

ReadirisTMPRO⁹

GEBRUIKERS- HANDLEIDING



Readiris Pro

© 2004 I.R.I.S. Alle rechten voorbehouden

OCR technologie door I.R.I.S.

Connectionist, AutoFormat en linguïstische technologie door I.R.I.S.

© 2004 I.R.I.S. Alle rechten voorbehouden

BESPAAR TIJD, VERTIK HET TE HERTIKKEN!

Gelukwensen bij de aanschaf van Readiris. Deze software zal u ongetwijfeld belangrijke hulp bieden bij het hernemen van uw teksten, tabellen en grafisch materiaal!

Hoe efficiënt computers ook mogen zijn, u moet de informatie altijd wel nog eerst invoeren. Als u ooit een rapport van 15 pagina's of een grote tabel heeft moeten overtypen, dan weet u ongetwijfeld hoe vervelend en tijdrovend dat kan zijn. Gebruik echter dit geavanceerde OCR-pakket om teksten automatisch in uw toepassingen in te voeren en u zult een ongekend niveau van efficiëntie en comfort halen!

Scan een gedrukt of getyped document, geef de te hernemen zones aan - of laat het systeem deze klus voor u klaren - en voer de karakterherkenning uit. Documenten die meerdere bladzijden bevatten worden in één beweging verwerkt. Enkele eenvoudige muisklikken overtreffen urenlang werk naarmate Readiris uw papieren documenten omzet in bewerkbare computerbestanden; dit gebeurt wel 40 keer sneller dan handmatige invoer!

De OCR wizard gidst u doorheen de herkenning: antwoord op enkele eenvoudige vragen en u haalt snelle resultaten met Readiris. U kunt de herkenningsresultaten tevens rechtstreeks naar uw tekstverwerker en uw spreadsheet sturen. Om faxen te herkennen en PDF documenten om te zetten kunt u de beeldbestanden van de Windows Verkenner naar het Readiris venster slepen. Of klik met de rechter muisknop op een beeld om het ogenblikkelijk met Readiris te openen.

Readiris herkent tabellen en hercreëert die als werkbladen of als tabelobjecten binnen uw tekstverwerker; uw cijfermateriaal is onmiddellijk klaar voor verdere verwerking met uw rekenblad.

Gebaseerd op de Connectionist technologie van I.R.I.S., heeft Readiris het beste wat OCR te bieden heeft. Omnifont herkenning wordt aangevuld met zelflerende technieken die afgeleid zijn van neurale netwerken. Het systeem kan



door de contextuele analyse nieuwe karakters aanleren: taalkundige kennis over lettergrepen en woorden verbetert de OCR-prestatie.

Readiris ondersteunt wel 107 talen: alle Amerikaanse en Europese talen worden herkend, inclusief de Centraal-Europese talen, de Baltische talen, het Grieks en de Cyrillische (“Russische”) talen. (Optioneel kunt vier Aziatische talen inlezen - Japans, vereenvoudigd en traditioneel Chinees en Koreaans.) Readiris verwerkt ook gemengde alfabetten: de software detecteert “Westerse” woorden die opduiken in Griekse, Cyrillische en Aziatische documenten - vele niet vertaalbare eigennamen, merknamen enz. zijn immers geschreven in het Latijnse alfabet.

Readiris maakt gebruik van taalkundige informatie *tijdens* de herkenningsfase, en niet nadien. Dit laat Readiris toe documenten van allerlei soort te herkennen met uiterste nauwkeurigheid, en dat geldt ook voor documenten met een lage drukkwaliteit, faxen en dot matrix afdrukken. Readiris herkent eveneens slecht gescande en slecht gekopieerde documenten met te lichte of te donkere lettertekens. Verbonden lettertekens (“ligaturen”) worden goed gelezen, gefragmenteerde vormen zoals dot matrix symbolen worden opnieuw samengesteld.

Controle door de gebruiker in “popup” stijl markeert de twijfelachtige karakters en verhoogt de precisie van het systeem. Alle oplossingen die door de gebruiker zijn bevestigd worden opgeslagen, wat de snelheid en betrouwbaarheid bij verder gebruik nogmaals verhoogt. Readiris gebruiken betekent de software telkens intelligenter maken! Deze krachtige leerfunctie laat u toe Readiris te trainen op allerhande speciale lettertekens zoals de wiskundige symbolen, “dingbats” en de vervormde, beschadigde lettertypes die u in echte documenten kunt terugvinden.

Om uw productiviteit nog meer te verhogen, herkent Readiris niet alleen uw teksten, maar hij kan ze ook in uw plaats *opmaken*! Maak gebruik van “autoformatting” en Readiris hercreëert een facsimile kopie van het gescande document: de woord-, alinea- en paginaformatting van het oorspronkelijke document blijven behouden.

Er worden gelijkaardige lettertypes, puntgroottes en stijlen gebruikt zoals in het brondocument. De positie van de kolommen, tekstblokken en grafieken is identiek aan het oorspronkelijke document. En aangezien Readiris moeiteloos grijswaarden en kleurscans ondersteunt kunt u tegelijk al uw grafisch materiaal

hernemen - het maakt niets uit of het zwart-wit illustraties, zwart-wit foto's of kleurfoto's betreft. Wanneer een document tabellen bevat, zal Readiris die "herorganiseren" in heuse cellen en de celgrenzen herscheppen naar het voorbeeld van de oorspronkelijke tabel.

Met andere woorden, met Readiris kunt u een conforme kopie van uw documenten archiveren, zij het als editeerbare, compacte tekstbestanden en niet als gescande beelden! Verscheidene niveau's van formattering zijn beschikbaar, de gebruiker maakt zijn keuze.

Readiris ondersteunt een breed gamma aan populaire scanners: talrijke vlakbedscanners, sheetfed scanners, "all-in-one" machines of "MFPs" ("multi-functional peripherals") en digitale camera's worden ondersteund. Readiris ondersteunt tevens de Twain scanner standaard en enkele scanplatformen.

INHOUDSTAFEL

Bespaar tijd, vertik het te hertikken!	III
Inhoudstafel	V
Auteursrechten	VII

Hoofdstuk 1: Installatie

Systeemvereisten	1-1
De Readiris software installeren	1-1
De Readiris software verwijderen	1-4
Het Readiris "uninstall" programma	1-5
De Windows (un)installwizard	1-5
Software-opties installeren	1-6
Verwante toepassingen installeren	1-8
Geïnstalleerde bestanden	1-10
"Lees Mij" bestand en documentatie	1-10
Scannerstuurbestanden	1-10
Registreer om mee te spreken!	1-10
Beroep doen op productondersteuning	1-12



Hoofdstuk 2: Een rondleiding

Readiris opstarten	2-1
De eerste sessie is uniek	2-2
De Readiris gebruikersinterface verkennen	2-3
Van start gaan met een eerste oefening	2-6
Inzoomen op beelden	2-10
Eén, een gescand beeld analyseren	2-13
Anderhalf, vensters sorteren	2-15
Twee, beelden manueel bevensteren	2-17
Drie, venstersjablonen opslaan	2-22
Met Readiris kunt u de hele wereld rond!	2-23
Readiris verandert tussendoor van taal	2-29
De kenmerken van het document aangeven	2-32
Readiris wordt telkens intelligenter!	2-34
Leren	2-37
Niet Leren	2-37
Wissen	2-38
Annuleren	2-38
Einde	2-39
Afbreken	2-39
De rol van woordenboeken	2-39
Tekst direct naar een toepassing sturen	2-42
Het resultaat in een tekstbestand opslaan	2-46
Universele bestanden aanmaken... ..	2-50
... of ze lezen!	2-55
Meerdere pagina's herkennen	2-57
Meerbladige documenten bewerken	2-65
Met een nieuw document beginnen	2-67
Tekstzones herkennen	2-68
De uitvoer opmaken	2-70
Uw scanner instellen	2-71
Geef uw scans wat kleur mee!	2-73
Andere apparaten, andere resolutie	2-76
Standaard instelling bewaren	2-80
Specifieke instellingen bewaren	2-81
Documenten scannen	2-82
De gescande beelden aanpassen	2-85

Laat de wizard voor u werken	2-89
Readiris herschept de originele lay-out	2-91
Kolommen aub, geen kaders!	2-95
Tekstformattering, deel 2	2-97
Tekst meerdere keren uitvoeren	2-98
Figuren apart opslaan	2-99
Faxen herkennen en uitgestelde herkenning	2-101
Tabellen herkennen	2-103
On-line hulp raadplegen	2-108

AUTEURSRECHTEN

De Readiris software werd ontworpen en ontwikkeld door I.R.I.S. De OCR, Connectionist, AutoFormat en linguïstische technologie werd ontwikkeld door I.R.I.S. I.R.I.S. bezit de auteursrechten op de Readiris software, de OCR technologie, de linguïstische technologie, het on-line hulpsysteem en deze handleiding.

AutoFormat, Cardiris, Connectionist, de I.R.I.S. linguïstische technologie, het I.R.I.S. logo en Readiris zijn handelsmerken van I.R.I.S.

De XML parser werd ontwikkeld door Apache. Dit product bevat software die werd ontwikkeld door de Apache Software Foundation (www.apache.org).

Acrobat en Reader zijn (geregistreerde) handelsmerken van Adobe. AsianBridge is een handelsmerk van TwinBridge. AsianSuite is een handelsmerk van UnionWay. Excel, Windows en Word zijn geregistreerde handelsmerken van Microsoft. Intel is een geregistreerd handelsmerk van Intel.



Hoofdstuk 1

INSTALLATIE

Dit hoofdstuk bespreekt de systeemvereisten en de installatie van de Readiris software.

SYSTEEMVEREISTEN

Dit zijn de minimale systeemvereisten om Readiris te gebruiken:

- ☐ een 486 Intel PC of compatibele computer. Een Pentium PC is aanbevolen.
- ☐ 32 MB RAM. 64 MB RAM is aanbevolen om beelden in grijswaarden en kleuren te verwerken.
- ☐ 110 MB vrije ruimte op de harde schijf. 95 MB schijfruimte volstaat wanneer u de voorbeeldbestanden op de CD-ROM laat staan.
- ☐ het besturingssysteem Windows XP, Windows ME, Windows 2000, Windows 98 of Windows NT 4.0.

Noteer dat sommige **scannerstuurbestanden** mogelijk niet werken onder de laatste versie(s) van Windows! Raadpleeg de documentatie die bij uw scanner geleverd werd om na te gaan welke besturingssystemen ondersteund worden.

DE READIRIS SOFTWARE INSTALLEREN

De Readiris software wordt uitsluitend geleverd op een **zelfstartende CD-ROM**. Om de software te installeren laadt u eenvoudig de CD-ROM in uw CD-

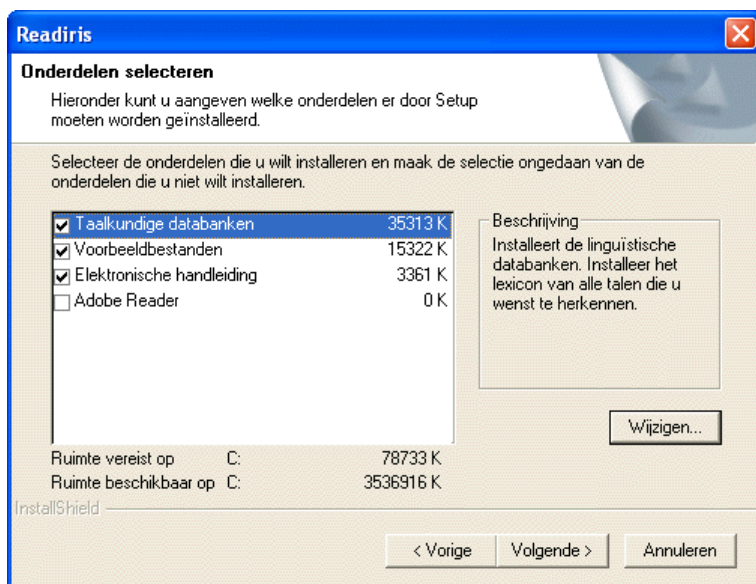


ROM station en wacht u tot het installatieprogramma begint te lopen. Voer de instructies uit die op het scherm verschijnen.

Mocht de installatie niet opstarten wanneer u de CD-ROM in de CD-ROM lezer laadt, start dan het installatieprogramma MENU.EXE op om de software te installeren.

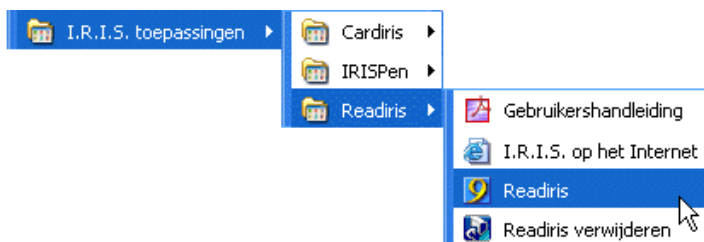
Gebruikers van Windows XP, Windows 2000 en Windows NT moeten nagaan of zij over de nodige **toegangsrechten** beschikken - contacteer zo nodig de systeembeheerder.

Er komen enkele opties kijken bij de installatie. Installeer in elk geval de **taalkundige databanken** van alle talen die u wenst in te lezen. Standaard worden alle woordenboeken geïnstalleerd. Wij raden u tevens aan de **voorbeeldbestanden** te installeren; die worden namelijk in deze handleiding gebruikt als lesbestanden.



Installeer tevens, mocht dat nodig blijken, de Adobe Reader software die toelaat de softwaredocumentatie te raadplegen. De **elektronische handleiding** wordt standaard op uw harde schijf gekopieerd, maar u kunt die ook op de CD-ROM laten staan.

Het installatieprogramma creëert automatisch het submenu "I.R.I.S. toepassingen - Readiris" onder het menu "Programma's".



Datzelfde geldt voor een **snelkoppeling** naar Readiris op het Windows bureaublad. Eén en ander betekent dat u de Readiris software onmiddellijk vanaf uw bureaublad kunt opstarten.

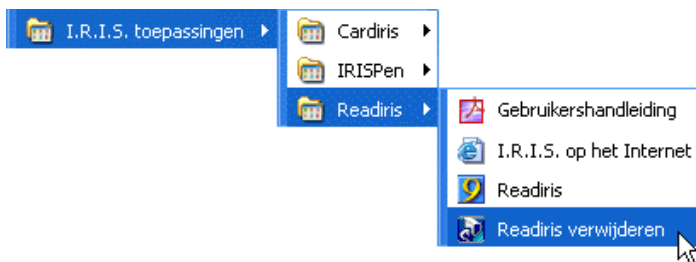


DE READIRIS SOFTWARE VERWIJDEREN

Er zijn slechts twee juiste methodes om de Readiris software te verwijderen: gebruik maken van de “uninstall” toepassing en gebruik maken van de Windows (un)installassistent. Het wordt ten stelligste afgeraden om de Readiris software of zijn softwaremodules te verwijderen door de bestanden van de toepassing manueel te verwijderen.

Het Readiris “uninstall” programma

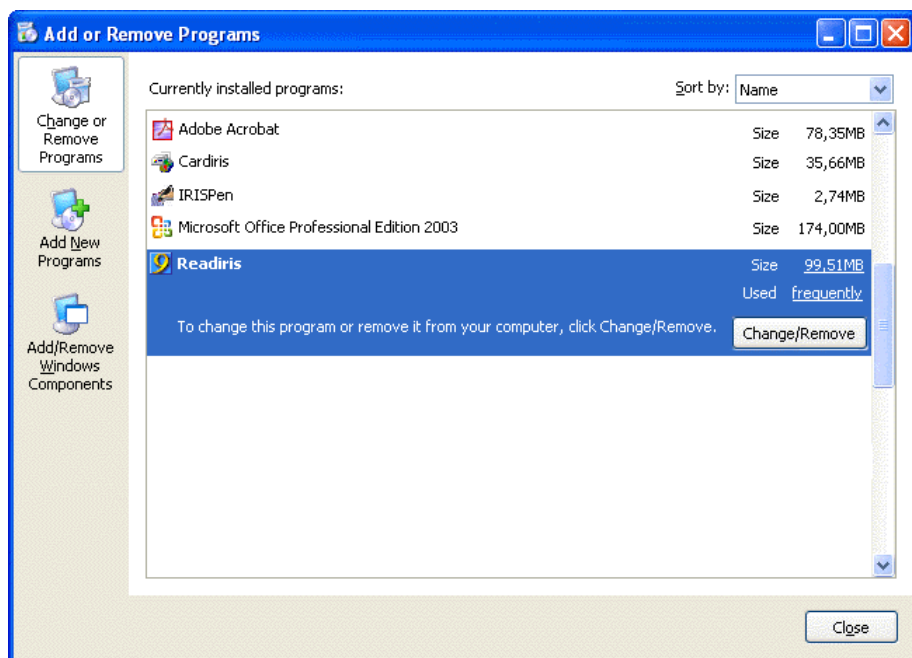
Selecteer "Readiris Verwijderen" in het submenu "I.R.I.S. toepassingen - Readiris" om de Readiris “uninstall”-toepassing op te starten. Volg de instructies die op het scherm verschijnen.



De Windows (un)installwizard

Voer de volgende stappen uit om gebruik te maken van de Windows (un)installassistent.

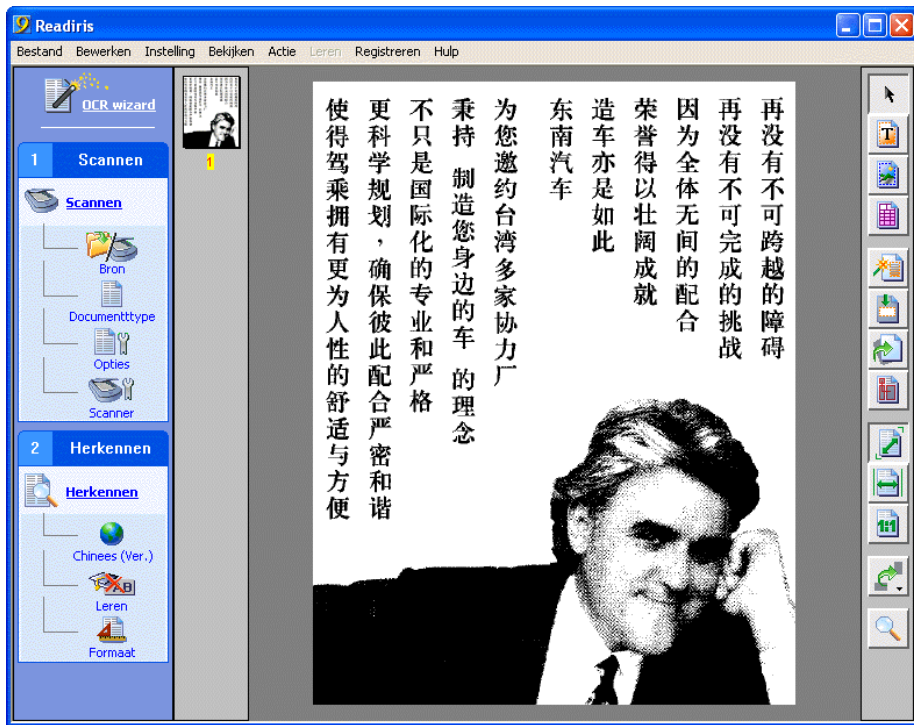
- ☐ Klik op "Instellingen" onder het "Start" menu van Windows en ga naar het "Configuratiescherm".
- ☐ Klik op het pictogram "Toepassingen Toevoegen/Wissen" in het "Configuratiescherm".



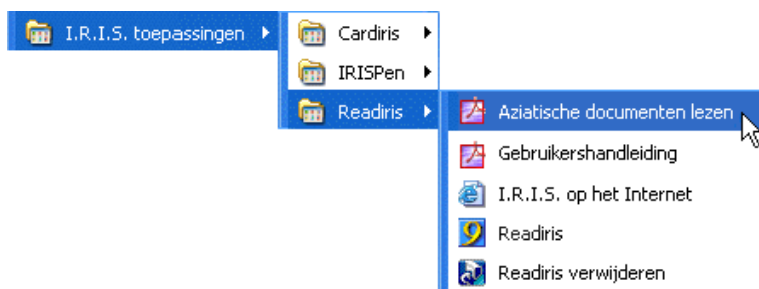
- ☐ Volg de instructies die op het scherm verschijnen om de Readiris software te verwijderen.

SOFTWARE-OPTIES INSTALLEREN

Er is één enkele software-optie beschikbaar voor Readiris: het complement “Aziatische OCR-module”. Deze extra software laat u toe het Japans, het traditioneel en vereenvoudigd Chinees en het Koreaans te herkennen. Ook die software wordt op een zelfstartende CD-ROM geleverd.



Wanneer u deze optie installeert, wordt er specifieke documentatie toegevoegd die uitlegt hoe u Aziatische documenten kunt herkennen.

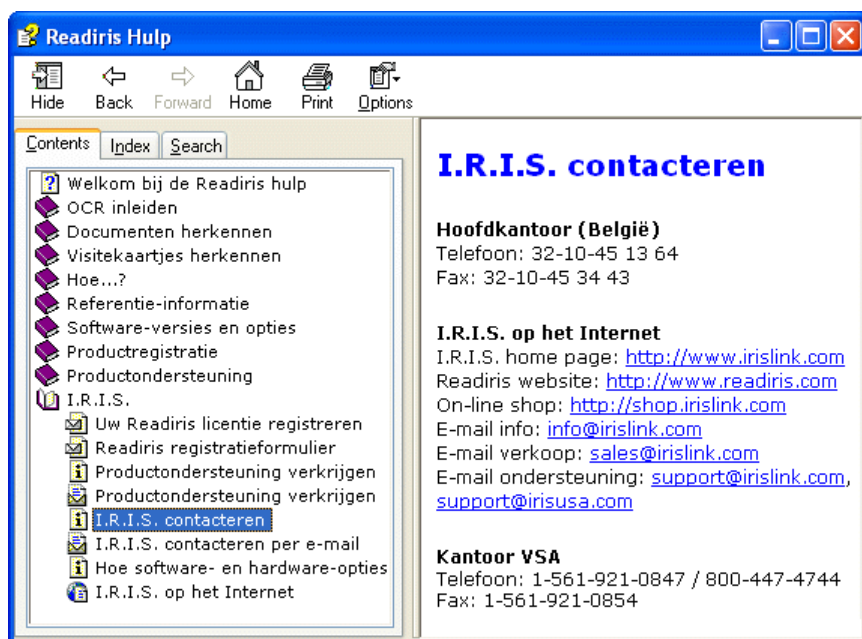


VERWANTE TOEPASSINGEN INSTALLEREN

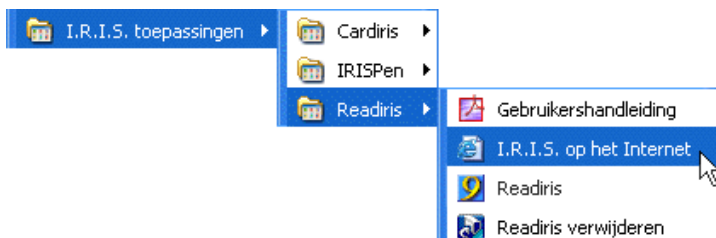
Naargelang de softwareversie die u aanschafte, kan Readiris vergezeld zijn van een evaluatie-versie van het aanverwante product Cardiris, een **organizer voor visitekaartjes**.

Indien dit gratis software-pakket meegeleverd is op de Readiris CD-ROM, wordt het net als Readiris geïnstalleerd door de zelfstartende CD-ROM in het station te laden en de instructies uit te voeren die op het scherm verschijnen.

Neem contact op met I.R.I.S. om meer te vernemen over dit complementair software-pakket; het commando "I.R.I.S. Contacteren" van het menu "Hulp" van Readiris vertelt u hoe u contact kunt opnemen met I.R.I.S.



Een pictogram in het submenu "I.R.I.S. toepassingen - Readiris" onder het menu "Programma's" brengt u rechtstreeks naar de I.R.I.S. **website**. Dat kan ook vanuit het startscherm van Readiris en aan de hand van het commando "I.R.I.S. op het Internet" onder het menu "Hulp" van Readiris.



GEÏNSTALLEERDE BESTANDEN

Het installatieprogramma maakt een map aan waarin de Readiris bestanden zich bevinden. Probeer nooit Readiris of sommige van zijn modules te verwijderen door handmatig programmabestanden te wissen. Gebruik hiervoor het Readiris “uninstall”-programma of de Windows (un)installwizard. Zie hierboven.

“Lees Mij” bestand en documentatie

README.HTM “Lees Mij” bestand (in HTML formaat)

MANUAL.PDF Gebruikershandleiding (in Adobe Acrobat formaat)

Scannerstuurbestanden

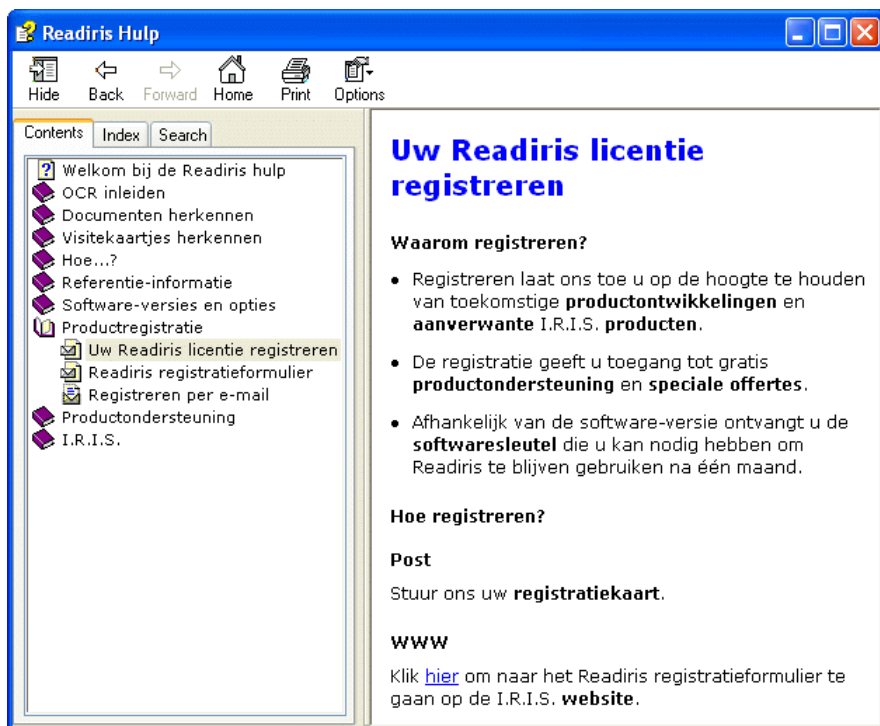
Aarzel niet uw scannerfabrikant of zijn vertegenwoordiger te contacteren mocht u problemen ondervinden met uw stuurbestand. Bij de meeste fabrikanten kunt u de laatste versies van de scannerstuurbestanden van de webstek halen.

REGISTREER OM MEE TE SPREKEN!

Vergeet niet uw Readiris licentie te registreren! Dit stelt ons in staat om u op de hoogte te houden van verdere productontwikkelingen en aanverwante I.R.I.S.

producten. De voordelen van de registratie, zoals gratis **productondersteuning** en **speciale offertes**, zijn exclusief voorbehouden aan geregisteerde gebruikers.

U kunt op vele verschillende manieren registreren: door uw registratiekaart op te sturen of door het elektronisch evenbeeld te faxen, door I.R.I.S. telefonisch te contacteren tijdens de kantooruren of door een registratieformulier in te vullen op de I.R.I.S. website!



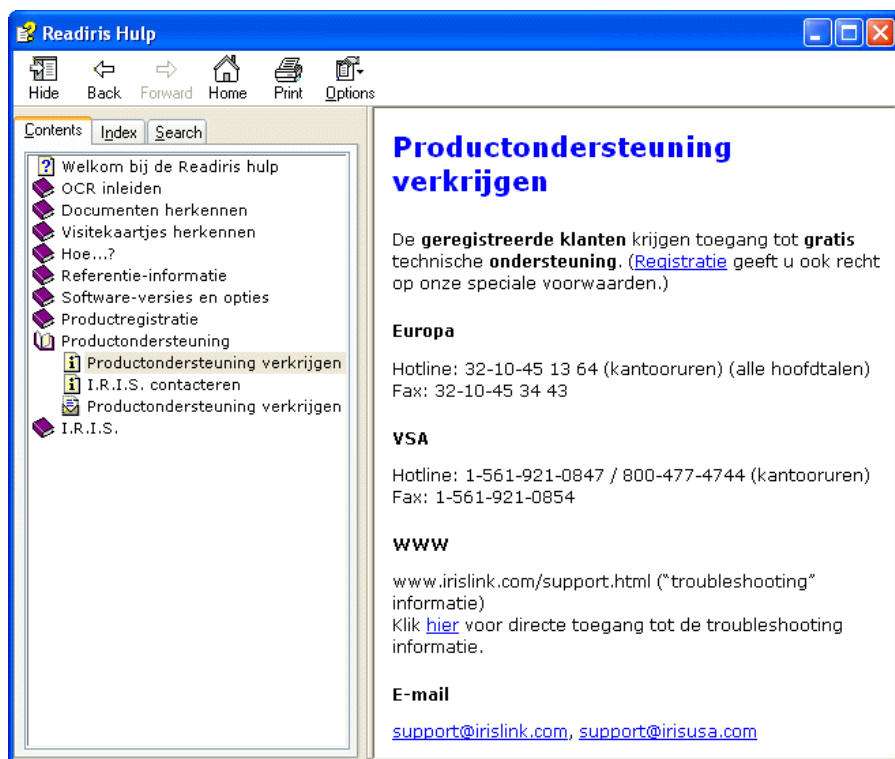
De Readiris **registratiewizard** die u vindt onder het menu "Registreren" van de Readiris software loodst u in luttele ogenblikken door het registratieproces.



Naargelang de softwareversie, ontvangt u in ruil de **softwaresleutel** die u kunt nodig hebben om de Readiris software na één maand te blijven gebruiken.

BEROEP DOEN OP PRODUCTONDERSTEUNING

Het commando "Productondersteuning" in het "Hulp" menu van Readiris beschrijft hoe u beroep kunt doen op de technische ondersteuning. Beschrijf duidelijk het probleem dat u ervaart en verschaf ons alle nuttige informatie betreffende Readiris, uw scanner en uw computersysteem.





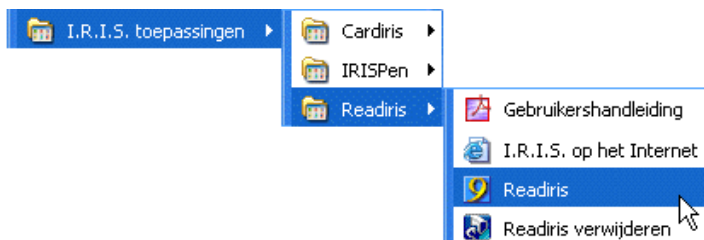
Hoofdstuk 2

EEN RONDLEIDING

Readiris is een geavanceerd OCR-pakket dat talrijke krachtige functies bevat. We zullen in dit hoofdstuk deze mogelijkheden doorlopen en vele tips betreffende het gebruik van Readiris toevoegen.

READIRIS OPSTARTEN

Klik op de Readiris applicatie in het submenu "I.R.I.S. Toepassingen - Readiris" of klik op de snelkoppeling van de Readiris applicatie op uw bureaublad.



Het Readiris startscherm en applicatievenster verschijnen op het scherm. Het startscherm geeft de versie en de auteursrechten van de Readiris software aan. Het verleent tevens rechtstreekse toegang tot de I.R.I.S. **website**: klik op de URL om de webstek van I.R.I.S. te bezoeken. Door met de muis elders te klikken verdwijnt het startscherm.



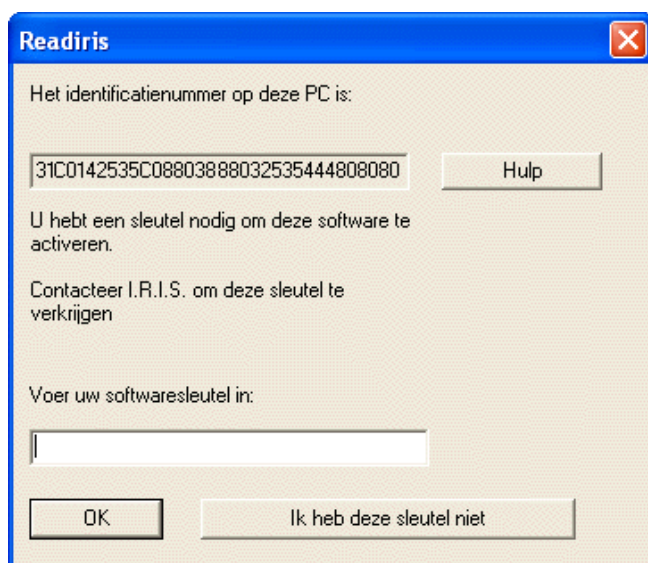
Het volgende venster betreft de OCR assistent; klik deze keer op "Annuleren" om die te doen verdwijnen.

DE EERSTE SESSIE IS UNIEK

Naargelang de versie van de software die u heeft aangekocht, kan de eerste sessie wat apart verlopen. U kunt namelijk uitgenodigd worden om uw licentie te registreren.

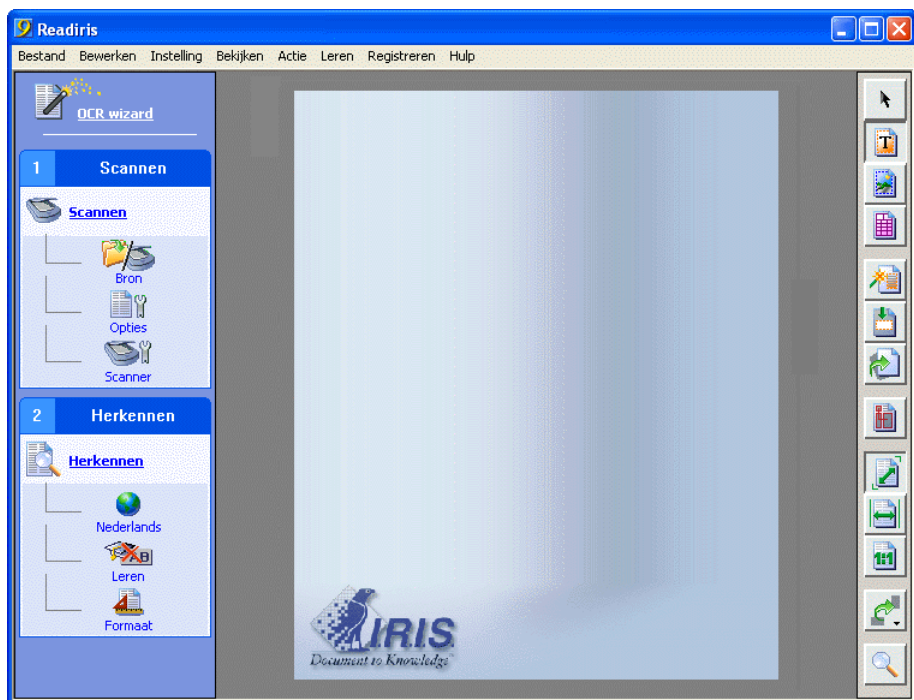
Als dat het geval is, is het gebruik van uw Readiris softwarelicentie beperkt tot één maand. Door te registreren ontvangt u van I.R.I.S. gratis de **softwaresleutel** die u toelaat om het programma na 30 dagen te blijven gebruiken.

U heeft uw **identificatienummer** nodig om de softwaresleutel te kunnen genereren; zorg ervoor dat u dit nummer onder uw bereik hebt of vermeldt wanneer u uw licentie registreert.



DE READIRIS GEBRUIKERSINTERFACE VERKENNEN

Het Readiris programmavenster bevat niet alleen de **commandomenu**'s, maar ook de knoppenbalken die snelle toegang geven tot alle frequent gebruikte bevelen. Aanvankelijk zijn sommige commando's grijs: ze betreffen de preview. Zolang er geen beeld geladen is, zijn ze niet voorhanden.



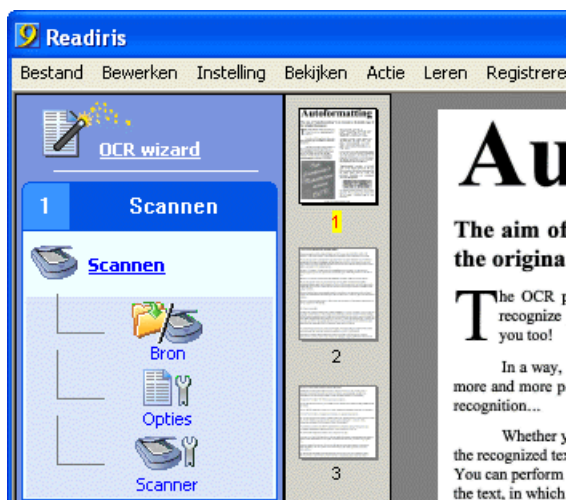
Dat geldt ook voor de **beeldknoppenbalk** aan de rechterkant van het programmavenster: deze balk bevat alle commando's die u tijdens de voorvertoning nodig heeft. De **algemene toolbalk** links geeft snelle toegang tot de frequente bevelen van algemene aard.

Om te vernemen welk bevel achter welke knop schuilt, kunt u uw muis een tijdje boven een knop houden: een **tooltip** zal u vertellen welke functie de knop vervult.



De **beeldzone** is de plaats waar de gescande beelden worden getoond. U kunt beeldbestanden naar deze zone (en naar het Readiris pictogram) slepen om ze te herkennen.

Van zodra een bladzijde wordt verwerkt, verschijnt een extra toolbalk, de **paginaknoppenbalk** links van de beeldzone: deze knoppenbalk stelt de verschillende bladzijden van het document voor en verleent met de rechter muisklik (het "Context" menu) toegang tot de bevelen die specifiek de bladzijden betreffen.



VAN START GAAN MET EEN EERSTE OEFENING

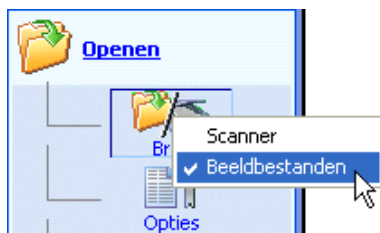
De beste manier om vertrouwd te raken met het gebruik van Readiris is ongetwijfeld de software gebruiken. Een aantal **vooraf gescande beelden** wordt bij de software geleverd; ze laten u toe om van start te gaan, zelfs als u nog geen scanner op uw PC aangesloten hebt. Laten we ermee van start gaan.

Met de knop "Bron" op de algemene knoppenbalk geeft u aan of u de scanner of een vooraf gescand beeld gebruikt als beeldbron.

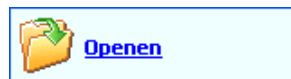
Readiris ondersteunt kleurbeelden, beelden in grijswaarden en zwart-wit beelden. Met Readiris kunt u Adobe Acrobat PDF documenten, JPEG beelden, Paintbrush (PCX) beelden, DCX faxen (een meerbladige versie van het Paintbrush formaat), (niet gecomprimeerde of LZW, PackBits, Groep 3 en Groep 4 gecomprimeerde) TIFF beelden, meerbladige TIFF beelden en Windows bitmaps (BMP) openen.

Deze mogelijkheid is bijzonder nuttig om **faxen** om te zetten in editeerbare tekstbestanden.

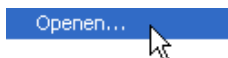
Om een vooraf gescand beeld te openen moet u "Beeldbestanden", en niet de scanner, als beeldbron kiezen met de knop "Bron".



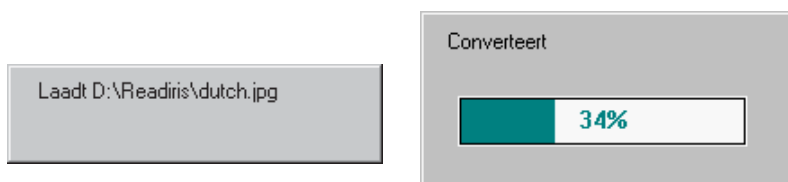
Daarna klikt u op knop "Openen". (Wanneer u de schijf als beeldbron kiest, wordt de knop "Scannen" vervangen door de knop "Openen" en wordt het corresponderende commando "Scannen" onder het menu "Actie" vervangen door het commando "Openen".)



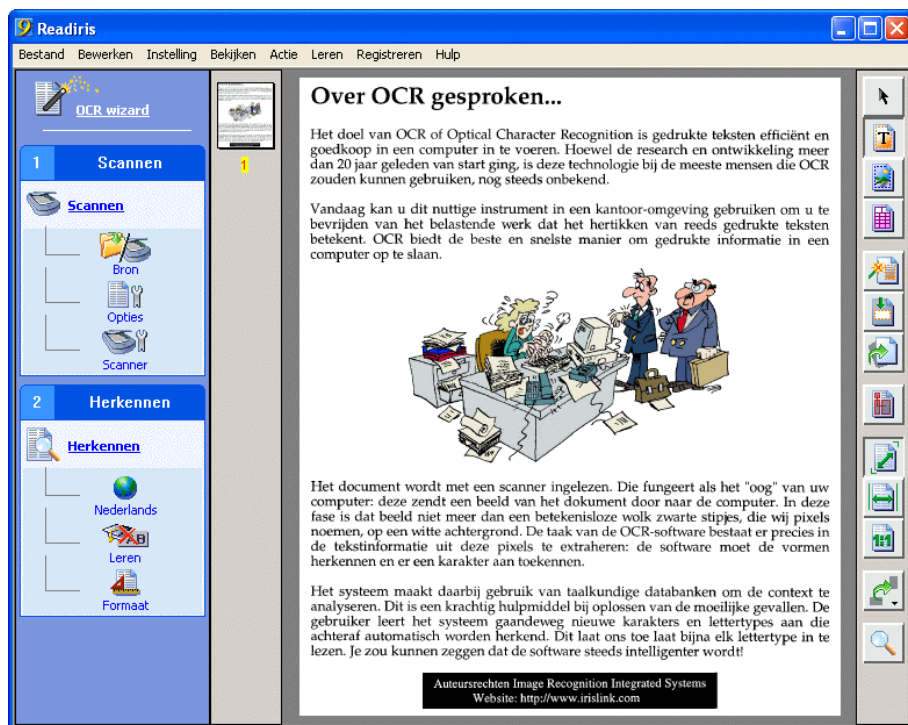
U kunt ook het commando "Openen" onder het menu "Bestand" gebruiken om een vooraf gescand beeld te openen - dat werkt zelfs wanneer uw scanner als beeldbron fungeert.



U wordt uitgenodigd om een beeldbestand te kiezen. Selecteer het bestand DUTCH.JPG in de Readiris map. Daar dit een kleurbeeld is, wordt het niet enkel van de harde schijf gelezen: er wordt tevens een zwart-wit versie aangemaakt voor het OCR-proces.



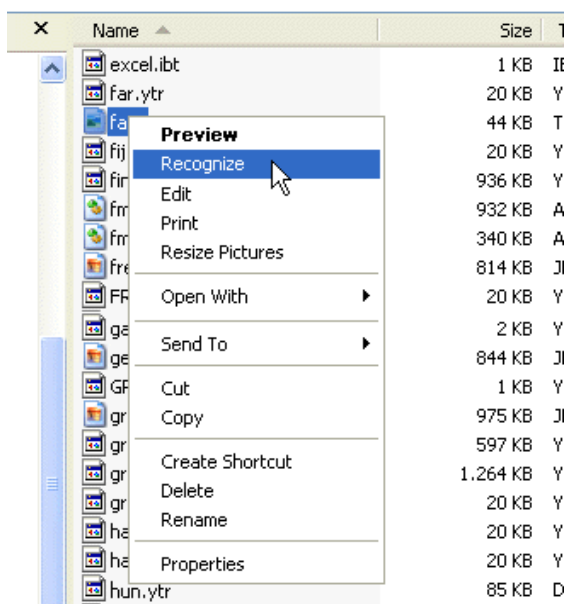
Tenslotte wordt het beeld getoond in de beeldzone. De paginaknoppenbalk geeft aan dat er één bladzijde in het geheugen geladen is.



Een derde manier om vooraf gescande beelden te openen is het gebruik van “**drag & drop**”: sleep beeldbestanden van de Windows Verkenner naar de beeldzone of naar het pictogram van Readiris en ze worden onmiddellijk geopend.



En er is nog een manier om beelden te openen vanuit de Windows Verkenner: **klik** op een beeldbestand **met de rechter muisknop** en selecteer het commando "Herkennen" van het "Context" menu. (Dit bevel verschijnt enkel indien het bestandstype door Readiris ondersteund wordt.)



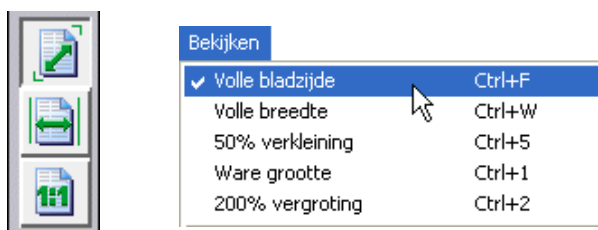
Eén en ander betekent niet dat de herkenning ogenblikkelijk wordt uitgevoerd: de gebruiker behoudt de volledige controle over het OCR-proces. Readiris wordt gewoon opgestart en het beeld wordt in het geheugen geladen.

De beeldknoppenbalk bevat alle commando's die nuttig zijn tijdens de beeldpreview: u vindt er tools om de zones van belang aan te duiden, om het beeld te roteren, om in- en uit te zoomen enz.

INZOOMEN OP BEELDEN

Readiris beschikt over enkele commando's waarmee u kunt **inzoomen** op het gescande beeld, bijvoorbeeld om de scankwaliteit te controleren.

De beeldknoppenbalk bevat knoppen waarmee u kunt inzoomen tot de werkelijke grootte, waarmee u het beeld kunt aanpassen aan de paginabreedte en waarmee u het volle beeld op het beeldscherm kunt brengen. Het menu "Bekijken" bevat diezelfde bevelen en voegt twee extra zoomniveau's toe: u kunt het beeld laten afbeelden op 50% en 200% van zijn ware grootte. Bij de werkelijke grootte beantwoordt één beeldpixel aan één scherpixel. (Er zijn overigens versnellingstoetsen beschikbaar voor alle de zoomniveau's.)



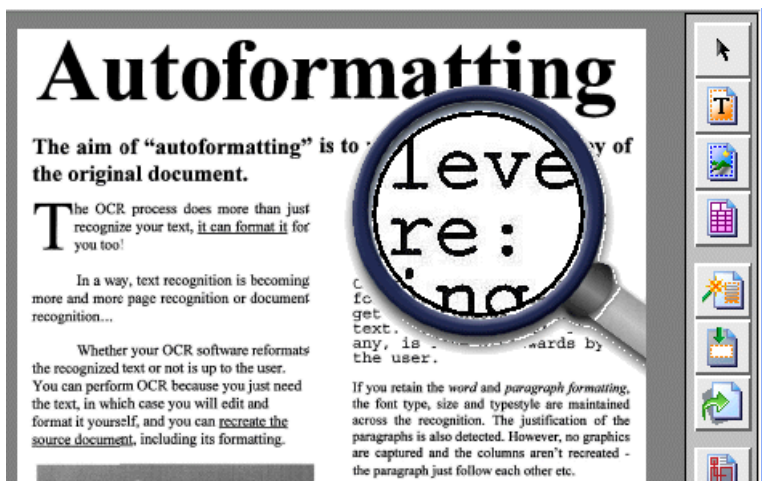
De zoomniveau's zijn overigens tevens beschikbaar via de rechter muisknop. Klik om het even waar met de rechtse muisknop om het "Context" menu op te roepen en selecteer er het gepaste zoomniveau.



U kunt tevens met de rechter muisknop *dubbeklikken* om op een deel van het gescande beeld in te zoomen op werkelijke grootte. Klik de rechter muisknop een tweede keer om opnieuw uit te zoomen.

Tenslotte kunt u van het **vergrootglas** gebruikmaken om in te zoomen op details van het gescand document. Het vergrootglas is ook beschikbaar op het "Context" menu wanneer u met de rechter muisknop boven het beeld klikt.

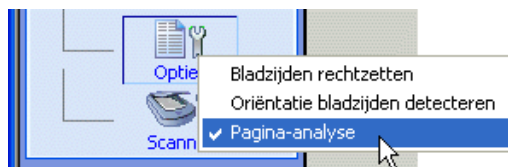




EÉN, EEN GESCAND BEELD ANALYSEREN

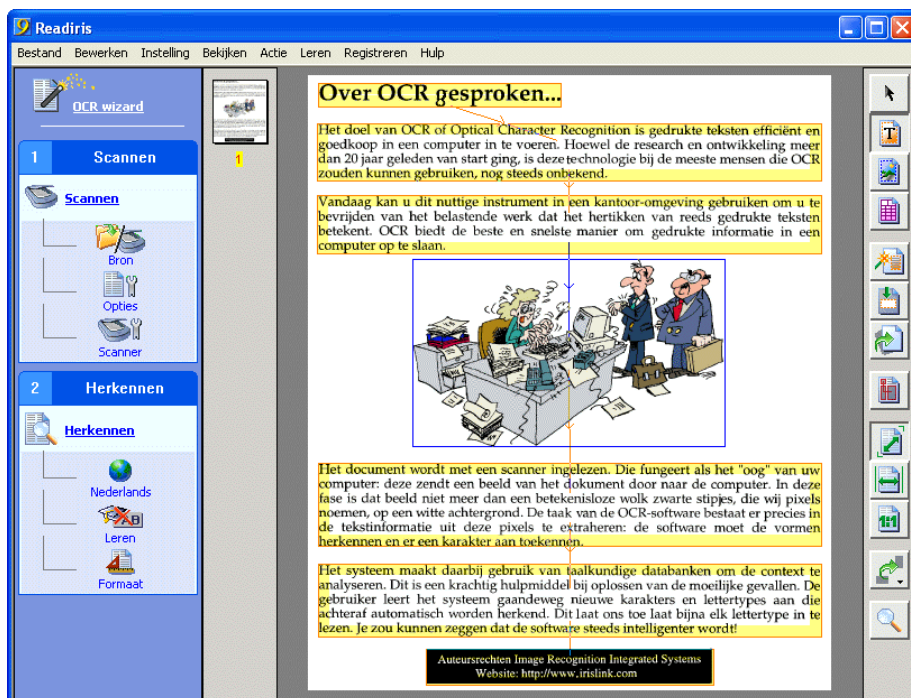
Nu het beeld gescand is, kunt u aangeven welke delen u wilt omzetten in bewerkbare tekst door kaders, zogenaamde "vensters", rond de zones van belang te tekenen.

Overigens kan Readiris dit automatisch voor u doen wanneer u de optie "Pagina-analyse" activeert met de knop "Opties" op de hoofdknoppenbalk (of onder het menu "Instelling").





De automatische pagina-analyse is bijzonder nuttig wanneer u **in kolommen geschikte tekst** moet herkennen en documenten met een ingewikkelde lay-out die mogelijkwerwijze illustraties en tabellen bevatten.



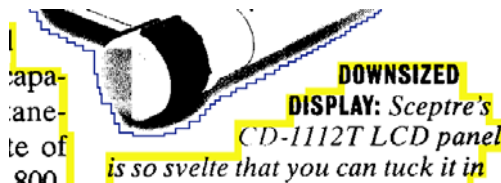
De pagina-analyse maakt gebruik van drie **venstertypes**: er zijn tekstvensters, grafische vensters en tabelvensters. Readiris onderscheidt de tekstblokken van de tabellen en de grafische zones die foto's, illustraties enz. bevatten. (We komen verder terug op de opslag van figuren en de herkenning van tabellen.)

Een **kleurcode** geeft aan met welk venstertype u te maken hebt: tekstvensters hebben een gele rand, grafische vensters hebben een blauwe rand en tabellen een purperen rand.

Het aantal vensters is altijd aangegeven in de tooltip van de tools "Tekstvenster", "Grafisch Venster" en "Tabelvenster".



De pagina-analyse verloopt snel, springt tolerant om met lijnkromming en is erg accuraat: de analyse vindt complexe, "onregelmatige" vormen terug.



De pagina-analyse zal zelfs zones detecteren met **witte tekst op een zwarte achtergrond**. Zulke tekstkaders herkennen vormt geen probleem: terwijl de voovertoning het gescande document getrouw op het beeldscherm brengt, zal Readiris tijdens de herkenningfase het beeld "negatief maken" om het kader te kunnen herkennen. (U kan ook de scanner *volledig* "negatieve" beelden laten aanmaken om zo bladzijden te verwerken met witte tekst op een zwarte achtergrond. Zie verderop.)

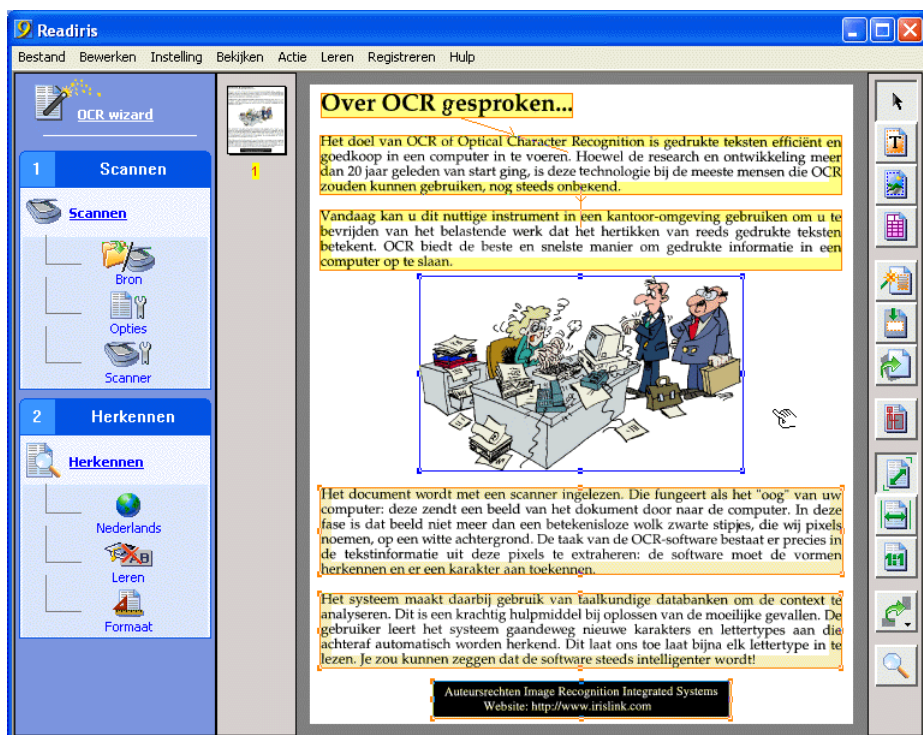
ANDERHALF, VENSTERS SORTEREN

Readiris detecteert niet alleen de verschillende zones, maar *sorteert* die ook: standaard worden de vensters van boven naar onder, van links naar rechts gesorteerd zodat in kolommen geschikte documenten goed verwerkt worden.

Natuurlijk kunt u de **sorteervolgorde** wijzigen. Om dit te doen klikt u op de knop "Sorteren" op de beeldknoppenbalk. Van zodra de "sorteermodus" actief is verandert de muiscursor in een hand met wijzende vinger.



Klik op de vensters die u wenst over te nemen. De vensters die u *niet* aanklikt zullen simpelweg niet herkend worden. Vernemen welke vensters wel en niet geselecteerd zijn is erg makkelijk: de geselecteerde vensters hebben hun volle kleur, de niet geselecteerde vensters zijn lichter van kleur.



Standaard is de pagina-analyse geactiveerd. Om Readiris te dwingen de huidige pagina te analyseren - u hebt bijvoorbeeld de pagina-analyse per ongeluk uitgeschakeld, u heeft sommige vensters gewist en nu wilt u de pagina-analyse herhalen enz. -, klikt u op de knop "Pagina Analyseren" in de beeldknoppenbalk.



Selecteer de taal van het document *eer* u de pagina-analyse uitvoert bij Aziatische documenten. Voor die talen worden er speciale routines gebruikt: de spatie tussen twee lijnen is meestal groter dan het geval is in Westerse documenten, de tekst bestaat uit kleine ikoontjes ("ideogrammen") die in Westerse documenten wel eens als grafische zone kunnen worden gezien, en de tekst kan van rechts naar links, van boven naar beneden lopen. En mocht u vergeten zijn de correcte taal te selecteren, selecteer die dan achteraf: Readiris voert de pagina-analyse vanzelf opnieuw uit!

Sommige documenten bevatten vlekjes op de bladzijde, zorgen voor een zwarte rand rond het eigenlijke beeld enz. U kunt alle kleine vensters wissen - we veronderstellen dat die geen tekst bevatten - en de overblijvende vensters opnieuw hersorteren met het bevel "Kleine Vensters Wissen" onder het menu "Bewerken".



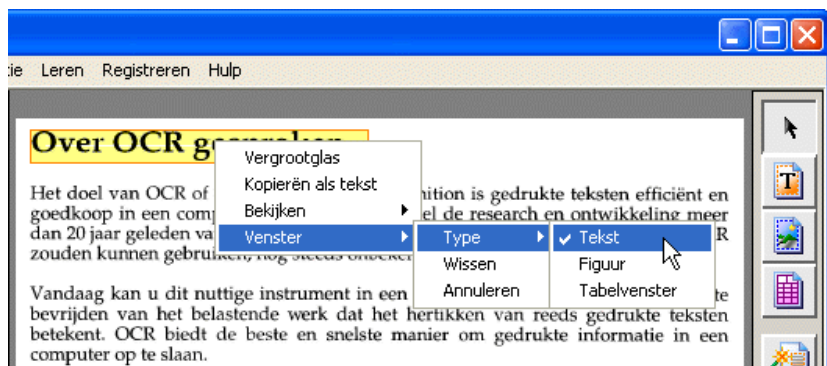
TWEE, BEELDEN MANUEEL BEVENSTEREN

De pagina-analyse is de automatische manier om gescande pagina's te bevensteren. Maar u kunt de zones van een beeld ook handmatig aangeven met de **venstertools** van Readiris!



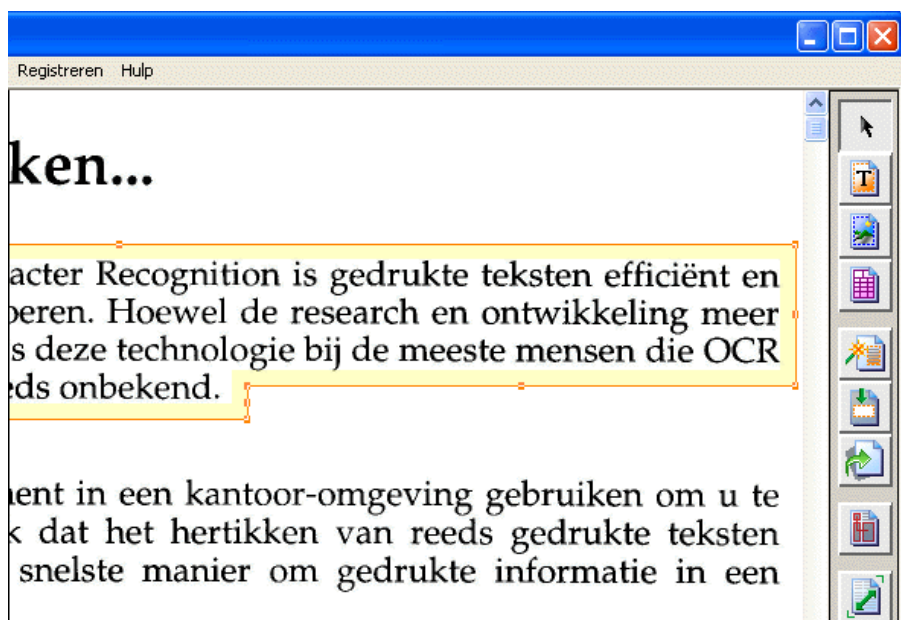
Om een rechthoekig kader te **tekenen** rond een te hernemen zone, selecteert u de corresponderende tool in de beeldknoppenbalk. Klik de cursor in de linker boven hoek van het venster, trek het venster open door de muis te verschuiven naar de tegenovergestelde hoek rechts onder en klik nogmaals. (Zijden kleiner dan 1 mm zijn niet toegelaten, ze zouden overigens niet één karakter bevatten.)

Maak u geen zorgen mocht u het verkeerde venstertype geselecteerd hebben: u kunt het zonetype snel wijzigen door met de rechter muisknop boven een venster te klikken. Selecteer het bevel "Venster - Type" in het "Context" menu.



De vensters worden automatisch gesorteerd naarmate ze gecreëerd worden: de pijlen geven de volgorde aan.

U kunt ook “onregelmatige” tekstblokken tekenen door **polygone vensters** rond dergelijke tekstzones te trekken. Veelhoekige vensters creëert men door rechthoekige vensters te versmelten: van zodra twee rechthoeken (van hetzelfde type) elkaar raken, worden ze automatisch omgezet in één venster! Het is alsof u een huis bouwt door de ene kamer na de andere toe te voegen... (Veelhoekige tabelvensters creëren heeft geen zin.)



Bovendien kunt u manuele bevenstering combineren met het sorteren van vensters; u kunt namelijk ook nieuwe vensters aanmaken wanneer de “sorteermodus” actief is. U sorteert dan een aantal vensters dat door de pagina-analyse werd gedetecteerd en u creëert nieuwe vensters waar de pagina-analyse geen bevredigend resultaat opleverde. Van zodra u een nieuw venster tekent in



de “sorteermodus”, worden alle vensters gewist die op dat ogenblik nog niet geselecteerd waren!

Om vensters te wijzigen, te verplaatsen en te wissen, moet u ze eerst **selecteren**. Om dit te doen, kiest u de knop "Vensters Selecteren" (de “pijl” knop) in de beeldknoppenbalk en klikt u op het venster. Vierkanten markeerpunten verschijnen op elke hoek en in het midden van de vensterzijden.

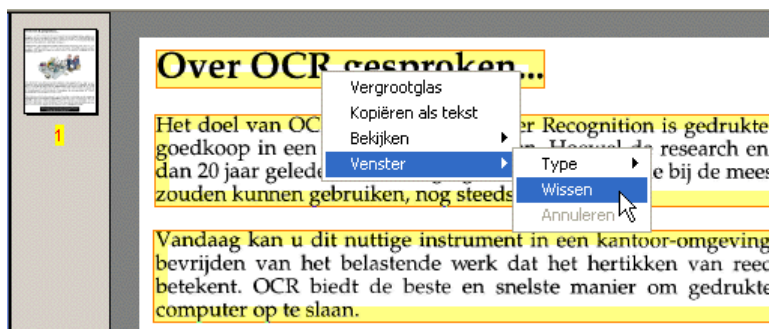


Klik de muisknop ergens anders om de **selectie** van vensters **ongedaan te maken**. Om **aanvullende vensters** te selecteren, houdt u de Shift knop ingedrukt terwijl u op de bijkomende vensters klikt. Om samen met een venster tegelijk de ingesloten vensters te selecteren (van een ander type), houdt u de Ctrl toets ingedrukt terwijl u op het hoofdvenster klikt.

Tot zover het selecteren van de vensters. Om een tekstvenster te **wijzigen**, selecteert u het, plaatst u uw muiscursor boven een markeerpunt en sleept u de vensterzijde zodat de afmetingen veranderen.

Om een venster te **verplaatsen**, selecteert u het en sleept u het naar een andere locatie.

Om vensters te **wissen**, selecteer u ze, klikt u met de rechter muisknop en selecteert u het bevel "Venster - Wissen" in het "Context" menu. Deze operatie wist alle geselecteerde venster én de zone onder de muiscursor.



U kunt tevens zones selecteren en het commando "Knippen" of "Wissen" van het menu "Bewerken" gebruiken. Het commando "Knippen" plaatst het venster(s) in een interne buffer, het commando "Wissen" wist het (de) venster(s) onherroepelijk. Wanneer u vensters plakt, worden ze ingevoegd worden op hun originele plaats; het is de taak van de gebruiker ze te verplaatsen naar hun nieuwe locatie.

Merk op dat *alle* klassieke opdrachten van het menu "Bewerken" van toepassing zijn op de vensters: vensters kunt u wissen, knippen, kopiëren en plakken! Het commando "Annuleren" speelt ook zijn rol: u komt op uw stappen terug wanneer u per ongeluk vensters heeft gewist, verplaatst, vergroot enz.

Bewerken	
Annuleren	Ctrl+Z
Knippen	Ctrl+X
Kopiëren	Ctrl+C
Plakken	Ctrl+V
Wissen	Delete
Kleine vensters wissen	Ctrl+M
Alles selecteren	Ctrl+A

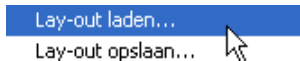
Merk tevens op dat er versnellingtoetsen beschikbaar zijn voor alle bevelen! Laten we een voorbeeld geven: om alle bestaande vensters te wissen, kiest u het



commando "Alles Selecteren" onder het menu "Bewerken" of de versnellingstoets Ctrl+A en klikt u op het comando "Wissen" of zijn versnellingstoets Del. U bent nu klaar om een nieuwe vensterlay-out aan te maken. Om de vorige lay-out te herstellen, kunt u "Annuleren" kiezen of zijn versnellingstoets Ctrl+Z.

DRIE, VENSTERSJABLONEN OPSLAAN

De resulterende vensterlay-outs kunnen opgeslagen worden als venstersjablonen met het commando "Lay-out Opslaan" onder het menu "Bestand" om in de toekomst opnieuw gebruikt te worden. U laadt een sjabloon in het geheugen met het commando "Lay-out Laden".



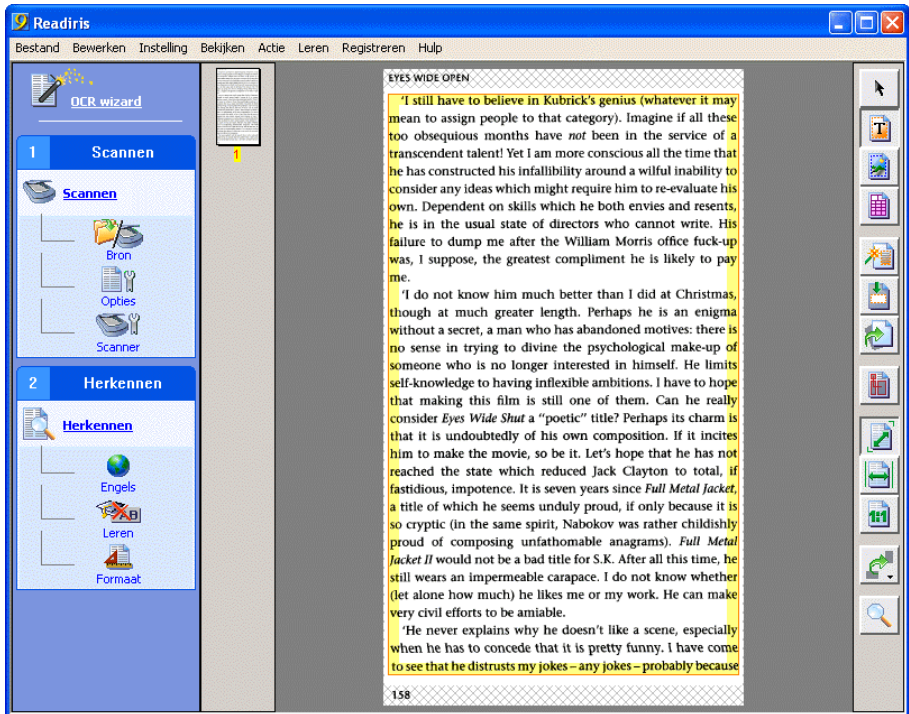
Als u documenten met een identieke opmaak wilt herkennen - bijvoorbeeld een rapport van 50 bladzijden waar de hoofding en de voetnoten om evidente redenen worden uitgesloten -, kunt u één sjabloon hanteren voor alle 50 pagina's.

Wanneer u een sjabloon in het geheugen laadt, wordt de pagina-analyse automatisch uitgeschakeld. De vensterlay-out blijft van kracht totdat u de pagina-analyse opnieuw activeert in de beeldknoppenbalk.

Overigens is er een leuk alternatief voorhanden voor zonesjablonen: de previewtool "Buitenkant van Zone Negeren" beperkt de pagina-analyse tot het "omkaderde" deel van het beeld.



Selecteer deze tool en trek een kader rond dat deel van het beeld dat u wilt verwerken. Wanneer u met een meerbladig document te maken hebt, kunt u dezelfde buitengrens hanteren voor alle bladzijden. (Voer de pagina-analyse opnieuw uit op het "bijsnijden" te annuleren, of pas de zones handmatig aan.)



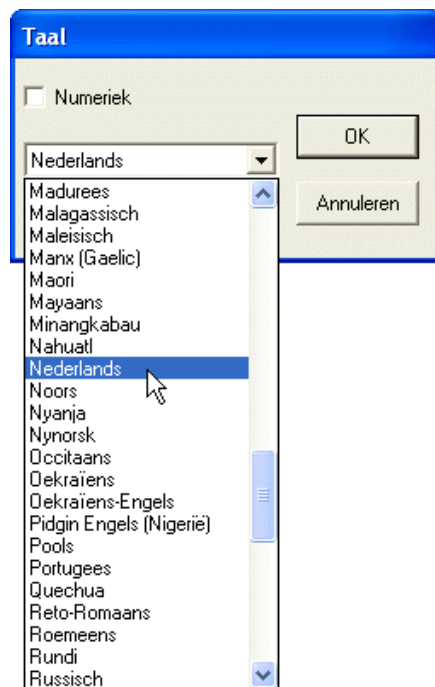
MET READIRIS KUNT U DE HELE WERELD ROND!

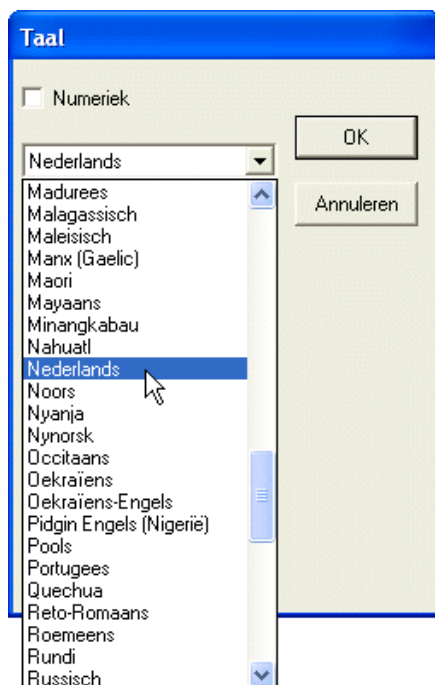
Nu de vensters correct getekend zijn, bent u bijna klaar om de karakterherkenning uit te voeren. We zeggen “bijna”, want u heeft de taal en de instellingen van het document nog niet gecontroleerd!

De taalinstelling vindt u op de hoofdknoppenbalk terug.



Klik op de knop "Taal" om de taal van het document aan te passen.





U kunt een lettertoets indrukken om onmiddellijk naar een taal te gaan: als u momenteel Nederlands geselecteerd hebt en u wilt Quechua selecteren drukt u op de "Q" toets op uw klavier om onmiddellijk naar het Quechua te gaan. Wanneer verschillende talen dezelfde beginletter hebben drukt u herhaalde malen op deze toets om de verschillende opties te doorlopen. Bijvoorbeeld: Readiris leest Deens en Duits. Door éénmaal op "D" te drukken, selecteert u het Deens, door een tweede maal op "D" te drukken selecteert u het Duits en door een derde maal op de "D" toets te drukken keert u terug naar het Deens. (Om naar een *andere* letter te gaan, bijvoorbeeld T, drukt u op de toets BackSpace voor u de "T" toets indrukt.)



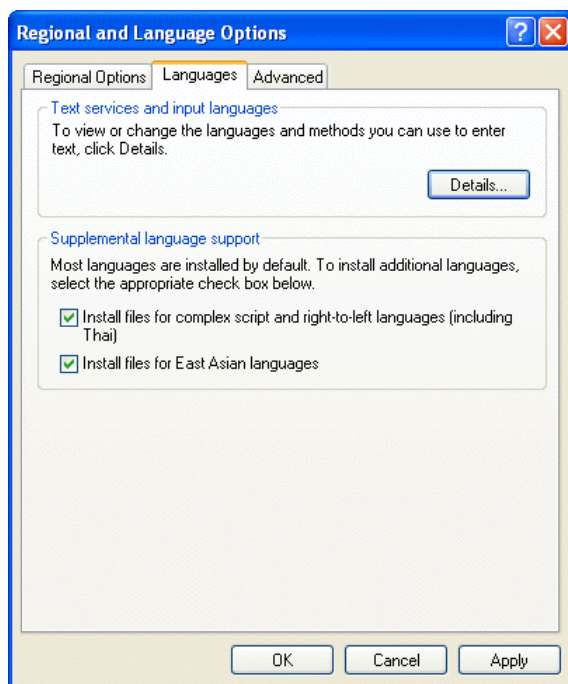
Readiris is helemaal niet beperkt tot het Nederlands: er worden wel 107 **talen** ondersteund! Alle Europese en Amerikaanse talen worden herkend, en daar horen ook de Centraal-Europese talen, het Grieks, het Turks, de Cyrillische talen (de “Russische” talen) en de Baltische talen bij.

Optioneel herkent Readiris **Aziatische documenten**: de extra module voor Aziatische herkenning leest het Japans, het vereenvoudigd en traditioneel Chinees en het Koreaans. (Het vereenvoudigd Chinees wordt gebruikt op het Chinees vasteland en in Singapore, het traditioneel Chinees in Hong Kong, Taiwan, Macau en de overzeese Chinese gemeenschappen.)

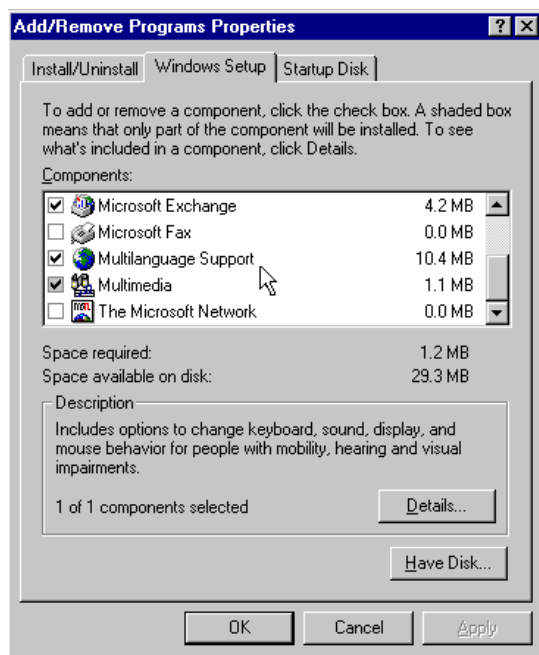
Merk ook op dat de Britse en Amerikaanse varianten van de Engelse taal - of zullen we eerder over “internationaal” Engels spreken in plaats van Amerikaans? - onderscheiden worden.

Merk op dat u een aangepaste configuratie van Windows nodig hebt om de Centraal-Europese, Griekse, Turkse, Cyrillische en Baltische karakters te ondersteunen. U dient de **veeltalige ondersteuning** van **Windows** te installeren opdat uw Windows systeem met deze talen kan omspringen.

Op een machine die uitgerust is met het besturingssysteem Windows XP, 2000 en Windows NT 4.0, selecteert u het pictogram "Regionale Instelling (en Talen)" onder het "Controlepaneel".



Indien u daarentegen met Windows ME of 98 werkt, selecteert u het pictogram "Software Toevoegen/Wissen" onder het "Controlepaneel" om uit te zoeken of de Windows module "Veeltalige Ondersteuning" al op uw PC geïnstalleerd is.



Om Aziatische documenten op het scherm af te beelden en te bewerken, kunt u een Aziatische versie van het Windows besturingssysteem installeren of gespecialiseerde “emulator”-software (zoals UnionWay AsianSuite of TwinBridge AsianBridge) installeren op een Westerse versie van Windows om de ideogrammen van deze Aziatische talen correct voor te stellen. Maar u kunt ook Word 2003, Word 2002 en Word 2000 gebruiken om dergelijke documenten te openen en bewerken: Office 2003 System, Office XP en Office 2000 werden specifiek ontworpen om te kunnen omspringen met documenten uit vele verschillende talen!

Raadpleeg het **“Read Me” bestand** van Readiris als u hierover meer wilt vernemen.

De correcte taal kiezen is van essentieel belang. Op basis van de geselecteerde taal, zal het programma weten welke **karakterset** dient herkend te worden. De veeltalige ondersteuning garandeert dat u “exotische” karakters zoals ß, ñ, γ en ø correct herkent.

Ten tweede maakt het programma uitgebreid gebruik van **linguïstische databanken** om de resultaten te valideren. Stel dat u het woord "president" moet lezen waarbij een inktvlek de "r" in een "f" heeft veranderd. Door het voorlopige resultaat op te zoeken in het lexicon zal Readiris autonoom ontdekken dat het woord "president" gelezen wordt en dat het dus geen zin heeft om het symbool als "f" te herkennen. Deze **zelflerende techniek** hangt natuurlijk sterk af van de linguïstische context.

De taalkunde vormt een belangrijk hulpmiddel om **ambiguë gevallen** op te lossen zoals een "O" die als een '0' kan worden gelezen. Of denk aan de letter "l" en het cijfer '1' die in heel wat lettertypes dezelfde vorm hebben - denk aan teksten die afkomstig zijn van een oude schrijfmachines! De linguïstische context helpt bepalen of u te maken heeft met een "l" of een '1'.

De onderstaande illustratie laat verschillende exemplaren zien van ééntjes en l-symbolen. De vormpjes op de eerste lijn zijn niet dubbelzinnig, de vormen op de tweede lijn zijn dat wel, maar de taalkunde biedt een uitweg. Wanneer de context geen uitsluitel biedt, moet de gebruiker tussenkomen.



193 1950s. 1hr
Well, Rossellini

READIRIS VERANDERT TUSSENDOR VAN TAAL

Maar het blijft hier niet bij: Readiris kan middenin een zin van taal veranderen zonder tussenkomst van de gebruiker! Wanneer Westerse woorden opduiken in

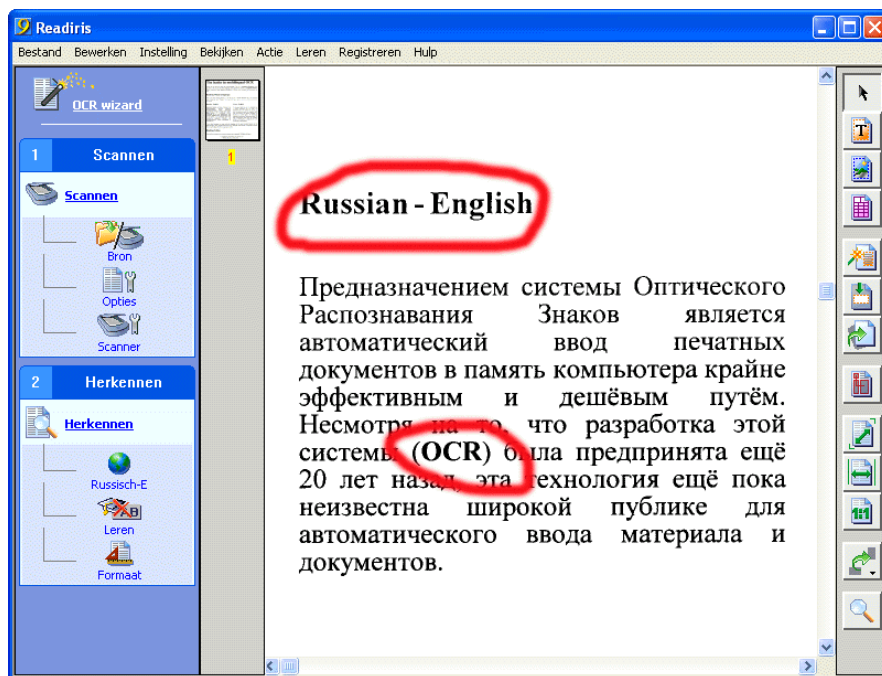


Griekse, Cyrillische en Aziatische documenten - vele eigennamen, merknamen enz. worden geschreven in het Latijns alfabet - zal Readiris automatisch naar het correcte alfabet grijpen. Met andere woorden, u kunt een **gemengd alfabet** activeren van Griekse, Cyrillische of Aziatische karakters in combinatie met Latijnse karakters.

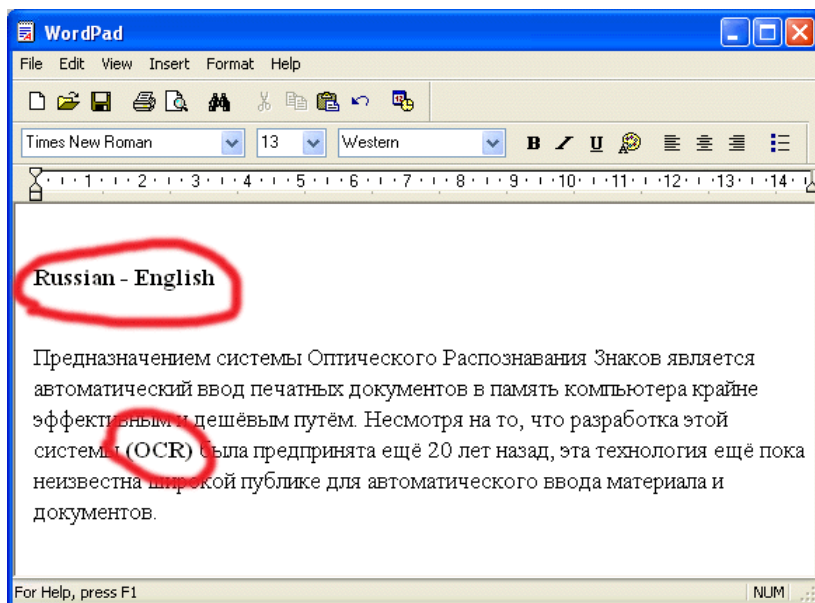
Wees er zeker van dat u "Grieks-Engels" geselecteerd hebt of een gepaste Cyrillische taalinstelling - bijvoorbeeld "Oekraïens-Engels". Met andere woorden: selecteer niet gewoon het "Grieks" of "Oekraïens" als taal van het document in de hoop dat de Westerse symbolen wel goed door de herkenning zullen komen.



Dit is een voorbeeld waar een Russische tekst enkele Engelse woorden bevat - open het beeld ALPHABETS.TIF mocht u het zelf willen proberen!



Wanneer u het resultaat met de tekstverwerker opent ziet het er zo uit - het kan gebeuren dat u een Cyrillisch **lettertype** moet activeren om de Russische tekst correct weer te geven op het beeldscherm.



Om **andere talen** te **mengen** moet u die taal selecteren met de meest uitgebreide karakterset. Stel dat u een document heeft waar de Franse vertaling naast de Nederlandse tekst staat: u moet het Frans selecteren om ervoor te zorgen dat de geaccentueerde karakters zoals ç, é en ù correct herkend worden.

DE KENMERKEN VAN HET DOCUMENT AANGEVEN

Nu de taal is ingesteld, gaan we over naar de andere kenmerken van het document. U kunt de herkenning aanpassen door enkele karakteristieken van het document aan te geven: het type en de karakterafstand van de lettertekens. (Deze commando's zijn niet van toepassing op Aziatische documenten.) Laten we verhelderen wat dit alles betekent.

We vangen aan met commando "Type Lettertekens" onder het menu "Instelling". Het "type" lettertekens onderscheidt de "normale" documenten van de **matrixgedrukte** documenten. "Draft" dot matrix symbolen - ook wel "9 pin" matrixdruk genoemd - zijn opgemaakt uit aparte vlekken, en er zijn sterk gespecialiseerde herkenningstechnieken nodig om dergelijke teksten te lezen.

ape-descended life

Matrixdruk met "briefkwaliteit", ook wel "25 pin" matrixdruk genoemd, vereist de "normale" instelling, en dat is ook het geval voor de **drukkwaliteiten** gedrukt, getikt, laserdruk en inkjetdruk.

De optie "Automatisch" betekent dat Readiris het lettertype automatisch zal terugvinden. Laat Readiris onder alle omstandigheden de drukkwaliteit automatisch detecteren - tenzij u zeker bent dat er enkel matrixgedrukte documenten worden ingelezen. ("Automatisch" is natuurlijk de standaard waarde.)



De drukkwaliteit wordt aangegeven in de tooltip van de knop "Herkennen": wanneer aan de tooltip geen bericht is toegevoegd, is de automatische detectie van de drukkwaliteit van toepassing, wanneer het bericht "Dot matrix" is toegevoegd, is de dot matrix leesmodus actief.



De **karakterafstand** kan gekozen worden met de opdracht "Karakterafstand" onder het menu "Instelling".



Bij *vaste* lettertypes hebben alle symbolen van één lettertype dezelfde breedte. Een "l" neemt evenveel horizontale plaats op binnen een lijn als een "w", zoals in deze zin het geval is. Neem bijvoorbeeld de documenten die afkomstig zijn van een schrijfmachine: de wagen van de schrijfmachine schuift telkens op met een vaste afstand.

Bij een *proportioneel* lettertype hangt de breedte van elk symbool van zijn specifieke vorm af. Symbolen zoals "m" en "w" zijn duidelijk breder, nemen meer plaats in binnen een lijn dan de dunne karakters "l" of "j". Zowat alle boeken, tijdschriften en kranten gebruiken exclusief proportionele lettertypes.

Dit is de simpelste oplossing: laat deze optie altijd op "Automatisch" staan, zodat Readiris de karakterspatiëring automatisch zal detecteren.

READIRIS WORDT TELKENS INTELLIGENTER!

Wanneer de taal en de kenmerken van het document aangegeven zijn, activeert u het interactief leren en klikt u op de knop "Herkennen".



De vooruitgang van het OCR-proces wordt op het scherm getoond. U kunt op de knop "Stop" klikken (en op Esc drukken) om de tekstherkenning af te breken.



Indien de leermodus ingeschakeld werd met de knop "Leren" op de hoofdknoppenbalk komt Readiris op het einde van de herkenning terecht in de interactieve fase.

(Het interactief leren is niet van toepassing op Aziatische documenten: het systeem trainen heeft geen zin bij talen die duizenden symbolen gebruiken - en dan gaan we er overigens van uit dat u de ideogrammen kunt intikken, geen makkelijke taak op een Westers toetsenbord!)

De leermodule kan in bepaalde gevallen de herkenning sterk verbeteren. Wanneer de gebruiker probeert om gebroken, misvormde lettertekens in te lezen zoals men die terugvindt in echte documenten, of wanneer men op gestileerde lettertekens botst die Readiris niet optimaal herkent, kan het **interactief leren** dit voorlopig "falen" overwinnen.

De leermodus wordt ook gebruikt om het systeem te trainen op **speciale symbolen** die Readiris initieel niet herkent, zoals wiskundige of wetenschappelijke symbolen en dingbats. Een paar voorbeelden: Readiris kan getraind worden om het symbool " π " te herkennen als "pi" of de dingbat "☎" als



"Tel". (Merk op dat de lijst van herkende symbolen echter niet kan uitgebreid worden met de symbolen " π " en "☎"!)

De herkende tekst wordt progressief doorlopen en het systeem stopt op de twijfelachtige karakters of - als we het over elkaar rakende karakters ("ligaturen") hebben - op de twijfelachtige karakterreeksen. De lettertekens worden altijd in hun context getoond, met de nodige klemtoon op de twijfelachtige karakters. Niet herkende karakters worden voorgesteld met een tilde (het symbool "~").



Het eerste wat u moet doen is nagaan of het correcte woordenboek en de goede woordenboekmodus geactiveerd zijn - deze zijn altijd aangegeven in de titel van het leervenster. Mocht dat niet het geval zijn, klik dan op "Afbreken" - het beeld van het document wordt weer op het scherm gebracht mét zijn bevenstering -, laad het juiste woordenboek of de woordenboekmodus en herhaal de herkenning. (De werking van de woordenboeken komt zo meteen aan bod.)

Voer indien dat nodig is een karakter (of een karakterreeks) in voor de verkeerde of niet herkende tekens en klik op één van de volgende knoppen.

Leren

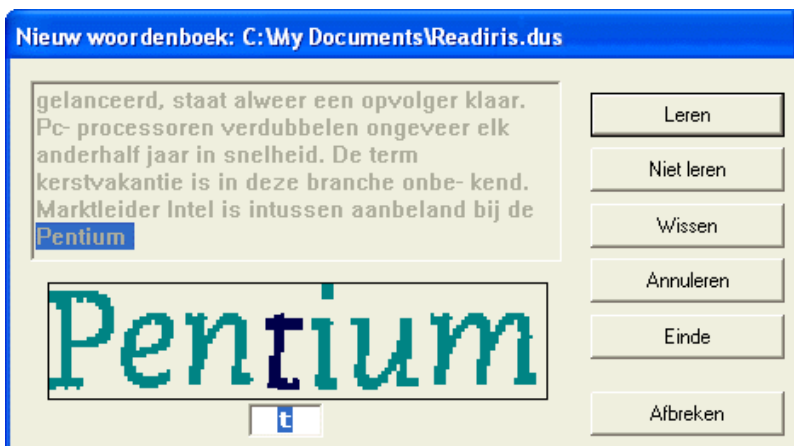
U gaat akkoord met de voorgestelde oplossing of u verbetert die. Het programma bewaart dit twijfelachtige karakter in het woordenboek als “zeker”, definitief. De verdere herkenning ervan zal geen tussenkomst meer vergen, de vorm wordt als voor ééns en altijd als aangeleerd beschouwd.

In het bovenstaand voorbeeld stopt het systeem op een twijfelachtige vorm, en we klikken op "Leren" om een vorm aan te leren die niet kan verward worden met andere karakters.

Niet Leren

U gaat akkoord met het voorgestelde oplossing of u verbetert ze. Het verschil met de knop "Leren" zit hierin dat de vorm de status “onzeker” krijgt in het woordenboek. Bij verdere herkenning zal het systeem u alvast wel de “aangeleerde” oplossing voorstellen, maar de herkenning zal nog steeds uw bevestiging vereisen.

Deze knop wordt gebruikt bij symbolen die kunnen verward worden met andere: een vervormde "e" die sterk op een "c" lijkt, een beschadigde "t" die veel weg heeft van een "r" enz.



De "t" hierboven is serieus beschadigd - in feite lijkt hij op een "r" -, en u klikt hier op "Niet Leren" om dit symbool niet met de "r" te verwarren.

Wissen

De twijfelachtige vorm wordt uit de uitvoer verwijderd. Gebruik deze knop om "ruis" die op de documenten voorkomt - inkt- of koffievlekken die herkend raken als punten of komma's en dergelijke - uit de uitvoer te verwijderen, of om een willekeurig ongewenst symbool te wissen.

Annuleren

U keert op uw stappen terug om fouten te verbeteren. Readiris houdt de 32 laatste operaties bij.

Einde

Het leerproces wordt gestopt maar het OCR-proces gaat verder in de automatische modus. Alle oplossingen van het systeem worden aanvaard zonder validatie door de gebruiker.

Klik op deze knop wanneer u ziet dat de herkenning erg goed is en geen gedetailleerde verificatie behoeft.

Afbreken

Verwar "Einde" niet met "Afbreken": bij "Afbreken" wordt er geen uitvoer gegenereerd en moet u de herkenning overdoen, met de knop "Einde" wordt de tekst wel aangemaakt, alleen neemt u die niet in detail door.

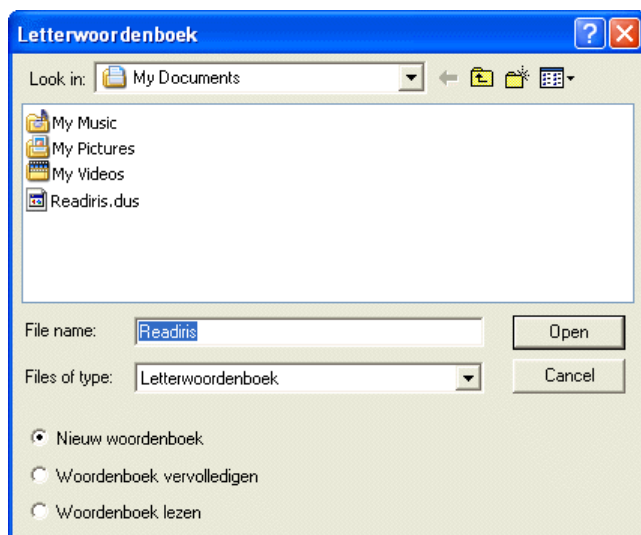
DE ROL VAN WOORDENBOEKEN

De resultaten van elke trainingssessie worden tijdelijk in het computergeheugen bewaard maar kunnen en moeten voor verder gebruik opgeslagen worden in bestanden die we “woordenboeken” noemen.

(Verwar letterwoordenboeken niet met (gebruikers)lexicons! Woordenboeken bevatten de vormen van lettertekens waarop u het systeem trainde, (gebruikers)lexicons zijn linguïstische databanken die de herkenning ondersteunen!)

De letterwoordenboeken worden in het geheugen geladen om bij gelijkaardige documenten van de extra intelligentie gebruik te maken. Readiris gebruiken betekent Readiris telkens intelligenter maken!

Hoe gaat dit in zijn werk? De werking van de woordenboeken wordt beheerd door het menu "Leren": u moet een woordenboek selecteren met het commando "Woordenboek" en zijn werkingsmodus bepalen.



Woordenboeken zijn beperkt tot 500 vormen; wij raden u aan aparte woordenboeken aan te maken voor specifieke applicaties - bijvoorbeeld per documenttype. Woordenboeken hebben standaard de extensie *.DUS. De training heeft geen effect meer wanneer het woordenboek vol is: de resultaten van het leerproces worden dan niet langer in het geheugen opgeslagen of naar het woordenboek weggeschreven.

U kunt de woordenboekmodus aanpassen met het commando "Woordenboek" of rechtstreeks onder het menu "Leren". Er zijn drie woordenboekmodi: nieuw, vervolledigen en lezen.



De optie "Nieuw Woordenboek" geeft aan dat de resultaten van de training zullen opgeslagen worden in een *nieuw* woordenboek. (Wanneer u een bestaand woordenboek selecteert wordt de inhoud gewist.)

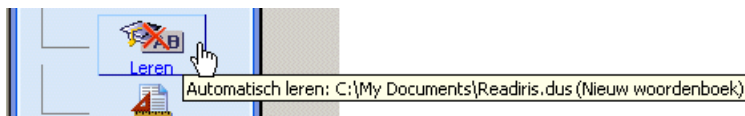
De uitbreidingsmodus geeft aan dat de trainingresultaten zullen bewaard worden in een *bestaand* woordenboek: de herkenning maakt al gebruik van de extra intelligentie die in het woordenboek aanwezig is, en u voegt nieuwe vormen aan het woordenboek toe. Het komt hierop neer dat u met deze optie een woordenboek opbouwt in verschillende stappen.

(Wanneer u een nieuwe bestandsnaam invult voor een woordenboek en u activeert de vervolledigingsmodus, zal een leeg woordenboek worden gecreëerd en gaandeweg ingevuld.)

Met de laatste optie, "Woordenboek Lezen", zal het woordenboek enkel gelezen, geraadpleegd worden: u maakt gebruik van het woordenboek *zonder* er wat dan ook aan toe te voegen ("read only").

Selecteer de modus "Nieuw" wanneer u één enkele bladzijde herkent. Om verschillende bladzijden van hetzelfde type - we bedoelen met hetzelfde lettertype en dezelfde drukkwaliteit - te herkennen, kiest u de modus "Nieuw" voor de eerste bladzijde, de vervolledigingsmodus voor enkele volgende bladzijden en de leesmodus voor de rest van het document.

Noteer dat de tooltip van de knop "Leren" op elk ogenblik aantoont welk woordenboek actief is en in welke werkingsmodus dat woordenboek zich bevindt.



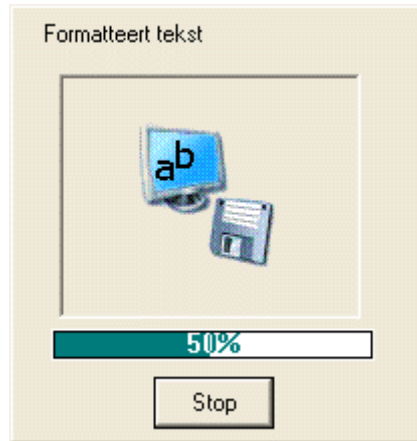
Wanneer u in het interactief leren terecht komt worden het woordenboek en zijn modus aangegeven in de titel van het leervenster, en u klikt op "Annuleren" om opnieuw te beginnen wanneer deze foutief blijken.



TEKST DIRECT NAAR EEN TOEPASSING STUREN

Het interactief leren beëindigt de karakterherkenning. Daar Microsoft Word fungeert als standaard doeltoepassing voor de uitvoer, wordt uw tekstverwerker na de herkenning (indien nodig) automatisch opgestart en wordt de herkende tekst ingevoegd.

Het kan voorkomen dat u de vooruitgang van de formattering van het document op het scherm kunt volgen. (Of dat daadwerkelijk het geval is hangt van de grootte van het document en de complexiteit van de formattering.)

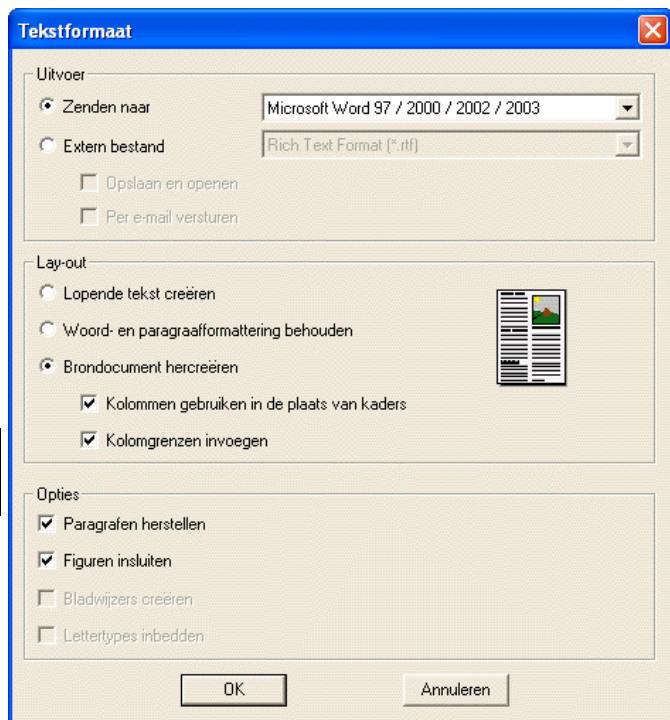


Het gescande beeld verschijnt opnieuw op het scherm mét de bevenstering. Het beeld is klaar voor verdere verwerking, het blijft afgebeeld tot u een nieuwe bladzijde inscant.

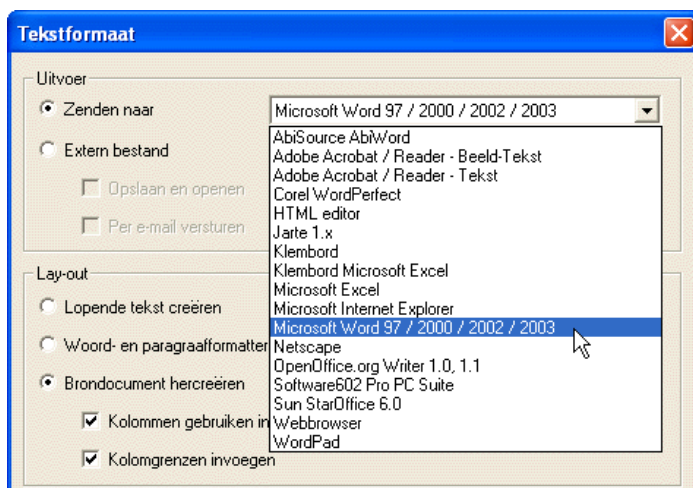
U heeft inderdaad een document omgezet in een editeerbaar computerbestand, en dit wel 40 maal vlugger dan indien u de tekst had hertikt! Ga uw gang en vergelijk het tekstresultaat met het beeld binnen het Readiris venster.

In feite biedt Readiris u drie verschillende methoden om de herkenningresultaten te bewaren: u kunt het herkende document rechtstreeks naar een doeltoepassing zenden, u kunt het resultaat opslaan in een extern bestand en in het Windows klembord stoppen.

De **uitvoermodus** wordt bepaald met de "Formaat" knop (of met het commando "Tekstformaat" onder het menu "Instelling").



De mogelijkheid om documenten “door te **zenden**” biedt u een rechtstreekse OCR link tussen uw scanner en uw Windows applicaties: u voert documenten onmiddellijk uit naar uw tekstverwerker, rekenblad of webbrowser, naar Adobe Reader enz.!



Na de herkenning wordt de doeltoepassing opgestart en worden de resultaten automatisch ingevoegd.

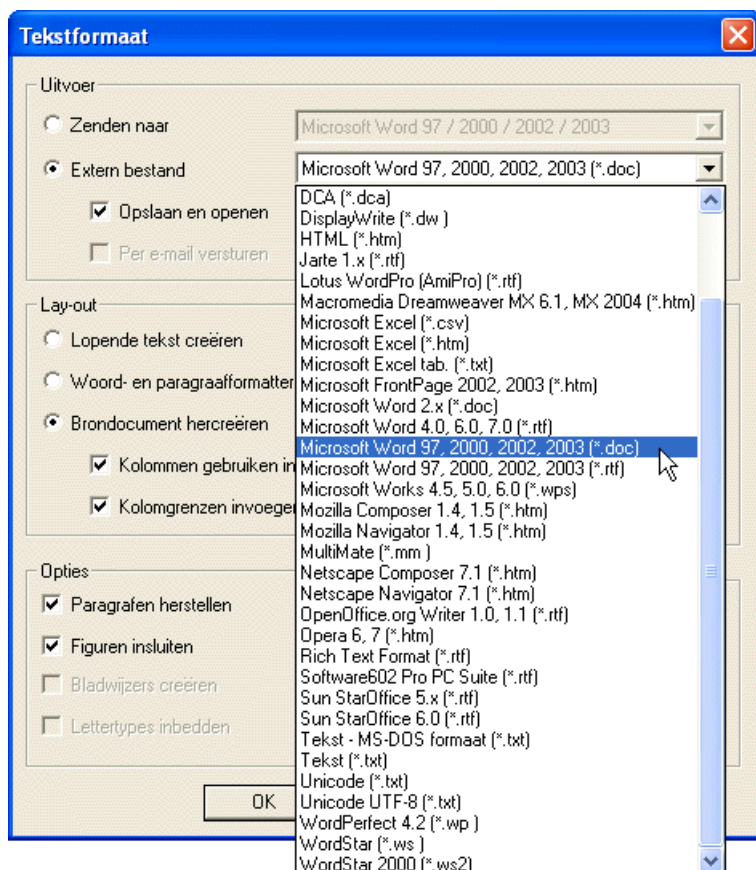




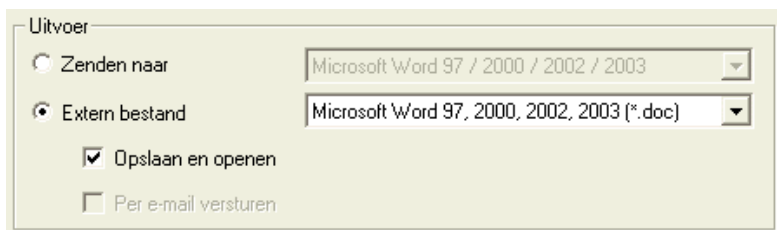
Vergeet echter niet dat de optie "Zenden naar" u ook teksten in het Windows **klembord** laat stoppen, er is dus geen strikte noodzaak om ze uit te voeren... of in een bestand op te slaan!

HET RESULTAAT IN EEN TEKSTBESTAND OPSLAAN

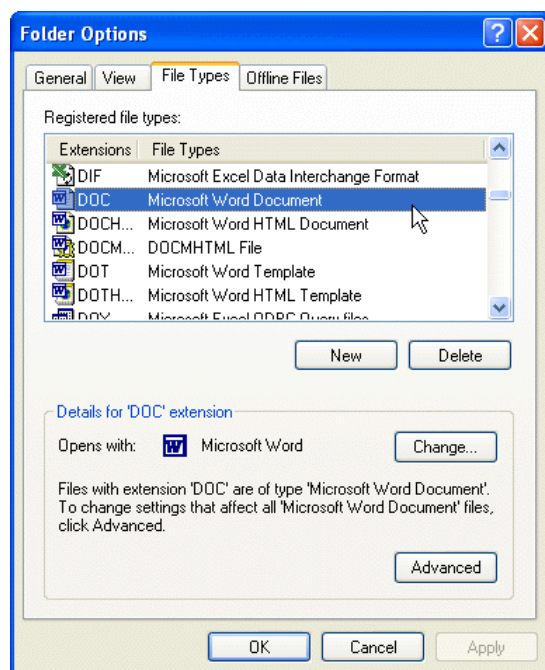
U kunt het herkenningsresultaat inderdaad bewaren in een “extern” bestand. Ook hier ondersteunt Readiris een breed gamma aan bestandsformaten dat alle populaire tekstverwerkers, rekenbladen, webtoepassingen enz. omvat.



De Optie "Opslaan en Openen" is in grote lijnen gelijkwaardig met de optie "Zenden naar": u opent het herkende document eens u het opgeslagen hebt.



De manier die wordt aangewend om de doeltoepassing te doen lopen is echter fundamenteel verschillend. Hier bepaalt het **Windows bestandstype** welke toepassing zal opgestart worden. Het is net alsof u het uitvoerbestand in de Windows Verkenner twee keer aangeklikt hebt... (De optie "Zenden naar" richt zich rechtstreeks naar specifieke doeltoepassingen.)



De optie "Per E-mail Versturen" creëert een nieuw e-mailbericht en voegt het herkend document toe als aangehecht bestand. Kent u een snellere manier om een papieren document vlug te verspreiden...?



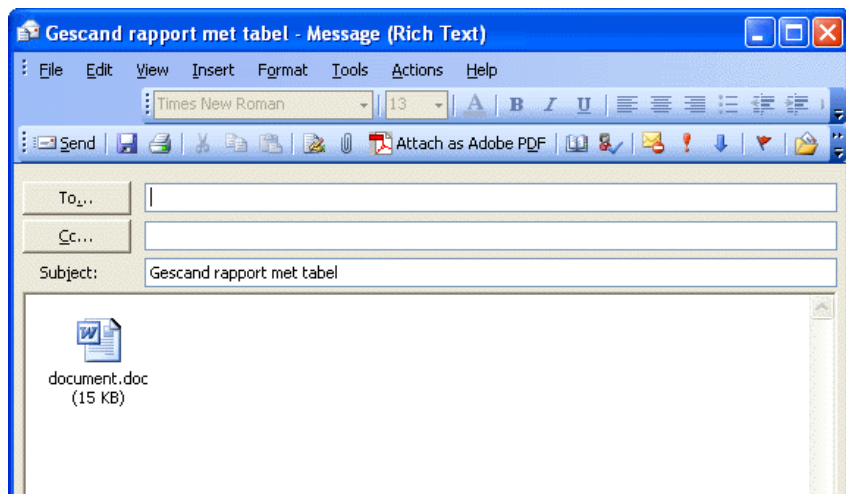
Uitvoer

☐ Zenden naar Microsoft Word 97 / 2000 / 2002 / 2003

☒ Extern bestand Microsoft Word 97, 2000, 2002, 2003 (*.rtf)

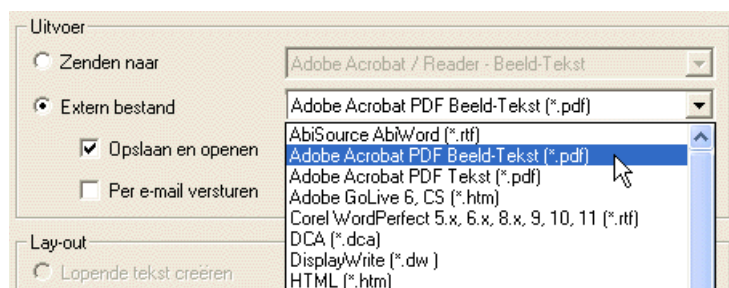
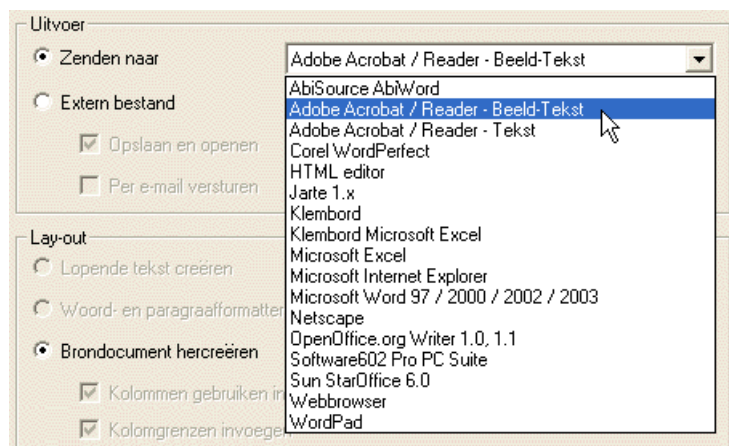
☐ Opslaan en openen

☒ Per e-mail versturen

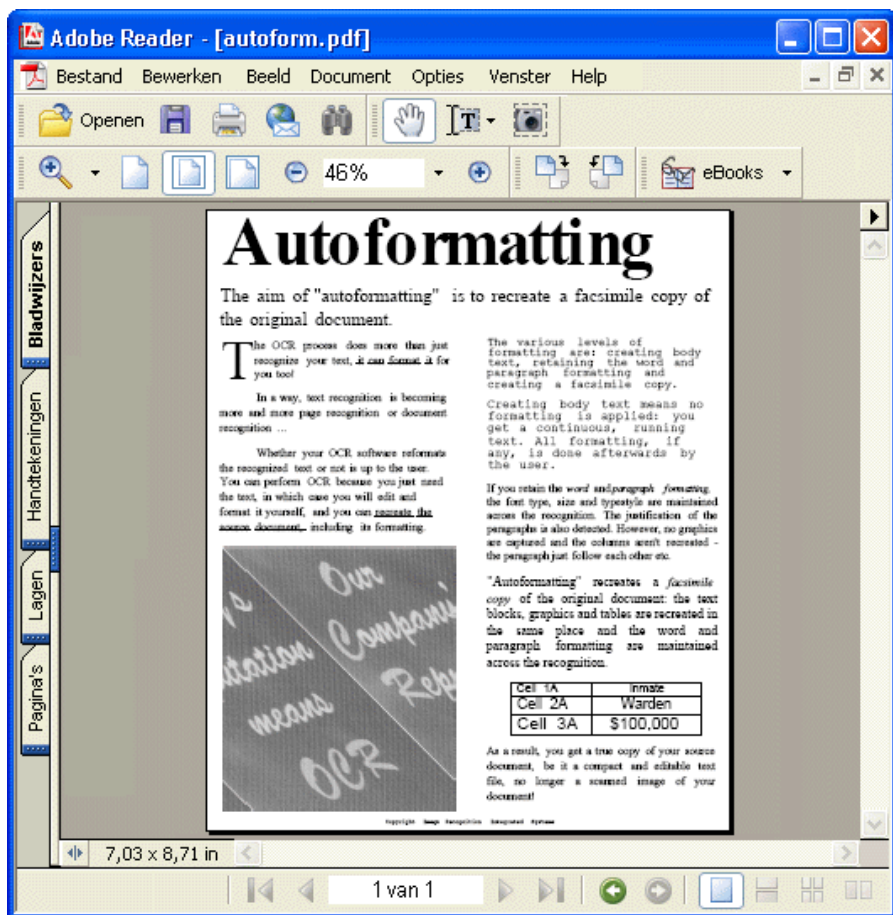


UNIVERSELE BESTANDEN AANMAKEN...

We gaan wat dieper in op één formaat: **Adobe Acrobat PDF**. Met Readiris kunt u twee types PDF documenten aanmaken - PDF Tekst en PDF Beeld-Tekst.

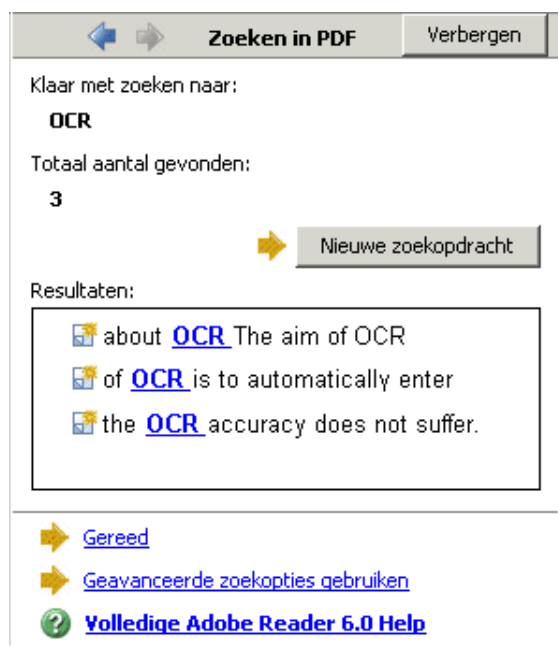


Wat is het verschil tussen beide formaten? Wanneer u de optie "PDF Tekst" kiest, maakt Readiris een PDF bestand aan dat het tekstresultaat bevat. (Er kunnen wel figuren in het bestand voorkomen maar enkel als die ook op de bladzijde voorkomen - foto's, illustraties enz.) Met andere woorden: het beeld van de volle bladzijde is *niet* opgenomen in het PDF bestand en het PDF bestand bevat slechts één laag!



Het formaat "PDF Beeld-Tekst" levert andere resultaten op: Readiris creëert een doorzoekbaar PDF bestand dat de herkende tekst *en* het beeld van de bladzijde

bevat. Het beeld van de bladzijde is boven de tekst opgenomen in het PDF bestand dat nu twee lagen bevat. Maak gebruik van de zoekfunctie van de Adobe Reader software en u zult snel het verschil begrijpen!



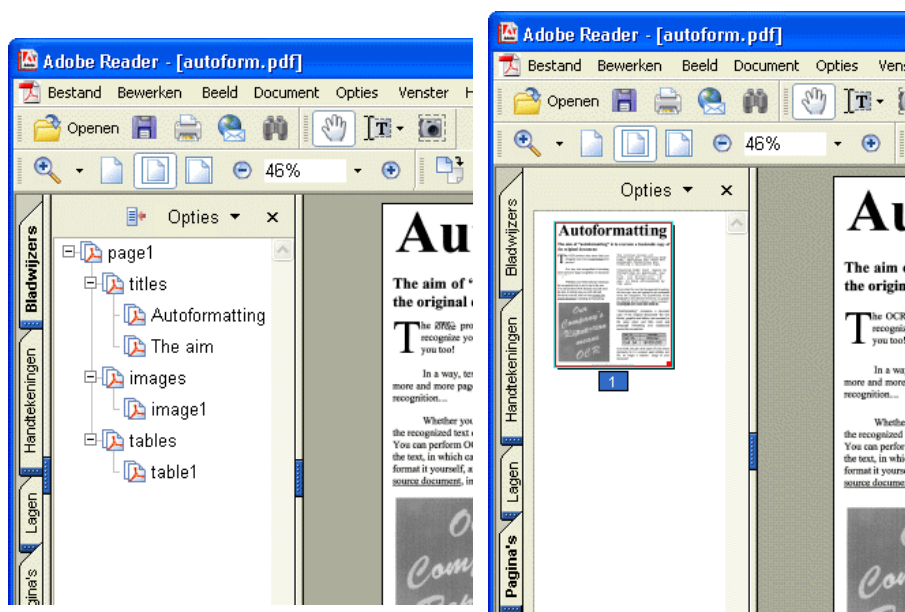
Klik op de knop "Formaat" om twee opties te ontdekken die het Acrobat PDF formaat betreffen: "Bladwijzers Creëren" en "Lettertypes Inbedden".



Opties

- ☒ Paragrafen herstellen
- ☒ Figuren insluiten
- ☒ Bladwijzers creëren
- ☒ Lettertypes inbedden

De optie "Bladwijzers Creëren" zorgt ervoor dat er een **bladwijzer** wordt aangemaakt voor elk element van het document - en de figuren zijn evengoed elementen als de tekstblokken en de tabellen. Voor de tekstvensters past Readiris een intelligent algoritme toe dat per venster een titel vindt, zeg maar een samenvatting maakt. De tabellen en figuren worden gewoon genummerd. Een ander navigatiemiddel van PDF documenten, **miniaturen** ("thumbnails"), kunt u dynamisch aanmaken met uw Adobe Reader software!)



De optie "Lettertypes Inbedden" sluit de lettertypes in de PDF bestanden in. Lettertypes inbedden belet de vervanging van lettertypes wanneer de lezer het herkende document op het scherm bekijkt en afdrukt. Het zorgt ervoor dat de lezer - welke computer hij ook moge gebruiken - de tekst te zien krijgen met de correcte, oorspronkelijke lettertypes. Lettertypes inbedden verhoogt wel (enigzins) de bestandsgrootte van de herkende documenten!

... OF ZE LEZEN!

We kijken even de andere richting uit. Aangezien Readiris het documentformaat Adobe Acrobat PDF volledig ondersteunt, kunt u dergelijke bestanden niet enkel aanmaken maar ook *inlezen* met Readiris!



PDF documenten “hieroriënteren” is één van de belangrijke objectieven van Readiris, en daar bestaan verschillende redenen voor. Het is bijvoorbeeld een manier om beelden in tekst om te zetten: laad op beeld gebaseerde PDF documenten in het geheugen, voer de herkenning uit en sla het herkenningsresultaat op in een tekstdocument (in het tekstformaat van uw keuze). Tekstbestanden zijn natuurlijk editeerbaar en beeldbestanden zijn dat niet.

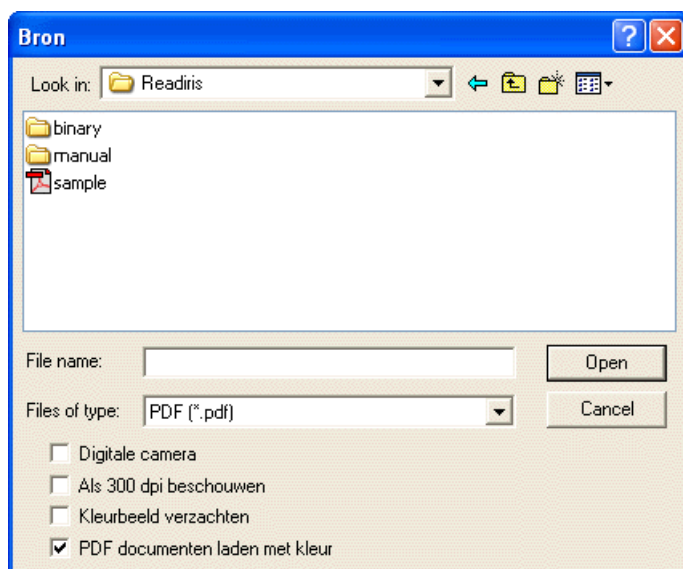
Tweede scenario: u zet op beeld gebaseerde PDF documenten om in PDF documenten die op tekst zijn gebaseerd. U voert dan de herkenning uit op PDF bestanden die enkel beelden bevatten en slaat de OCR resultaten op... in PDF *tekstdocumenten*! Moeten we het nog herhalen? Op tekst gebaseerde PDF bestanden zijn doorzoekbaar en bewerkbaar, PDF bestanden die enkel beelden bevatten zijn dat niet.

Tot slot is de conversie van PDF documenten een manier om hun inhoud toegankelijk te maken. U kunt namelijk “read-only” PDF bestanden waarvan de inhoud normaal gezien buiten uw bereikt valt, herkennen. Bij PDF documenten wiens inhoud niet beschermd is, kan u bij de inhoud komen (om die naar het klembord te kopiëren en in een tekstbestand op te slaan). Bij “read-only” bestanden kunt u geen gebruik maken van de inhoud. Dergelijke documenten kunnen namelijk enkel op het scherm worden bekeken en afgedrukt!

We voegen een belangrijke nuance toe: Readiris opent geen PDF documenten die met een paswoord beschermd zijn, ook al worden alle andere beveiligingen door Readiris gesloopt!

Ga te werk zoals gewoonlijk: laad PDF bestanden in het geheugen zoals u vooraf gescande beelden - faxen, snapshots die u met uw digitaal fototoestel nam enz. - inlaadt. Klik op de knop "Stoppen" of druk op de toets Escape om het laadproces tussen twee bladzijden te onderbreken.

Er is één optie die specifiek PDF bestanden betreft: u kunt die openen als kleurdocument of als zwart-wit document. Het nut van deze optie bestaat hierin dat het veel meer tijd kost om kleurdocumenten in bitmaps om te zetten dan het geval is bij zwart-wit documenten!



MEERDERE PAGINA'S HERKENNEN

Na de herkenning wordt het gescand document opnieuw afgebeeld met zijn bevenstering; het document is beschikbaar voor verdere verwerking.

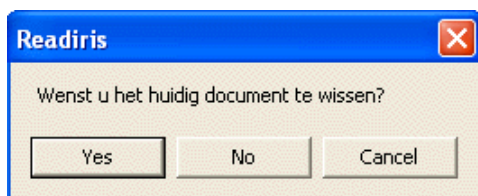
U kunt nu de herkende tekst openen met uw tekstverwerker of teksteditor, in uw DTP-software of een andere tekstapplicatie invoeren. Ga uw gang en vergelijk het tekstresultaat met het beeld dat door Readiris afgebeeld wordt.

Onze oefening van één bladzijde eindigt hier. Maar hoe slaat u de tekst op van de volgende bladzijden? Met andere woorden: hoe verwerkt u documenten die meerdere pagina's beslaan? Het is eigenlijk heel simpel: ga rustig verder met de herkenning van de volgende bladzijden en sla de resultaten op in hetzelfde tekstbestand. (Wees er dan zeker van dat die tekst op dat ogenblik niet geopend

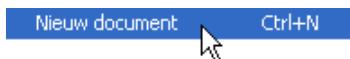


is want zoiets belet u in dat bestand te schrijven!) Vergeet ook niet het letterwoordenboek in de uitbreidingsmodus te plaatsen zodat u de training kunt verderzetten.

Van zodra u bladzijden scant (of beeldbestanden opent) *binnen* een document, moet u beslissen of u met een nieuw document wilt aanvangen of het huidige document wilt aanvullen.

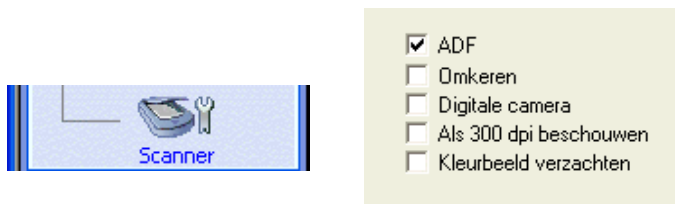


Kies "nee" om bladzijden toe te voegen aan het huidige document, kies "ja" om een nieuw document aan te maken. Dit laatste antwoord heeft hetzelfde effect als wanneer u het bevel "Nieuw Document" onder het menu "Bestand" gebruikt.



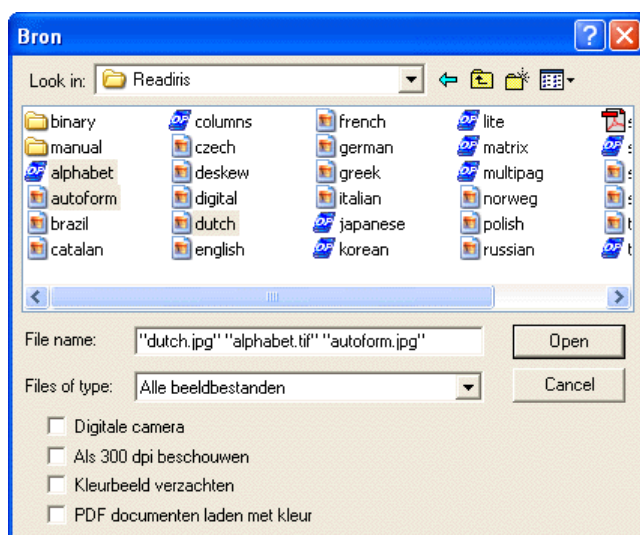
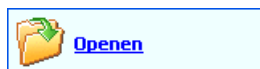
Maar er is een veel efficiëntere manier om meerdere bladzijden over te nemen dan die één voor één te herkennen. U kunt namelijk gewoon meteen **meerbladige documenten** verwerken.

Om een document van meerdere bladzijden te scannen, schakelt u de documentinvoer van uw scanner in met de optie "ADF" onder de "Scanner" knop.

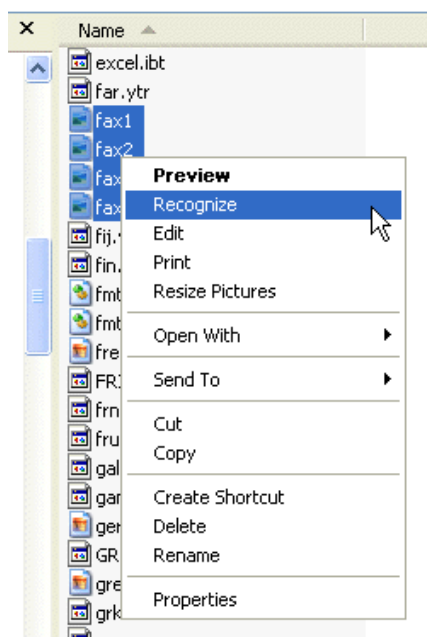


Leg de bladzijden in de documentinvoer en start het scannen: Readiris blijft scannen tot de documentdoorvoer leeg is.

U kunt ook meerdere vooraf gescande beeldbestanden *openen*. Om meerdere beeldbestanden in het geheugen te laden, selecteert u het eerste beeldbestand en houdt u de Ctrl toets ingedrukt terwijl u extra beeldbestanden selecteert. Om een ononderbroken reeks van beelden te selecteren, selecteert u het eerste beeldbestand en houdt u de Shift toets ingedrukt terwijl u het laatste bestand selecteert.



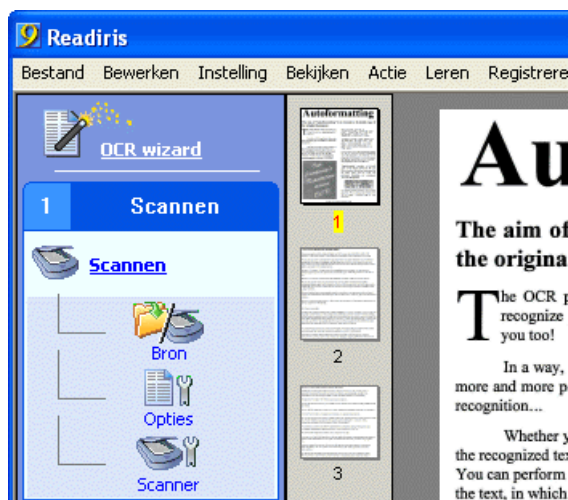
U kunt hetzelfde resultaat bereiken vanuit de Windows Verkenner: selecteer meerdere bestanden, klik met de rechter muisknop en kies het bevel "Herkennen" in het "Context" menu. U kunt deze operatie overigens herhalen: alle beeldbestanden die u zo in Readiris invoert vullen het huidig document aan tot u op het bevel "Nieuw Document" klikt!



U kan zelfs meerdere vooraf gescande beelden van de Windows Verkenner naar het Readiris venster *slepen*! En ook hier geldt: alle beeldbestanden die u naar Readiris sleept vullen het huidige document aan tot u het commando "Nieuw Document" aanklikt.

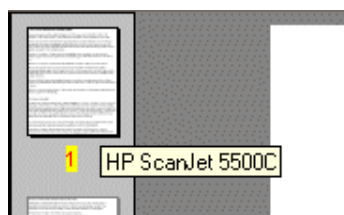
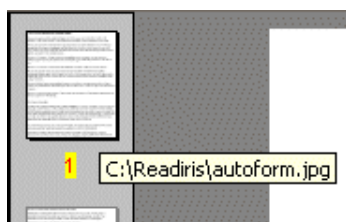
Readiris sorteert de beelden automatisch - beeld 001.tif gaat vooraf aan beeld 002.tif dat vooraf gaat aan beeld 003.tif enz.

De **paginaknoppenbalk** links van het beeld - deze knoppenbalk verschijnt van zodra er een bladzijde in het geheugen geladen wordt - stelt de verschillende pagina's van het document voor. Een klik van de rechter muisknop roept de bevelen op die specifiek de bladzijden betreffen.



De huidige bladzijde wordt beklemtoond in de paginaknoppenbalk en is aangegeven in de titelbalk van Readiris.

De paginaknoppenbalk is uitgerust met een tooltip: houd de muiscursor boven de miniatuur van een bladzijde en u verneemt aan welk beeldbestand deze pagina beantwoordt. (Indien het om een meerbladig beeldbestand gaat is er natuurlijk maar één bestand voor alle bladzijden.) Wanneer u meerbladige documenten *scant*, vermeldt de tooltip welk scannermodel u daarvoor gebruikt.





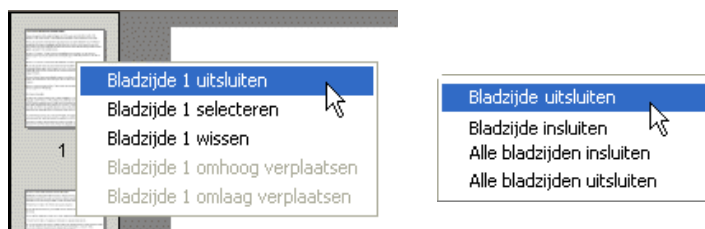
Mocht u een overzicht willen van uw document, dan kunt u vlug de gescande **beelden afdrukken** met het bevel "Beelden Afdrukken" onder het menu "Bestand".



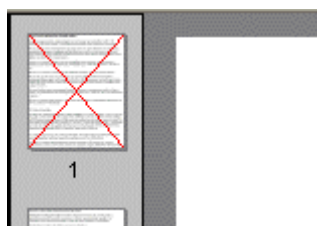
U kunt de huidige bladzijde of alle bladzijden afdrukken. Kies het aantal bladzijden of miniatures dat u op één pagina wenst af te drukken.



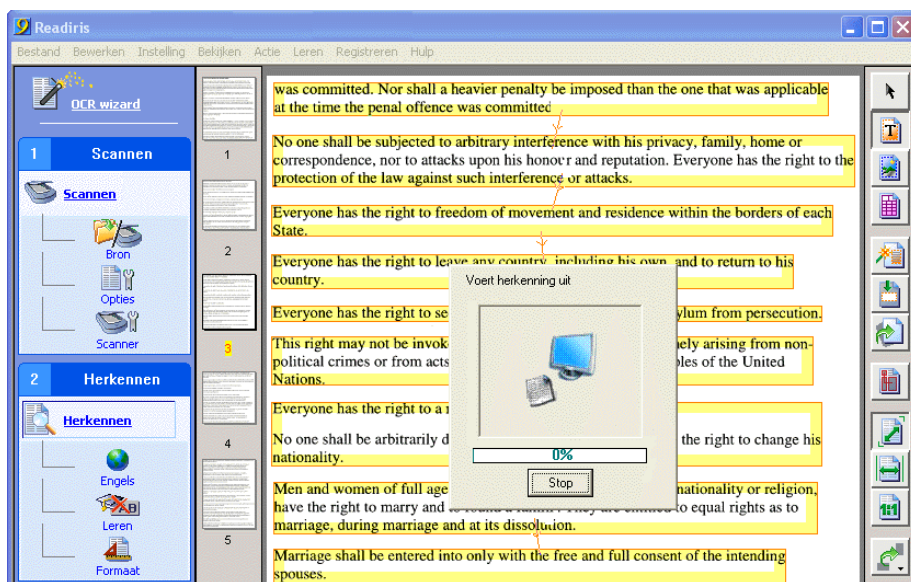
Niet dat u alle bladzijden moet afdrukken: met de paginaknoppenbalk (en het menu "Bewerken") kunt u bepaalde bladzijden (tijdelijk) negeren. Klik met de rechter muisknop op een paginaknoppenbalk en selecteer het bevel "Bladzijde Uitsluiten" op het "Context" menu of haal een bladzijde naar het scherm en selecteer het bevel "Bladzijde Uitsluiten" onder het menu "Bewerken" om het uit te sluiten van het druk- (en OCR-)proces. Selecteer het bevel "Bladzijde Insluiten" om de uitsluiting te annuleren. Omwille van de flexibiliteit biedt het menu "Bewerken" nog enkele handige commando's die op *alle* bladzijden slaan.



De miniatures van de uitgesloten bladzijden zijn doorgehaald. Let op: de huidige bladzijde drukken lukt altijd, zelfs al is die pagina op het ogenblik uitgesloten!



Open het beeldbestand MULTIPAGE.TIF en voer de herkenning uit. De bladzijden worden één na één afgebeeld, de titelbalk van Readiris geeft het paginanummer aan.



Als het interactief leren ingeschakeld is, doorloopt u de herkenning en de leerfase voor elke bladzijde. De woordenboekmodus "nieuw" wordt gebruikt voor de eerste bladzijde, de toevoegingsmodus voor de volgende bladzijden.

Door op de knop "Einde" te klikken worden alle oplossingen van het systeem aanvaard zonder validatie door de gebruiker. Met andere woorden, het interactief leren wordt dan afgebroken voor *alle* bladzijden, het OCR-proces gaat verder in de automatische modus.

Het herkenningsresultaat van meerbladige documenten wordt opgeslagen in één enkel uitvoerbestand. (Wanneer u de uitvoer naar een doeltoepassing stuurt, worden meerdere pagina's aangemaakt binnen één document.)

Tenminste, dat is het geval wanneer u de optie "Eén Document Creëren per Bladzijde" uitgeschakeld is op het ogenblik dat u het herkende document opslaat.

Deze optie zorgt er namelijk voor dat elk blad van een meerbladig document wordt opgeslagen in een apart bestand. Indien de gebruiker het bestand tekst.doc noemt, zullen de bestanden tekst-1.doc, tekst-2.doc enz. heten.

File name:

Save as type:

☐ Opslaan en openen

☐ Per e-mail versturen

☒ Eén bestand creëren per bladzijde
(Multipage-1, Multipage-2...)

MEERBLADIGE DOCUMENTEN BEWERKEN

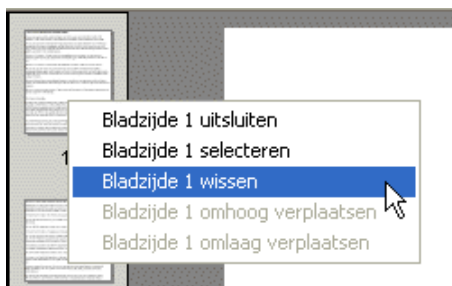
De gebruiker kan meerbladige documenten bewerken, hoofdzakelijk om scanfouten te corrigeren: hij kan bladzijden uit het document verwijderen en de volgorde van de pagina's binnen het document wijzigen.

Maar we beginnen met de navigatie. Om *naar een bladzijde te gaan*, klikt u op de bladzijde in de paginaknoppenbalk of houdt u de muiscursor boven de miniatuur van de bladzijde, roept u het "Context" menu op door met de rechter muisknop te klikken en gebruikt u het commando "Bladzijde Selecteren". Om naar de vorige bladzijde te verspringen kan u op de toets PageUp duwen, om naar de volgende bladzijde te gaan, op de toets PageDn. Druk op Home op naar de eerste bladzijde te gaan, druk op End op naar de laatste bladzijde te gaan. Of gebruik de overeenkomstige commando's uit menu "Bekijken".

Eerste bladzijde	Home
Vorige bladzijde	PageUp
Volgende bladzijde	PageDown
Laatste bladzijde	End

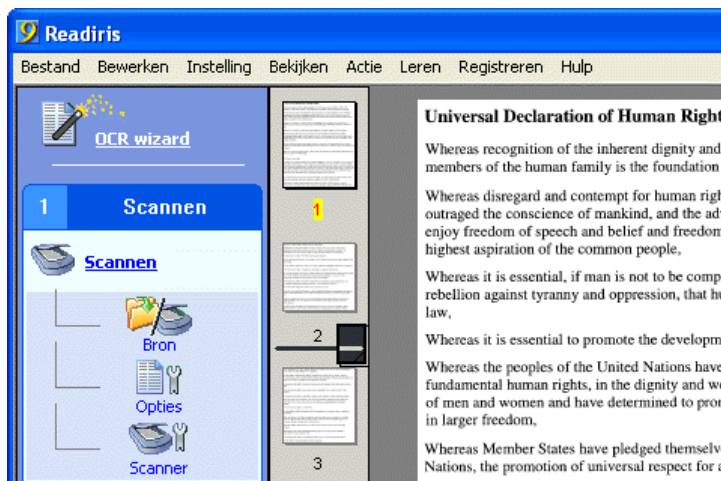


Het document wijzigen gaat als volgt. Om een *bladzijde* te *wissen*, houdt u de muiscursor boven de miniatuur van de bladzijde, roept u de "Context" menu op door met de rechter muisknop te klikken en gebruikt u het commando "Bladzijde Wissen". (We herinneren u eraan dat u specifieke bladzijden tijdelijk kunt uitsluiten - zonder ze te wissen - uit het herkenningsproces (en het afdrukken): de paginaknoppenbalk (en het menu "Bewerken") bevatten de nodige bevelen.



Om een *bladzijde naar boven* te *schuiven*, een hoger paginanummer te geven, gebruikt u het commando "Bladzijde Omhoog Verplaatsen", om een *bladzijde naar beneden* te *verschuiven*, gebruikt u het commando "Bladzijde Omlaag Verplaatsen".

Om een *bladzijde* te *verhuizen*, namelijk ze een heel andere plek te geven in het document, sleept u haar pictogram naar een andere locatie.



MET EEN NIEUW DOCUMENT BEGINNEN

Met het bevel "Nieuw Document" onder het menu "Bestand" sluit u het huidige document af.



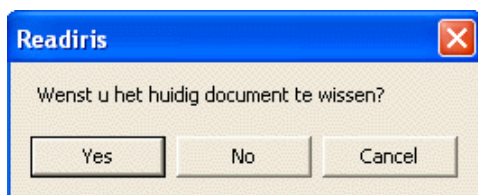
Dit commando wist alles uit het geheugen. Het in het geheugen geladen document (dat één of meerdere bladzijden bevat) wordt gewist. U bent nu klaar om een nieuw document aan te maken.

Maar u kunt ook met een nieuw document beginnen vanuit het huidige document. Zolang de herkenning niet werd uitgevoerd, veronderstelt het systeem dat u nog bladzijden aan het document wilt toevoegen. U kunt bijvoorbeeld alle bladzijden in de documentinvoer inscannen, de documentdoorvoer bijvullen en opnieuw beginnen. Of u scant enkele documenten en voegt daar enkele beeldbestanden, zeg maar faxen, aan toe. Deze bladzijden vormen één docu-



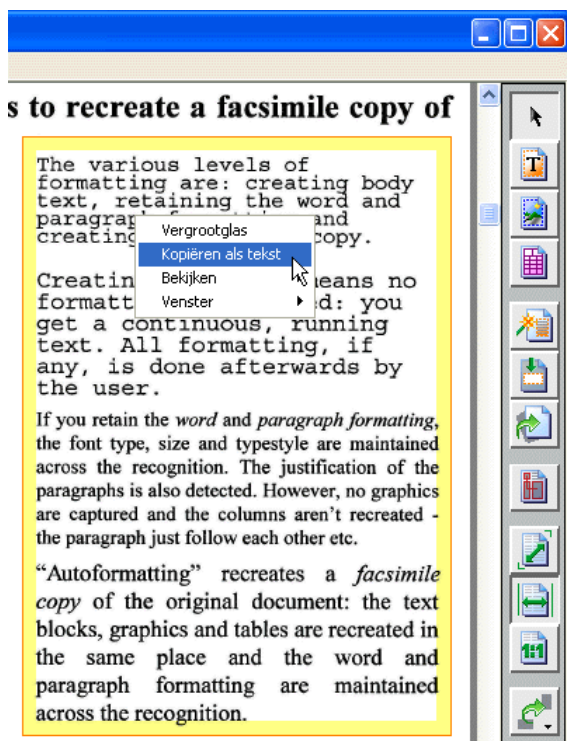
ment, het enige wat u moet doen is tussendoor de beeldbron wijzigen met de knop "Bron".

Wanneer de herkenning wel al was uitgevoerd en u begint opnieuw te scannen (of beelden te laden), zal het systeem u vragen of u met een nieuw document wilt beginnen of met het huidige document wilt verdergaan.



TEKSTZONES HERKENNEN

We weten nu hoe men een bladzijde herkent en hoe men meerbladige documenten verwerkt. Maar kunnen we ook makkelijk delen van een tekstbladzijde herkennen? Dat kan: klik met de rechter muisknop en selecteer het bevel "Kopiëren Als Tekst" in het "Context" menu: het tekstvenster onder de muiscursor wordt herkend en het tekstresultaat wordt in het klembord gestopt.



De huidige systeeminstelling - taal, type letterteken enz. - is van toepassing. Het OCR-resultaat wordt naar het klembord gestuurd als lopende, niet geformatteerde tekst.

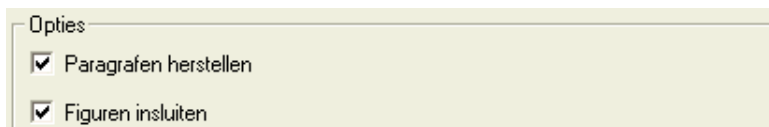


DE UITVOER OPMAKEN

Een tekst opslaan of uitvoeren houdt meer in dan een uitvoermethode kiezen of een bestandsnaam opgeven voor de tekstuitvoer. U moet ook een tekstformaat selecteren en de lay-out van de herkende tekst gaan bepalen. Kortom, u moet beslissen waar u met de tekst heen wilt eer u de herkenning uitvoert.

Sommige opties van de "Formaat" knop laten u toe om het uitzicht van de uitgevoerde tekst te beïnvloeden.

De “**tekstflow**” van de uitvoer wordt onmiddellijk beïnvloed door de optie "Paragrafen Herstellen".



Schakel deze optie in om Readiris de **paragrafen** te laten **detecteren**: Readiris zal dan de normale “wordwrap” toepassen die typisch is voor tekstverwerkers, zoniet wordt een carriage return (Enter) toegevoegd na iedere lijn en blijven gesplitste woorden gesplitst! De paragraafdetectie is standaard actief.

Laten we een voorbeeld geven om één en ander te verduidelijken. Wanneer de eerste drie lijnen van een kolom "De nieuwe presi-", "dent wuift van op het terras." en "Zijn vrouw vergezelt hem." zijn, zorgt de paragraafdetectie voor het volgende resultaat: "De nieuwe **president** wuift van op het terras. **Zijn** vrouw vergezelt hem." De twee stukken van het woord "president" die door een koppelteken gesplitst zijn, zijn opnieuw aaneengelijmd, en er werd een spatie ingevoegd op het einde van de eerste zin. Kortom, u krijgt een vlotte, continue tekst.

Was de paragraafdetectie *niet* geactiveerd, dan zou de originele lay-out behouden zijn, met een carriage return (Enter) aan het einde van elke zin.

Deze optie is *niet* beschikbaar wanneer u een PDF formaat kiest: Adobe Acrobat bestanden slaan tekst altijd lijn per lijn op.

(De "Formaat" knop bevat tevens formatteeropties die we nog niet besproken hebben - we zullen dit straks doen.)

UW SCANNER INSTELLEN

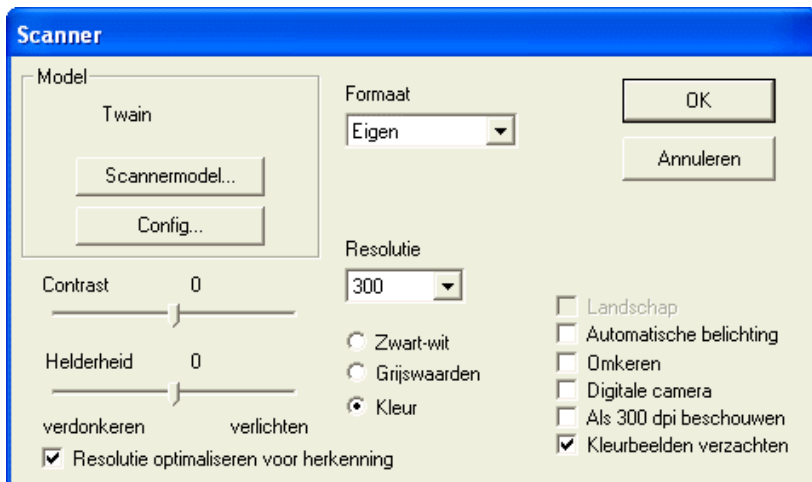
Laten we uw scanner gaan instellen. We veronderstellen nu wel dat uw scanner en de nodige stuurbestanden goed werden geïnstalleerd.

Als uw Readiris licentie gebundeld is met een scanner, kan deze stap overbodig blijken omdat uw scanner dan allicht al onder Readiris geconfigureerd is.

Klik op de "Scanner" knop in de hoofdknoppenbalk.



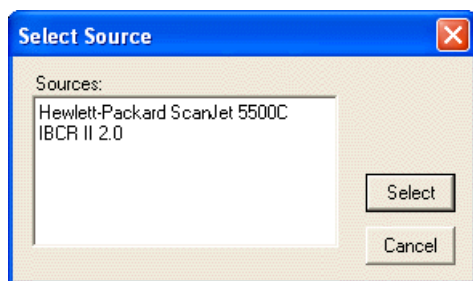
Klik op de knop "Scannermodel" om uw **scannermodel** te kiezen.





Wanneer u de optie "<Image>" (beeld) als "scanner" selecteert, fungeren vooraf gescande beelden ten allen tijde als beeldbron – u hoeft zelfs de schijf niet te selecteren met de knop "Bron" op de hoofdknoppenbalk.

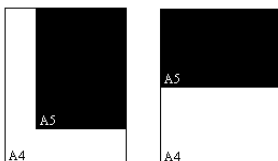
De knop "Configureren" is enkel beschikbaar wanneer uw scanner het toelaat. Hij geeft toegang tot sommige geavanceerde scanparameters; bij Twain scanners kunt u deze knop gebruiken om de Twain bron te selecteren. (U kunt dat ook met het commando "Bron Selecteren" onder het menu "Bestand".)



Wanneer de scanner geselecteerd is, zal hetzelfde venster u toelaten om de scanresolutie, het papierformaat en -oriëntatie, de helderheid en het contrast in te stellen. U kunt tevens aangeven of u van de documentinvoer wilt gebruik maken. Bij Twain scanners worden alle instellingen meestal binnen de Twain interface geregeld.

Stel de **helderheid** in en, wanneer dat mogelijk is, **het contrast**.

Door de optie "Landschap" te selecteren geeft u aan dat de pagina-oriëntatie liggend is en niet staand ("portret"). Deze pagina-oriëntatie is eigenlijk enkel van toepassing op gereduceerde formaten: bij een A4 flatbed scanner kunt u bijvoorbeeld A5 bladzijden (de helft van een A4 blad) in portret- en in landschaporiëntatie scannen, maar volle A4 bladzijden kunt u natuurlijk alleen in één richting inscannen!



De optie "Omkeren" tenslotte laat de scanner volledig **“negatieve” beelden** laten aanmaken in de zwart-wit scanmodus - activeer deze optie om bladzijden te verwerken met witte tekst op een zwarte achtergrond.

GEEF UW SCANS WAT KLEUR MEE!

Readiris ondersteunt zwart-witte scans, scans in grijswaarden en in kleur. U kan dus de **kleurmodus** kiezen die u het best bevalt. Om zwart-wit illustraties in de documenten in te sluiten, scant u in zwart-wit, om zwart-wit foto's over te nemen, scant u in grijswaarden en om kleurfoto's over te nemen, scant u in kleur.

Maar waarom zou u de bitdiepte van de beelden bij het scannen herleiden? Het spreekt voor zich dat het scannen van grijswaarden en kleur trager is en meer RAM-geheugen vereist dan zwart-wit beelden.

Grijswaarden en kleur gebruiken zijn niet enkel nuttig om de figuren met voldoende kwaliteit te hernemen; in een aantal gevallen hebt u grijswaarden of kleuren gewoon nodig voor een goede herkenning! Bij teksten die afgedrukt zijn op een gekleurde achtergrond kan een kleurenscaan de kleurverschillen creëren die in het zwart-wit beeld afwezig zijn. Wanneer er slechts weinig contrast is tussen de tekst en de achtergrond, kan de achtergrond “ruis” opleveren die de herkenning bemoeilijkt of gewoon onmogelijk maakt!

Nemen we bijvoorbeeld een zwarte tekst op een donkere achtergrond. Hoezeer u de helderheid ook bijstelt, wanneer u zo'n document in zwart-wit scant zult u er allicht niet in slagen de achtergrond te filteren zonder dat de tekst eveneens wegvalt.



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 creëert een zwart-witte kopie van elk kleur- en grijswaardenbeeld. De software maakt daarbij gebruik van intelligente routines, zodat zelfs de moeilijke gevallen worden opgelost. Hieronder ziet u hoe ons "lastig" beeld wordt omgezet...

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.

Om de zwart-witte versie van een gescand beeld af te beelden, schakelt u de optie "Document in Kleur Afbeelden" onder het menu "Bekijken" uit.

✓ Document in kleur afbeelden Ctrl+O



Wanneer deze optie actief is, zult u nooit zwart-wit beelden op het scherm zien - zelfs niet wanneer u eigenlijk zwart-witbeelden scant! Dat komt hierdoor: de optie "Afbeelden met Hoge Kwaliteit" uit het menu "Bekijken" zorgt voor optimaal leesbare beelden op uw computerscherm.

✓ Afbeelden met hoge kwaliteit



Deze gespecialiseerde hoge-resolutie displaytechniek zet zwart-wit beelden om in grijswaardenbeelden.

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

Grijswaarden- en kleurbeelden daarentegen worden zachter gemaakt.

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.

Dit alles zorgt ervoor dat u niet hoeft in te zoomen, zelfs op een draagbare PC met een LCD scherm of op een desktop-PC met een bescheiden 14" scherm. De beeldweergave met hoge kwaliteit is standaard ingeschakeld, maar deze optie kan overbodig blijken op schermen met een hoge resolutie.

ANDERE APPARATEN, ANDERE RESOLUTIE

Wat uw kleurmodus ook moge wezen, kies een **resolutie** van 300 dpi voor normale toepassingen en de hogere resolutie van 400 dpi voor kleine druk (kleiner dan 10 punt) en wanneer het document erg gedegradéerd is.

Readiris leest **puntgroottes** van 6 tot 72 punt (0,21 tot 2,54 cm.).

6 punt

72

Readiris leest eveneens **kapitalen** ("dropletters"), hoofdletters die meerdere lijnen beslaan. (Die kunnen uiteraard niet groter zijn dan 72 punt!)

Readiris leest "drop letters" die meerdere lijnen bedekken en kent die toe aan de eerste lijn van de paragraaf.

Daar optimale OCR een resolutie vereist tussen 300 dpi en 400 dpi, waarschuwt Readiris u wanneer u beelden poogt te herkennen met een resolutie die lager ligt dan 200 dpi of hoger dan 800 dpi. Merk op dat Readiris scans met teveel details automatisch kan corrigeren! Activeer hiervoor de optie "Resolutie Optimaliseren voor Herkenning" in de scannerinstelling. Van zodra de beeldresolutie van uw scans 600 dpi overschrijdt, wordt hun resolutie automatisch verminderd voor de herkenning.

☒ Resolutie optimaliseren voor herkenning

Er zijn andere manieren om deze waarschuwing te vermijden: u leest bijvoorbeeld **faxen** - die hebben een resolutie van 100 of 200 dpi -, u herkent beelden die afkomstig zijn van een digitale foto-toestel - de resolutie is dan onbekend -, of u opent beeldbestanden waarvan de hoofding een verkeerde resolutie aangeeft. Om zulke beelden succesvol te verwerken activeert u de optie "Als 300 dpi Beschouwen". Deze parameter geldt evengoed voor rechtstreekse scans als voor vooraf gescande beelden die u in het geheugen laadt.

- ☐ ADF
- ☐ Omkeren
- ☐ Digitale camera
- ☒ Als 300 dpi beschouwen
- ☐ Kleurbeeld verzachten

Files of type:

- ☐ Digitale camera
- ☒ Als 300 dpi beschouwen
- ☐ Kleurbeeld verzachten
- ☐ PDF documenten laden met kleur

Wanneer de beelden werden aangemaakt met een **digitale camera** in plaats van een scanner, moet u een speciale optie inschakelen. Ook deze optie is van toepassing op rechtstreekse scans en beelden die u opent.

- ☐ ADF
- ☐ Omkeren
- ☒ Digitale camera
- ☒ Als 300 dpi beschouwen
- ☐ Kleurbeeld verzachten

Files of type:

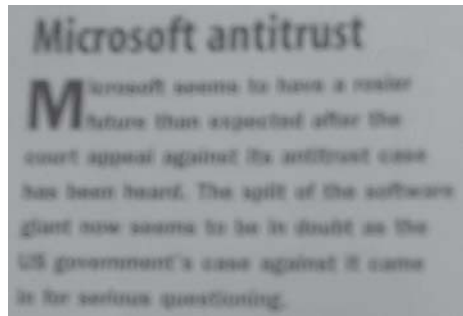
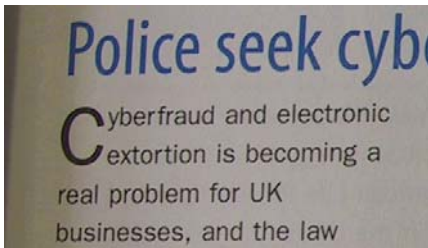
- ☒ Digitale camera
- ☒ Als 300 dpi beschouwen
- ☐ Kleurbeeld verzachten
- ☐ PDF documenten laden met kleur



Door dit te doen verbetert u de kwaliteit van het beeld eer de herkenning plaatsvindt. Beelden die afkomstig zijn van digitale fototoestellen stellen OCR-programma's op proef: digitale camera's produceren beelden met een lage resolutie - zelfs wanneer de u camera erg dicht boven uw document houdt - en de beeldresolutie is bovendien onbekend.

Er zijn enkele "details" waarvan u hoogte moet nemen om beelden die met een digitale camera werden genomen succesvol te herkennen.

Selecteer eerst en vooral de hoogst mogelijke resolutie. Schiet bijvoorbeeld plaatjes van 2.048 x 1.536 pixels wanneer u ook beelden van 1.024 x 768 pixels en van 640 x 480 pixels kunt schieten. Activeer tevens de "macromodus" van uw camera zodat u closeups neemt. Documenten fotograferen betekent immers altijd dat u closeups neemt. (Deze modus werd eigenlijk ontworpen om bloemen, insecten enz. te fotograferen.) Als u dat niet doet zijn de beelden zo onscherp dat ze onleesbaar worden.

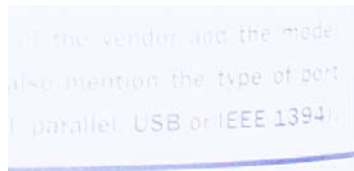


Pas weinig of geen compressie toe: significante compressie herleidt de scherpte van de genomen foto. Zoom handmatig om uw document te omkaderen; sommige camera's zijn uitgerust met software die de aparte stukken van foto's aan elkaar "naait" ("photo stitching"), maar het loont niet de moeite dit te doen bij OCR-toepassingen.

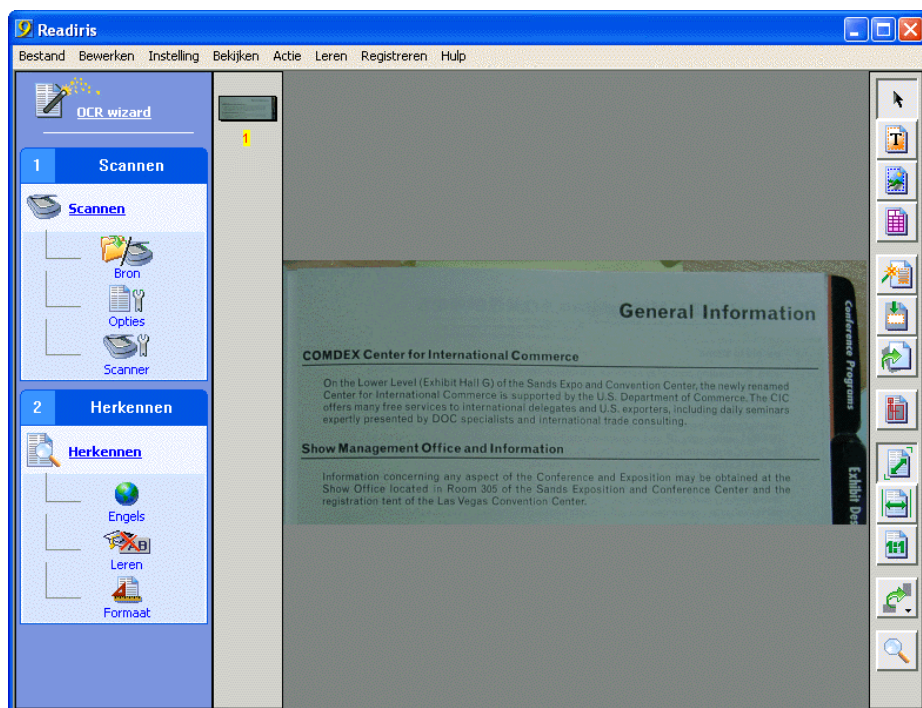
Houd de camera recht boven het document, scan het document niet onder een hoek. Maar probeer wel om met uw handen geen schaduw op het document te

werpen! Maak stabiele beelden aan. U kunt de camera desnoods op een statief plaatsen.

Schakel het flitslicht uit wanneer u glanzend papier fotografeert, zoniet kan het beeld veel te licht uitvallen. Door de bank genomen kunt de helderheid en het contrast van een digitale camera aan de omgeving aanpassen - er zijn vaak aparte instellingen voor daglicht, kunstlicht, neonlicht enz. (Sommige camera's kunt u zelfs calibreren door een wit document te fotograferen!)



Als u het even wilt proberen, kunt u het beeldbestand **DIGITAL.JPG** in de Readiris map openen en de herkenning uitvoeren.



STANDAARD INSTELLING BEWAREN

Stel de scanparameters goed in en klik op het commando "Standaard Instelling Opslaan" onder het menu "Bestand" om de actuele instellingen te bewaren als standaard **instelling** voor verder gebruik.

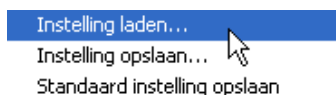
Standaard instelling opslaan



Configuratiebestanden bevatten meer dan enkel de scanopties: ze bepalen ook of u het interactief leren gaat gebruiken, welke taal de documenten bevatten, welke uitvoermethode en lay-outopties u gebruikt - bijvoorbeeld lopende tekst naar WordPad sturen enz. Kortom, *alle* operationele parameters van Readiris worden in de instellingen bewaard.

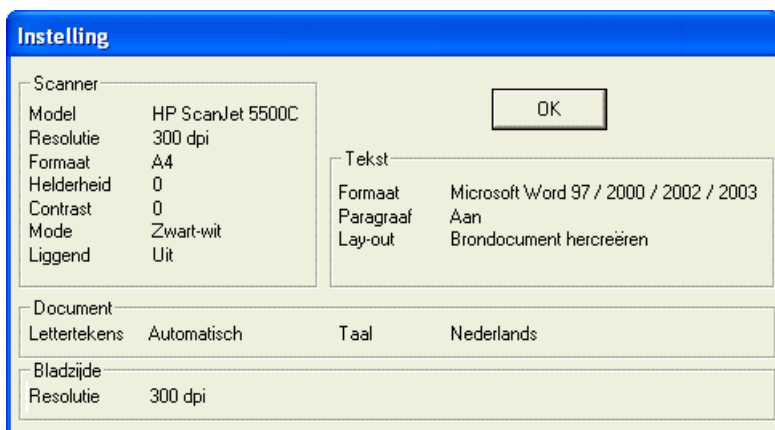
SPECIFIEKE INSTELLINGEN BEWAREN

De standaard instelling zal natuurlijk gebruikt worden telkens u Readiris opstart, maar u kunt specifieke configuraties bewaren om te vermijden dat u operationele parameters moet herdefiniëren. De opdrachten "Instelling Opslaan" en "Instelling Laden" onder het "Bestand" menu nemen deze taak voor hun rekening.



Laten we een voorbeeld geven: als u regelmatig Duitse documenten moet herkennen met een specifieke lay-out, dan wordt u aangeraden om een configuratiebestand aan te maken voor dit documenttype. U zult dan met name "Duits" als documenttaal selecteren, een aangepaste vensterlay-out laden om te vermijden dat u steeds dezelfde vensters moet hertekenen, het leren uitschakelen maar een woordenboek laden in de leesmodus omdat telkens dezelfde lettertypes terugkeren enz.

Als u twijfels heeft over de huidige instelling, hoeft u niet in elk menu en elk commando te duiken om ze te achterhalen. U kunt gewoon het commando "Informatie" onder het "Bestand" menu aanklikken om een praktisch overzicht te krijgen.



DOCUMENTEN SCANNEN

Nu de scanner is ingesteld, zijn we klaar om te gaan scannen. Wat het scannen betreft, moeten we wel op enkele punten wijzen.

Besteed enige aandacht aan de regelkromming. De pagina-analyse en de herkenning springen wel tolerant om met schuine regels, maar het kan fout aflopen wanneer het document te schuin werd ingescand. Beperkte regelkromming (die kleiner is dan $0,5^\circ$) kunt u negeren omdat die geen invloed heeft op de herkenning.

De optie "Bladzijden Rechtzetten" onder de knop "Opties" (en onder het menu "Instelling") bepaalt of die documenten die schuin werden ingescand automatisch zullen **rechtgezet** worden. Beperkte regelkromming wordt daarbij genegeerd; deze optie is standaard uitgeschakeld.



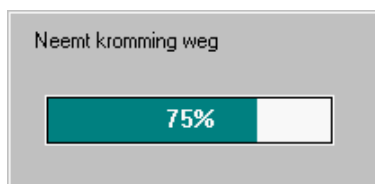
Indien u vergat deze optie in te schakelen, gebruikt u de knop "Bladzijde Rechtzetten" op de beeldknoppenbalk (of het overeenkomstig commando onder het menu "Actie") om de regelschuinte te corrigeren.



Bladzijde rechtzetten



Een blad rechtzetten duurt enkele seconden. Eerst wordt het beeld geanalyseerd om de hoek van de regelschuinte te detecteren - als die er is tenminste -, daarna wordt het kleur- of grijswaardenbeeld én de zwart-witte versie rechtgezet en wordt de pagina-analyse opnieuw uitgevoerd.



Het kan ook nodig zijn dat u de pagina-oriëntatie moet aanpassen. Maak hiertoe gebruik van de rotatieknoppen op de beeldknoppenbalk. (Overeenkomstige opdrachten bevinden zich in het menu "Bekijken".) Er zijn drie rotatierichtingen: rotatie naar links, naar rechts en met 180°. De **rotatie** neemt een paar seconden in beslag: het achterliggende beeld wordt herschreven, en niet enkel de afbeelding ervan op het scherm.



Links roteren

Rechts roteren

180 graden roteren

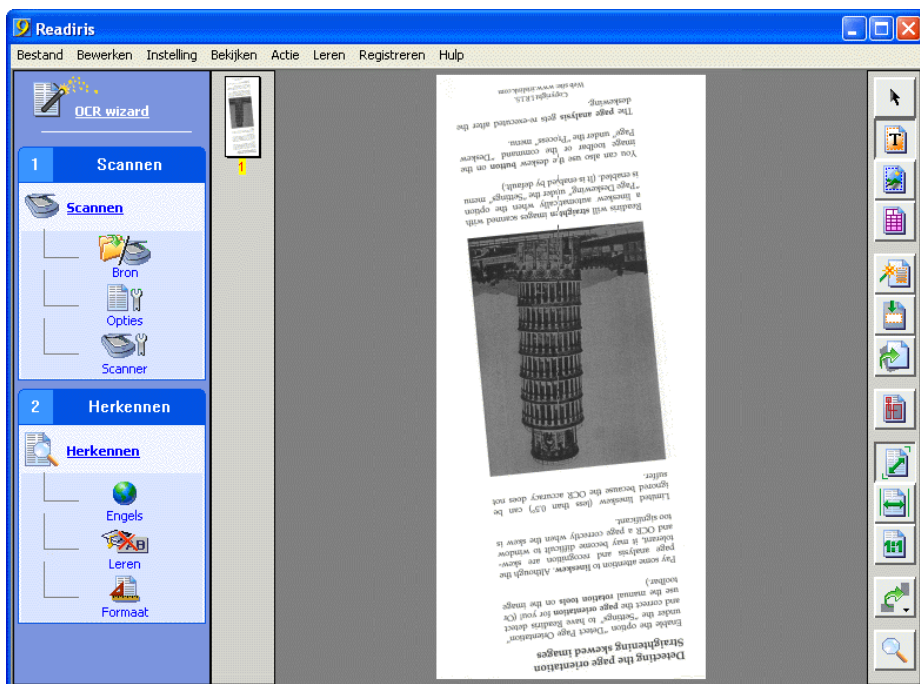




Merk overigens op dat Readiris bladzijden met een verkeerde oriëntatie voor u kan rechtzetten. Schakel de optie "Oriëntatie Bladzijden Detecteren" in met de knop "Opties" (of onder het menu "Instelling") en Readiris zal wanneer dat nodig is de richting van de bladzijde corrigeren.



U kunt dit alles eens proberen met het beeldbestand DESKEW.JPG in de Readiris map. Schakel de opties "Bladzijden Rechtzetten" en "Oriëntatie Bladzijden Detecteren" in eer u het beeld laadt en Readiris zal de toren van Pisa “herstellen” zoals we die graag zien.



DE GESCANDE BEELDEN AANPASSEN

Zoals al aangestipt werd, worden krachtige, intelligente routines gebruikt om kleur- en grijswaardenbeelden automatisch om te zetten in zwart-wit beelden. Maar de gebruiker kan, mocht dit nodig blijken, het beeld verder optimaliseren voor het OCR-proces. Dit gebeurt met het bevel "Beeld Aanpassen" onder het menu "Actie".



Beeld aanpassen... Ctrl+J

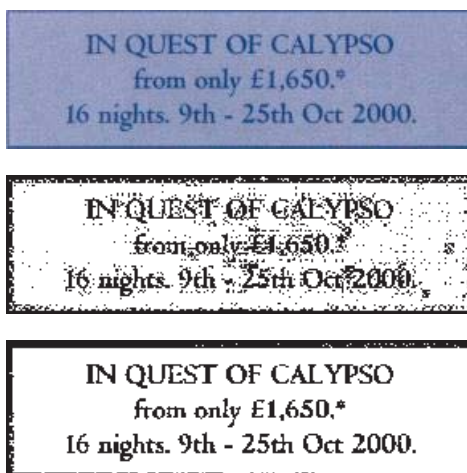


Wanneer u dit commando aanklikt, wordt het zwart-wit beeld automatisch op het scherm gebracht, het is alsof u de optie "Document in Kleur Afbeelden" uitschakelde. Dit bevel hanteert enkele moeilijke begrippen, en we zullen er in detail op ingaan.



De optie "Kleurbeeld Verzachten" maakt de kleur- en grijsbeelden homogener door ze "af te vlakken", door beperkte verschillen in intensiteit te weg te nemen. Het resultaat van deze bewerking is dat er een groter contrast wordt gecreëerd tussen de voorgrond - de tekst - en de achtergrond - een steunkleur, foto enz.

Deze optie **optimaliseert** het beeld voor de herkenning. Deze functie ziet er misschien erg technisch uit en moeilijk om te begrijpen, maar ze heeft wel degelijk haar belang: bij sommige scannermodellen moeten we de scherpte verminderen om kleur- en grijswaardenscans succesvol te herkennen. De beelden verzachten is soms de enige manier om tekst te onderscheiden van de gekleurde achtergrond! Hieronder ziet u een voorbeeld dat zonder verzachting gewoon onleesbaar is.



Beelden verzachten kan ook wanneer u voorafgescande beeldbestanden in het geheugen laadt.

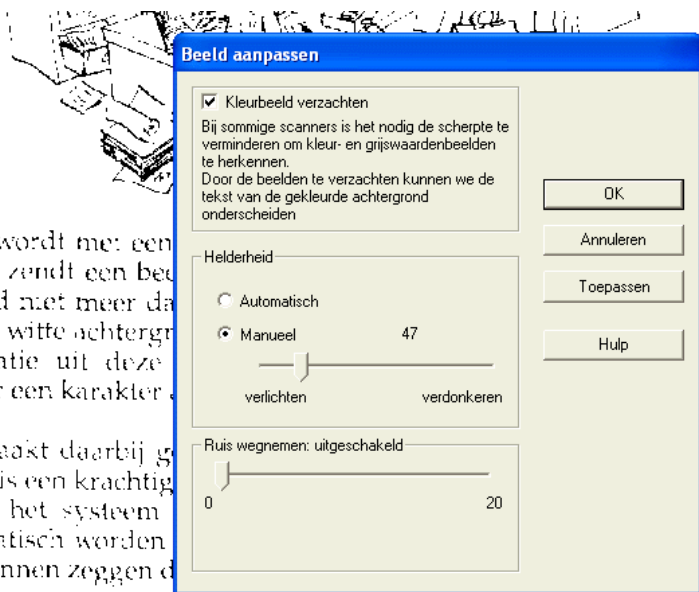


En wat bedoelen we met de helderheid? Met "helderheid" bedoelen we eigenlijk de zwart-wit drempel. De optie "Automatisch" bepaalt de zwart-wit drempel automatisch. Pas desnoods een andere drempel toe door het zwart-wit beeld te



verdonkeren of te verhelderen. Wanneer u het beeld verdonkert, worden meer pixels in de zwart-wit versie zwart, wanneer u het beeld verheldert, worden minder pixels in het zwart-wit beeld zwart.

Let wel, het beeld wordt pas echt aangepast wanneer u de knop "Toepassen" aanklikt. Door op "OK" te klikken past u het beeld aan én sluit u het venster. Hier hebt u een voorbeeld waarbij het zwart-wit beeld dramatisch werd verlicht - zij het dat we geen goede OCR-resultaten verwachten van dit beeld!

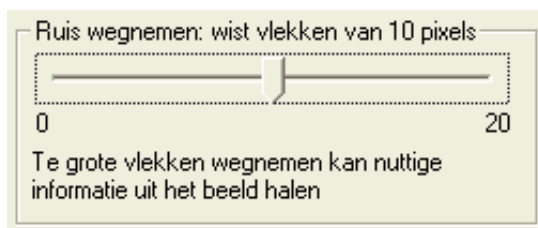


De eerste twee opties betreffen kleur- en grijsbeelden. De laatste optie, "Ruis Wegnemen", slaat exclusief op de zwart-wit beelden. "Ruis" wegnemen betekent dat de toevallige, "parasitaire" pixels uit het zwart-wit beeld worden verwijderd.

**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.**

Zorg ervoor dat u geen te grote vlekken uit het beeld verwijdert, want het zou wel eens kunnen voorkomen dat u de punten op de "i" wegneemt of delen van matrix letters wist!

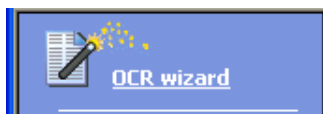


Dit is de beste manier om een beeld te optimaliseren: plaats het venster waarin u het beeld aanpast in een hoek van het scherm zodat u goed kunt beoordelen welk effect de aanpassingen die u uitvoert hebben op het beeld. Pas de parameters aan, klik telkens op "Uitvoeren" tot u een scherp en "proper" beeld op het scherm hebt.

LAAT DE WIZARD VOOR U WERKEN

Laten we van start gaan met het scannen van documenten. In plaats van alle parameters te doorlopen zullen we echter gebruik maken van de **OCR assistent**, een erg makkelijke manier om bladzijden te herkennen.

Klik op de knop "OCR Wizard" op de hoofdknoppenbalk of selecteer het commando "OCR Wizard" onder het menu "Actie".



OCR wizard...

Ctrl+D

De assistent gidst u comfortabel doorheen de herkenning: het volstaat op enkele eenvoudige vragen te antwoorden om snelle resultaten halen met Readiris.



Merk op dat de OCR assistent begint te lopen telkens u Readiris opstart; u kunt dit vermijden door de optie "Wizard Activeren bij Opstarten" in het eerste scherm van de wizard uit te schakelen. (U kunt ook de gelijkwaardige optie onder het menu "Instelling" gebruiken.)

READIRIS HERSCHEPT DE ORIGINELE LAY-OUT

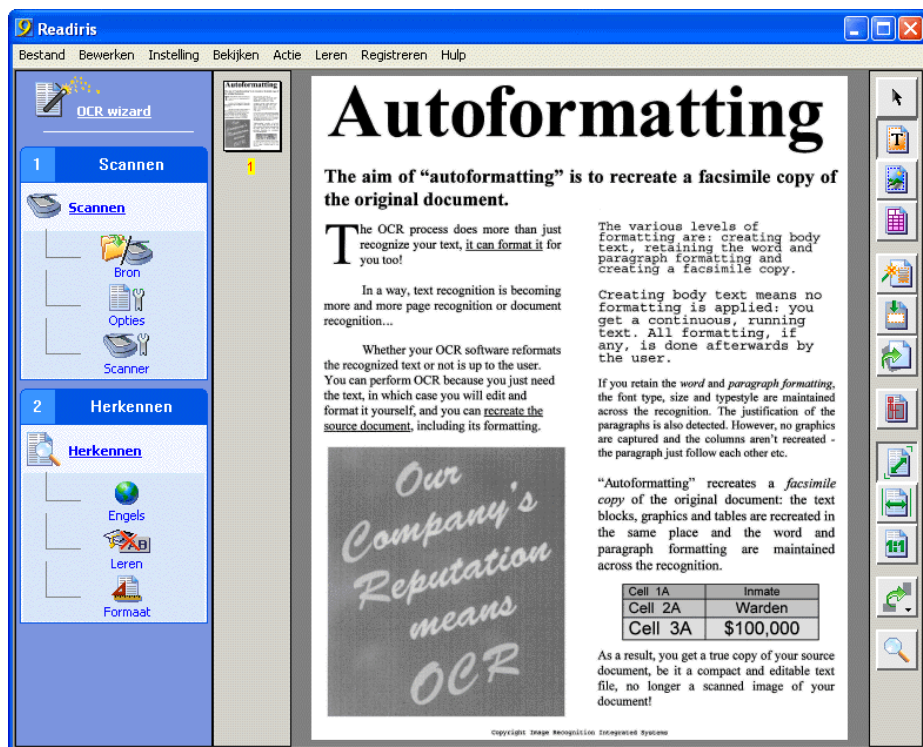
De OCR assistent automatiseert de herkenning in hoge mate, maar u mag deze “automatische” herkenning niet verwarren met automatische formattering! “Autoformattering” houdt met name in dat Readiris een **getrouwe kopie** aanmaakt van het gescande document: de woord-, paragraaf- en bladformattering van het oorspronkelijk document keren terug in de uitvoer.

Gelijkaardige lettertypes als in het brondocument (met schreef of schreefloos, proportioneel of vast, normaal of samengedrukt) worden gebruikt, de puntgroottes en stijlen (vet, cursief en onderlijnd, superscript en subscript) worden behouden tijdens de herkenning. De tabulatie en de uitlijning (links, gecentreerd, rechts en uitgelijnd) van elke tekstblok worden gerecreëerd. Ook de opsommingstekens (“bullets”) en de nummering worden herschapen. Alle e-mailadressen en URLs van webbladzijden worden gedetecteerd en als hyperlinks in de uitvoer geplaatst. De plaats van de kolommen, tekstblokken en figuren is gebaseerd op het origineel document.

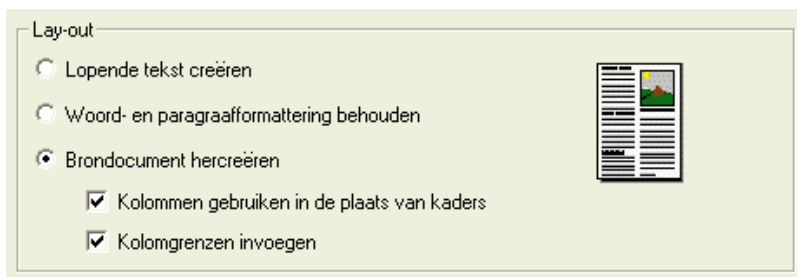
Met andere woorden, Readiris laat u toe om een kopie te archiveren van uw documenten, zij het dat u dit keer te maken hebt met een editeerbaar, compact tekstbestand, en niet langer met een gescand beeld!

Dit alles houdt in dat bij “autoformattering” de sortering van vensters slechts *gedeeltelijk* geldt: u kunt zones insluiten en uitsluiten, maar als u de sorteerorde wijzigt, wordt dat genegeerd!

Laten we bekijken hoe dit in zijn werk gaat. Om vertrouwd te raken met “autoformattering” opent u best het beeld AUTOFORMAT.JPG in de Readiris map.

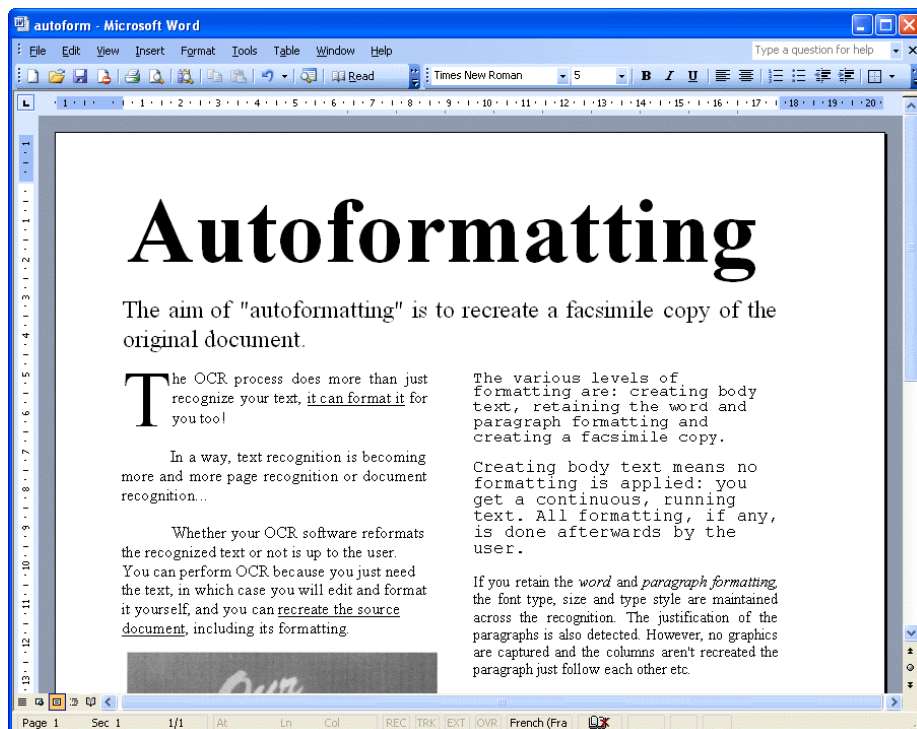


Klik op de "Formaat" knop op de hoofdknoppenbalk en kies ervoor het herkenningsresultaat naar Microsoft Word te zenden of selecteer het uitvoerformaat Word (DOC) of RTF (Rich Text Format). Activeer eveneens de lay-outoptie "Brondocument Hercreëren". (De optie "Paragrafen Herstellen" wordt standaard ingeschakeld om binnen de paragrafen "wordwrap" toe te passen.)

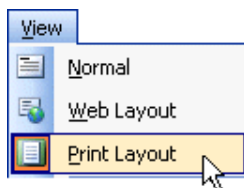


Of de opmaakreconstructie beschikbaar is, hangt af van de gekozen uitvoermodus. Sommige “arme” tekstformaten die “platte” tekst (“plain text”) genereren zoals Tekst (ANSI), MS-DOS Tekst (ASCII) enz. ondersteunen geen geavanceerde formattercodes en kunnen dan ook geen “autoformatting” bieden. Het Adobe Acrobat PDF formaat aan de andere kant werd specifiek ontworpen om de lay-out van uw documenten te bewaren: PDF documenten impliceren dus van nature uit autoformatting!

Wanneer de herkende tekst geopend wordt met een heuse tekstverwerker ziet hij er als volgt uit zonder enige interventie van de gebruiker.



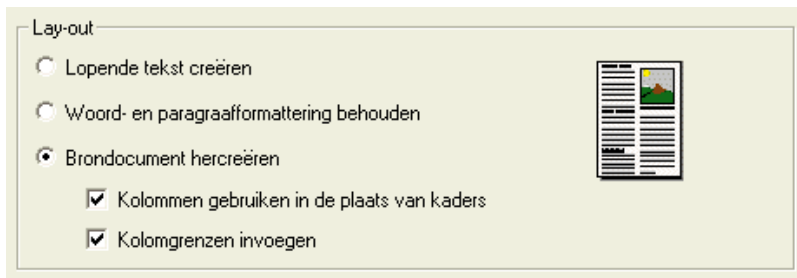
Om de lay-out correct te beoordelen moet u wel de “WYSIWIG”-modus van uw tekstverwerker activeren; meestal is er sprake van een “pagina-lay-out” optie. Let wel: wanneer u het herkende document rechtstreeks naar Microsoft Word zendt, wordt de “pagina-lay-out” optie automatisch geactiveerd!



Samengevat, Readiris herkent niet alleen uw teksten, maar kan ze ook formatteren. OCR is tegenwoordig meer dan tekstherkenning, OCR wordt meer en meer blad- of **documentherkenning**!

KOLOMMEN AUB, GEEN KADERS!

De formatteeroptie "Kolommen Gebruiken in de Plaats van Kaders" bepaalt *hoe* de “autoformattering” plaatsvindt: de tekstblokken, tabellen en figuren kunnen in kaders of in editeerbare **kolommen** worden opgeslagen.



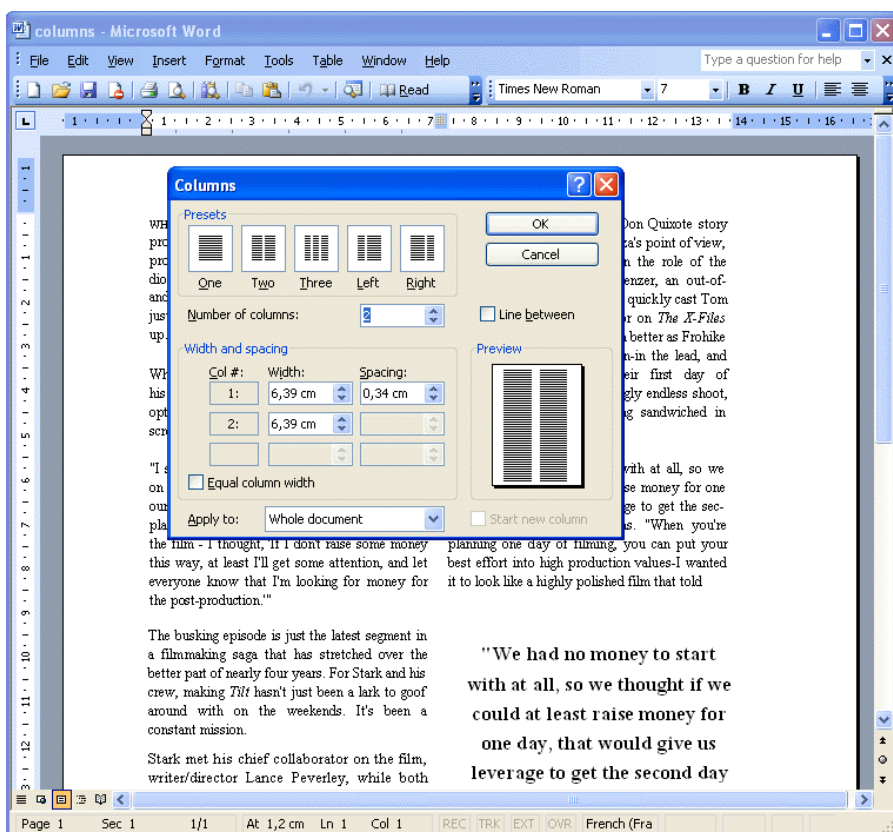
Kaders zijn aparte “containers” zodat je meerdere tekstblokken, figuren en tabellen op een bladzijde kan plaatsen. Bij kolommen ligt dat anders: de tekst vloeit natuurlijk van de ene kolom over in de volgende kolom. Tekst die in kolommen is geschikt is veel makkelijk te wijzigen dan tekst die over meerdere kaders is verdeeld.

We veronderstellen nu wel dat er wel degelijk kolommen voorkomen op het gescande document: wanneer het systeem niet in staat is kolommen terug te



vinden in het brondocument, valt deze lay-outmodus op het gebruik van kaders terug!

Laad het beeldbestand COLMUNS.TIF uit de Readiris map in het geheugen als u dit alles eens wilt proberen.



De optie "Kolomgrenzen Invoegen" verfijnt het gebruik van heuse kolommen verder: deze optie bepaalt of u een kolomgrens toevoegt aan het einde van elke kolom.

Wanneer er kolomgrenzen toegevoegd worden, blijft elke tekst u die u wijzigt, toevoegt of verwijdert binnen de kolom; tekst zal nooit vanzelf over een kolomgrens heen vloeien. Alle tekst die na een kolomgrens komt verschuift naar het begin van een volgende kolom!

Schakel deze optie in wanneer u de kolomgrenzen wilt behouden waar die in het brondocument werden gedetecteerd - welke wijzigingen u later ook aanbrengt na de herkenning. In kranten en tijdschriften behoren de kolommen vaak aan verschillende artikelen toe. De tekst stiekem van de ene kolom in de andere laten overlopen is dan niet zo'n goed idee!

Schakel deze optie uit wanneer u een doorlopende tekst hebt die in kolommen is geschikt; u zorgt er dan voor dat de tekst op natuurlijke wijze van de ene kolom naar de andere overloopt.

TEKSTFORMATTERING, DEEL 2

De andere lay-outopties zijn "Lopende Tekst Creëren" en "Woord- en Paragraafformattering Behouden".

Zoals het pictogram aan de rechterkant illustreert, betekent **lopende tekst creëren** dat u een niet geformatteerde, "continue" tekst aanmaakt. De tekst wordt hernomen, maar van zijn lay-out wordt abstractie gemaakt. Gebruik deze optie wanneer u gewoon een document wilt overnemen als broodtekst maar geen belang hecht aan zijn opmaak.



Een lopende tekst is tevens wat u krijgt wanneer u snel een tekstzone herkent door er met de rechter muisknop op te klikken en het bevel "Kopiëren als Tekst" te selecteren: wanneer de herkenning van de zone klaar is, kunt u een lopende tekst in uw teksttoepassing plakken.

De optie "Woord- en Paragraafformattering Behouden" is de middenweg: de **woordformattering** - het lettertype, de puntgrootte en stijl - wordt behouden gedurende de herkenning. Dat is ook het geval met de **paragraafformattering** - de tabs en de uitlijning.

Verwar deze formatteeroptie echter niet met de "volle" autoformattering: deze optie plaatst de paragrafen gewoon na elkaar, ze creëert geen kolommen en herschept evenmin de relatieve positie van de verschillende zones.

TEKST MEERDERE KEREN UITVOEREN

Overigens kunt u de herkenningresultaten meerdere keren exporteren zonder de herkenning opnieuw uit te voeren! Verander het tekstformaat en de formatteeropties met de knop "Formaat" en klik opnieuw op de knop "Herkennen". Deze keer wordt er geen herkenning uitgevoerd - tenzij u nieuwe vensters tekende of bestaande vensters wijzigde. Zoniet zal Readiris gewoon de herkende tekst opnieuw formatteren en opslaan in het nieuwe formaat of naar de nieuwe doeltoepassing sturen.

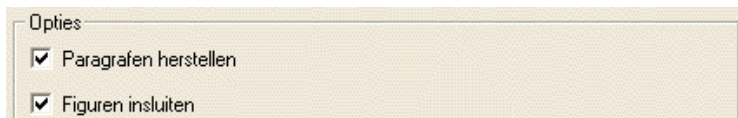


Datzelfde principe geldt voor enig ander element dat u wijzigt: wanneer u een bladzijde toevoegt aan uw document, zal enkel die pagina herkend worden. Wanneer u een nieuwe tekstzone toevoegt op eender welke bladzijde, zal enkel die zone worden herkend eer de resultaten worden opgeslagen.

Stel, u herkent een document van 10 bladzijden en slaat het op in een Word bestand. Daarna scant u de samenvatting op de coverbladzijde en u die stuurt die per e-mail naar een ongeduldige collega. Tenslotte scant u de bijlage - een tabel - en u slaat alle bladzijden op in een HTML bestand dat u op de webstek van uw onderneming plaatst.

FIGUREN APART OPSLAAN

In het vorig voorbeeld was de figuur opgenomen in het tekstbestand; of dit al dan niet het geval is, hangt af van de formatteeroptie "Figuren Insluiten". Of het mogelijk is figuren binnen de tekst te bewaren hangt van de uitvoermodus af: "arme" tekstformaten zoals Tekst (ANSI) enz. slaan geen figuren op!



Maar u kunt met Readiris ook figuren en opslaan zonder aan tekstherkenning te doen! En aangezien Readiris zwart-wit beelden, grijswaarden en kleuren ondersteunt, kunt u probleemloos zwart-wit grafisch materiaal en foto's scannen.

Hoe? Teken een grafische zone rond de illustratie, de tekening enz. die u wilt opslaan. Dat gebeurt op dezelfde wijze als wanneer u tekst- en tabelvensters tekent, alleen selecteert u nu de tool "Grafisch Venster".

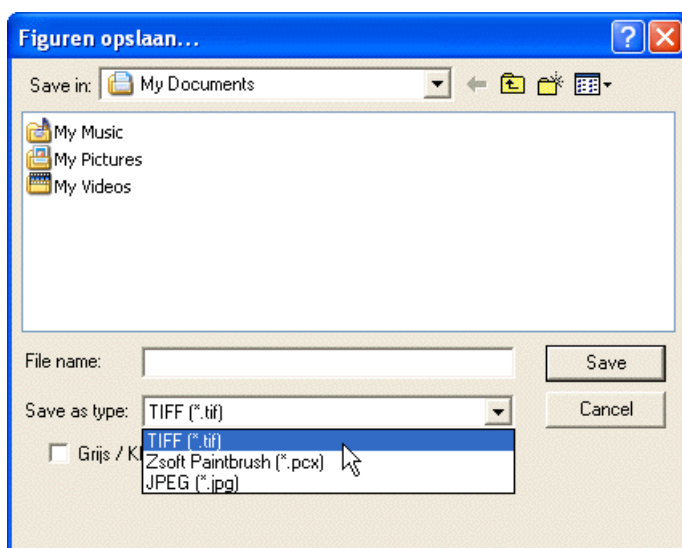




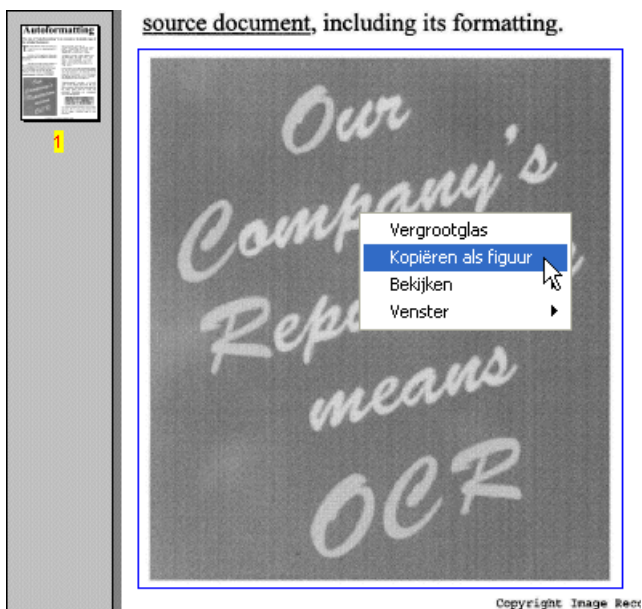
Kies daarna het commando "Figuren Opslaan" onder het menu "Bestand".

Figuren opslaan...

U moet nu een bestandsnaam opgeven. Bepaal welk grafisch formaat u wilt gebruiken. Selecteer een bestandsformaat dat uw fotoretouchesoftware ondersteunt. U hebt de keuze tussen de formaten JPEG, TIFF en Paintbrush (PCX). Schakel de optie "Grijs/Kleur" in om de figuur in kleur of grijswaarden op te slaan.



Om een figuur naar het klembord te sturen i.p.v. die op te slaan in een beeldbestand, klikt u met de rechter muisknop boven een grafisch venster en selecteert u het bevel "Kopiëren als Figuur": de grafische zone onder de muiscursor is klaar om geplakt te worden!



FAXEN HERKENNEN EN UITGESTELDE HERKENNING

Gescande documenten als beeldbestanden bewaren opent een uitdagende mogelijkheid: u kunt de *volledige* pagina bewaren en er later **uitgestelde OCR** op uitvoeren. Dat is eigenlijk wat we eerder deden met de vooraf gescande beelden van onze oefeningen.

Scan een document. Selecteer het commando "Volledige Bladzijde als Beeld Opslaan" onder het menu "Bestand" om een enkele bladzijde op te slaan. U wordt uitgenodigd de volledige bladzijde te bewaren als een TIFF of Paintbrush (PCX) bestand.



Volledige bladzijde als beeld opslaan...

Alle bladzijden als beeld opslaan...



Selecteer het commando "Alle Bladzijden als Beeld Opslaan" om een meerbladig document op te slaan. Er is één bestandsformaat beschikbaar: het meerbladig TIFF formaat.

U kunt nu de schijf selecteren als bronbeeld en het beeld openen met de knop "Openen" (of met het overeenkomstige bevel onder het menu "Actie"). (Als u het commando "Openen" onder het "Bestand" menu gebruikt moet u de beeldbron zelfs niet aanpassen.)

Daar Readiris beelden in kleur, grijswaarden en zwart-wit ondersteunt, kunt u Adobe Acrobat PDF bestanden, JPEG beelden, Paintbrush (PCX) beelden, DCX faxen (meerbladige versie van het Paintbrush formaat), (niet gecomprimeerde of LZW, PackBits, Groep 3 en Groep 4 gecomprimeerde) TIFF beelden, meerbladige TIFF beelden en Windows bitmaps (BMP) openen.

Deze mogelijkheid is vooral handig om **faxen** om te zetten in editeerbare tekstbestanden! Readiris gebruikt overigens extra intelligentie om faxen te herkennen: de software detecteert de typische faxresoluties - 100 x 200 dpi ("normale kwaliteit"), 200 x 200 dpi ("fijne kwaliteit") en 200 x 400 dpi ("superfijne kwaliteit") - en optimaliseert deze beelden vóór de herkenning automatisch om een zo goed mogelijk herkenningsresultaat te behalen.

Niettemin is het een goed idee om uw correspondenten te vragen faxen te sturen met de "fijne" kwaliteit - deze faxen hebben de hogere resolutie van 200 dpi en zullen een beter OCR resultaat opleveren.

Vergeet overigens niet dat u met de rechter muisknop op beelden kunt klikken in de Windows Verkenner en het bevel "Herkennen" van het "Context" menu kunt aanklikken. U kunt ook gebruik maken van een "drag & drop"-operatie om beelden te openen: sleep beeldbestanden van de Windows Verkenner naar de beeldzone of naar het pictogram van Readiris om ze onmiddellijk te openen!

TABELLEN HERKENNEN

Tot nog toe hebben we teksten en faxen herkend en figuren opgeslagen. Laten we nu een tabel gaan verwerken. Neem een tabel en scan die of open het voorbeeld TABLES.JPG in uw Readiris map.

Het beeld TABLES.JPG bevat eigenlijk twee tabellen en dat is geen toeval! De pagina-analyse detecteert beide tabellen en Readiris zal ze hercreëren door ze cel per cel te reconstrueren in uw rekenblad of door een tabelobject in te voegen in uw tekstbestanden.

Laten we de verschillende mogelijkheden doorlopen. We starten met de “omkaderde” tabel, haar cellen zijn omgeven door een rand.



Readiris - C:\Program Files\Readiris\Tables.jpg (bladzijde 1 van 1)

Bestand Bewerken Instelling Bekijken Actie Leren Registreren Hulp

Reading Tables

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

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

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

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

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

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

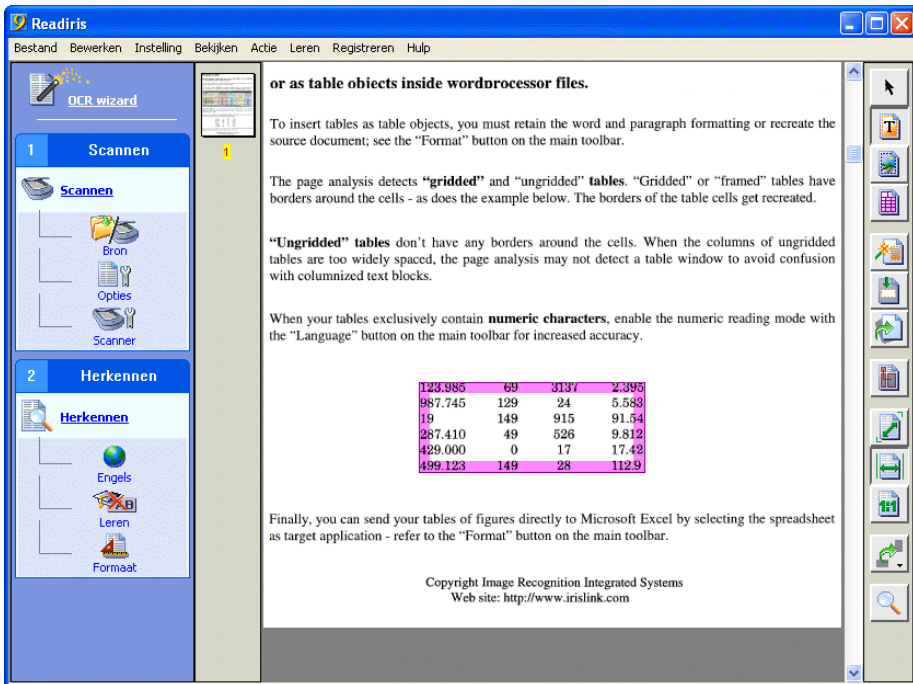
When your tables exclusively contain **numeric characters**, enable the numeric reading mode with the "Language" button on the main toolbar for increased accuracy.

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

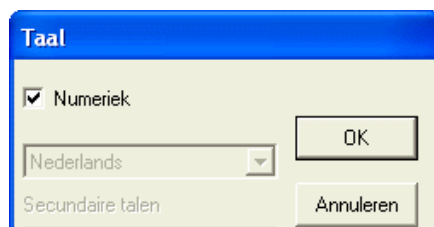
Voer de herkenning uit terwijl de lay-outoptie "Woord- en Paragraafformattering Behouden" of "Brondocument Hercreëren" ingeschakeld is en de tabel wordt herschapen. Open uw tekstverwerker om het resultaat even te bekijken: de cellen en randen werden één voor één door Readiris herschapen! (U had trouwens de overige tekstblokken in de uitvoer kunnen insluiten.)

Performance test optical media	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				

Nu de “niet omkaderde” tabel. Die heeft geen randen rond de cellen, maar niettemin detecteert de pagina-analyse haar!

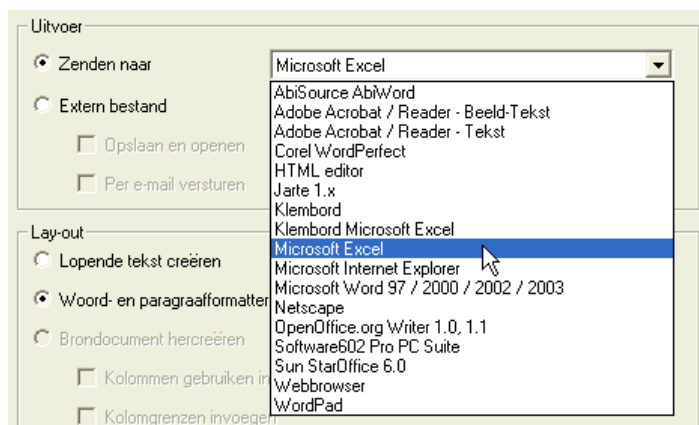


Voor een optimale OCR precisie, beperkt u de herkenning best tot de **numerieke symbolen** met de "Taal" knop. (De numerieke modus bevat niet enkel de getallen, maar ook de symbolen +, *, /, %, , (komma), . (punt), (,), -, =, \$, £, ¥ en het € symbool.)



Aangezien u dit alleen kunt doen wanneer de tabel geen alfabetische symbolen bevat - zoniet worden de tekstdelen niet goed herkend - kunnen we hier de numerieke modus activeren maar konden we dat eerder niet voor de eerste tabel.

Deze keer zullen we het herkenningresultaat rechtstreeks naar het rekenblad Microsoft Excel sturen, we kiezen dus Excel als doeltoepassing met de "Formaat" knop.



Het rekenblad wordt automatisch opgestart en het resultaat ziet er zo uit: de typische tabelstructuur met rijen en kolommen werd herschapen, u bent dus onmiddellijk klaar om de data te verwerken.



	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					

Het kan voorkomen dat u niet omkaderde tabellen ontmoet die de pagina-analyse niet als tabelvensters markeert omdat er teveel ruimte zit tussen de kolommen. Readiris tracht namelijk de verwarring met tekstkolommen te vermijden. Om tabelvensters manueel te creëren, klikt u op de tool "Tabelvenster" en gaat u zoals gewoonlijk te werk. De tooltip van de knop "Tabelvensters" geeft het aantal tabelvensters aan.

Tabelvensters tekenen: 0



ON-LINE HULP RAADPLEGEN

Hier eindigt onze rondleiding van Readiris. Het kan voorkomen dat bepaalde informatie van het laatste ogenblik niet in deze handleiding is opgenomen. We raden u dan ook aan om de on-line hulp te raadplegen als u meer wilt weten over Readiris.

Ga hiervoor naar het "Hulp" menu. Het commando "Hulpthema's" en zijn versnellingsstoets F1 laten u navigeren door de vele hulpbladzijden.



De andere bevelen uit het menu "Hulp" vertellen u hoe u op onze productondersteuning kunt beroep doen, hoe u I.R.I.S. kunt contacteren, geven u rechtstreeks toegang tot de homepage van I.R.I.S. enz.

