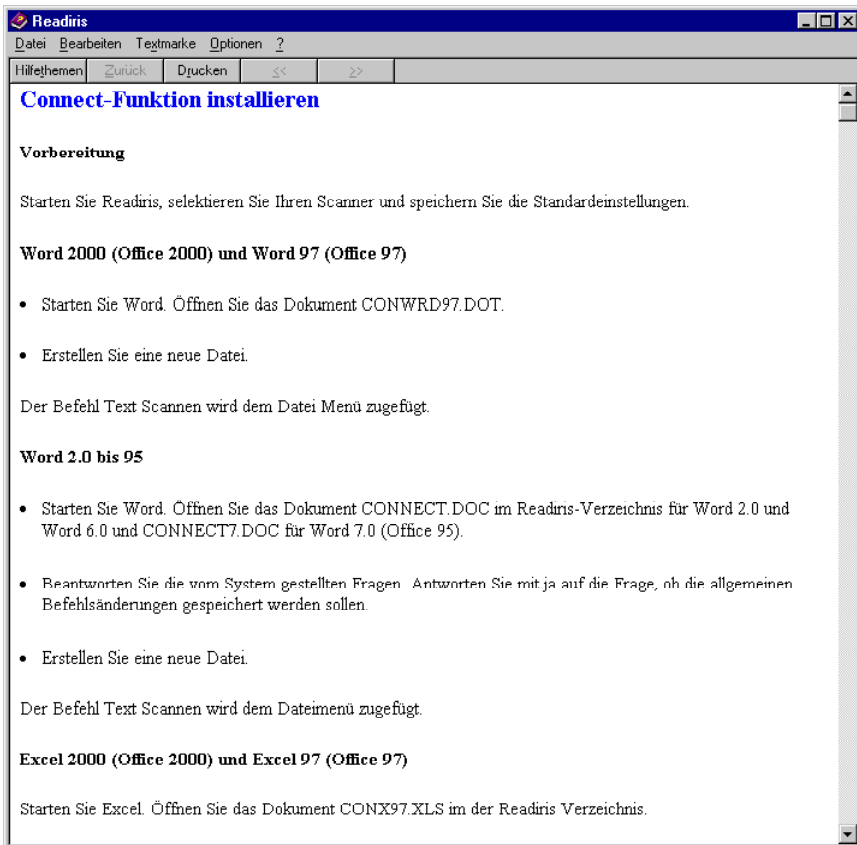


DIE “CONNECT” FUNKTION, EINE WEITERE DIREKTE VERBINDUNG

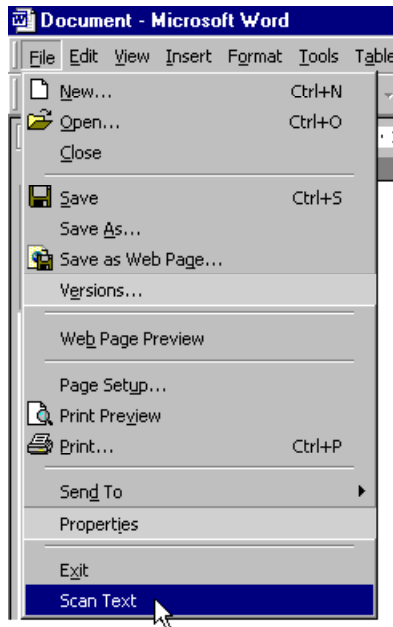
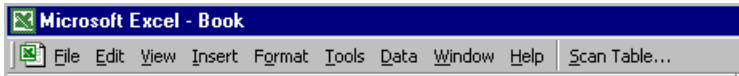
Es gibt noch einen anderen Weg, um Zeichentabellen zu scannen, nämlich direkt über den Spreadsheet.

Ähnlich wie bei der "Senden nach" Funktion, die die Ergebnisse automatisch exportiert, bietet die “Connect” Funktion eine direkte OCR Verbindung zwischen Ihrem Scanner und Ihrer Windows Anwendung. Der einzige Unterschied ist die “umgekehrte” Vorgehensweise: Sie erkennen Ihre Dokumente nicht in Readiris, um die Ergebnisse dann in die gewünschte Anwendung zu exportieren, sondern Readiris wird direkt von Ihrer Anwendung aus aufgerufen!

Bei der **Installation** gibt es einige Schritte die manuell ausgeführt werden müssen. Lesen Sie hierzu die Datei CONNREAD.DOC in Ihrem Readirisverzeichnis. (Diese “Lese Mich“-Datei erscheint automatisch am Ende des Installationsprogramms.)



Bei korrekter Installation fügt das “Connect“-Feature Ihrer Anwendung einen neuen Befehl zu. In den unten gezeigten Beispielen wird der Befehl "Text Scannen" dem Word Menü "Datei" zugefügt und der Befehl "Tabelle Scannen" wird der Excel 97 Menüleiste zugefügt.



So funktioniert es: Readiris darf *nicht* im Hintergrund laufen. Um Readiris zu starten, klicken Sie den neuen Befehl in Ihrem Applikationsmenü an. Die “Connect“-Schnittstelle wird Ihnen einige Fragen stellen bevor Readiris startet. Gehen Sie vor wie gewöhnlich und verlassen Sie die “Connect“-Schnittstelle nach Beendigung. Die Leseergebnisse werden der derzeitigen Cursorposition hinzugefügt.



ON-LINE 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 On-Line-Hilfe-System zu konsultieren wenn Sie zusätzliche Informationen über Readiris brauchen.

Selektieren Sie das Menü "Hilfe". Der Befehl "Hilfethemen" oder den Short-cut F1 gestattet Ihnen den Zugriff auf vielen On-Line-Hilfe-Informationen.

