

Readiris™ PRO 10

GUIDE DE L'UTILISATEUR



Readiris Corporate

© 2005 I.R.I.S. Tous droits réservés
Technologie OCR par I.R.I.S.
Technologie Connectionist, AutoFormat et linguistique par I.R.I.S.
Technologie ICR et lecture de codes à barres
Technologie BCR et d'analyse du champs par I.R.I.S.

© 2005 I.R.I.S. Tous droits réservés

GAGNEZ DU TEMPS, ENTREZ SANS FRAPPER!

Nous vous félicitons pour votre acquisition de Readiris. Ce logiciel vous sera très utile pour la saisie de vos textes, tables, graphiques, cartes de visite, codes à barres et même vos textes manuscrits.

Aussi efficaces que puissent être les ordinateurs, il vous faut toujours d'abord encoder les informations. S'il vous est déjà arrivé de devoir refrapper un rapport de 15 pages ou un grand tableau de chiffres, vous savez comme cette tâche peut être longue et fastidieuse. Utilisez ce logiciel à la pointe du progrès pour encoder automatiquement du texte dans vos applications et vous atteindrez un niveau d'efficacité et de confort jamais atteint!

Numérisez un document imprimé ou tapé à la machine, indiquez les zones à lire - ou laissez le système les détecter pour vous - et effectuez la reconnaissance de caractères. Des documents composés de plusieurs pages seront traités du début à la fin avec un effort minime. Quelques clics vous épargnent de longues heures de travail car Readiris convertit vos documents papier en fichiers éditables par l'ordinateur: c'est jusqu'à 40 fois plus rapide que la reffrappe manuelle.

L'assistant OCR vous mène à travers le processus d'OCR d'une manière confortable: vous répondez à quelques questions simples pour obtenir des résultats rapides avec Readiris. Vous pouvez envoyer les résultats de la lecture directement dans votre traitement de texte ou dans votre tableur. Pour reconnaître des télécopies et convertir des documents PDF, vous pouvez glisser-déposer les fichiers images depuis l'Explorateur de Windows vers la fenêtre de l'application de Readiris. Ou cliquez avec le bouton droit de la souris sur une image pour l'envoyer aussitôt dans Readiris.

Readiris reconnaît les données insérées dans des tables et les recrée comme feuille de calcul ou comme objet table dans votre traitement de texte; vos données numériques sont immédiatement prêtes pour les manipulations ultérieures.

Basé sur la technologie Connectionist d'I.R.I.S., Readiris représente ce que l'OCR vous offre de mieux. La reconnaissance omnifont est complétée par des techniques d'auto-apprentissage basées sur un réseau neuronal propriétaire. Le



système peut apprendre des nouveaux caractères grâce à l'analyse contextuelle: des connaissances linguistiques à propos des syllabes et des mots améliorent les performances de l'OCR.

Readiris reconnaît jusqu'à 117 langues: toutes les langues américaines et européennes, en ce y compris les langues de l'Europe de l'Est, les langues baltiques, le grec et les langues cyrilliques ("russes"). (De manière optionnelle, vous pouvez lire les documents en hébreu et lire quatre langues asiatiques: le japonais, le chinois simplifié, le chinois traditionnel et le coréen.). Readiris supporte aussi des mélanges d'alphabets: le logiciel détecte les mots "occidentaux" qui surviennent dans des textes grecs, cyrilliques, hébreux ou asiatiques - de nombreux noms propres, noms de marques, etc. intraduisibles sont écrits avec des symboles occidentaux.

Readiris utilise ces connaissances linguistiques *pendant* la phase de reconnaissance, pas après. Cela permet à Readiris de reconnaître des documents de tous types avec une précision maximale, en ce compris les documents de mauvaise qualité tels que les télécopies et les textes issus d'imprimantes à aiguille. Readiris se charge de documents mal numérisés et de copies de documents contenant des formes trop claires ou trop sombres. Des caractères collés ("ligatures") sont reconnus avec succès et des formes fragmentées, tels que les symboles issus d'une imprimante matricielle, sont recomposés.

La vérification avec "contexte dans l'image" par l'utilisateur non seulement signale les caractères douteux, mais augmente également la précision du système. Toutes les solutions confirmées par l'utilisateur sont mémorisées, ce qui augmente la vitesse et la fiabilité au fur et à mesure de votre progression. Utiliser Readiris signifie le rendre plus intelligent à chaque fois! Ce puissant outil d'apprentissage vous permet non seulement d'apprendre à Readiris les caractères spéciaux tels que les symboles mathématiques et les dingbats mais aussi de se charger des polices tordues que vous pourrez trouver dans des documents réels.

Pour augmenter encore votre productivité, Readiris non seulement reconnaît vos textes, mais peut aussi les *formater* pour vous! Utilisez "l'autoformatage" et Readiris recrée une copie facsimile du document numérisé: le formatage du mot, du paragraphe et de la page du document original sont retenus.

Les caractères sont reproduits dans une police similaire, la taille et le style des caractères sont également préservés lors de la reconnaissance. L'emplacement des colonnes, des blocs de texte et des graphiques est fidèle à celui de vos documents originaux. Readiris peut même inclure dans le document reconnu l'image qui se trouve dans l'arrière-plan d'une page numérisée! Et puisque Readiris reconnaît sans problème les numérisations en échelle de gris et en couleur, vous pouvez saisir n'importe quel graphique - qu'il s'agisse de modes traits, de photos en noir et blanc ou d'illustrations colorées. Lorsqu'un document contient des tables, Readiris les réorganise dans de vraies cellules et recrée les bordures des tables originales.

En d'autres mots, Readiris vous permet d'archiver une copie exacte de votre document, qui sera un fichier texte éditable et compact au lieu d'une image scannée! Plusieurs niveaux de formatage sont disponibles, le choix appartient à l'utilisateur.

Les codes à barres qui figurent sur les pages numérisées peuvent également être lus, ainsi que les textes manuscrits, à condition que les lettres "imprimées" majuscules soient suffisamment espacées.

Readiris peut même reconnaître des cartes de visite: numérisez vos cartes de visite, reconnaissez-les et convertissez-les en informations d'une base de données. Rappelez-vous de votre dernière foire commerciale, lorsque vous êtes revenu avec une pile de cartes de visite que votre secrétaire a mis deux jours à encoder!

Les données de la carte sont extraites automatiquement de l'image et les données reconnues sont attribuées à des champs spécifiques de la base de données. Readiris utilise une base de données de connaissance, et dispose de cette façon de l'intelligence nécessaire pour distinguer le prénom du nom, une ville et son pays, un numéro de téléphone et un numéro de télécopie, etc. Vous pouvez envoyer directement les résultats dans un logiciel de gestion de contacts comme Microsoft Outlook (Express) ou dans n'importe quelle application supportant le standard vCard.

Readiris supporte une large variété de numériseurs: de nombreux scanners à plat, à alimentation, les systèmes hybrides ou "tout en un" (également appelés "MFPs", "périphériques multifonctionnels") et les appareils photos numériques. Readiris supporte également le standard de saisie Twain, de même que certaines



plates-formes de numérisation. Si vous ne disposez pas d'une alimentation automatique, la numérisation avec pause vous permettra toutefois d'acquérir rapidement des documents multipages. (Readiris Corporate supporte des scanners à haute vitesse et exécute l'OCR en lot sur un grand nombre d'images. Des pages blanches sont utilisées pour segmenter les lots en divers documents. Ces derniers sont correctement indexés grâce au système de lecture automatique des codes à barres.)

TABLE DES MATIÈRES

Gagnez du temps, entrez sans frapper!	III
Table des matières	VI
Droits d'auteur	VIII

Chapitre 1: Installation

Exigences du système	1-1
Installation de Readiris	1-1
Désinstallation de Readiris	1-3
Programme de désinstallation de Readiris	1-3
Assistant de (dés)installation de Windows	1-4
Installation des options logicielles	1-5
Fichiers installés	1-8
Fichier "Lisez-moi" et documentation	1-8
Formulaire pour les textes manuscrits	1-8
Pilotes de numériseur	1-8
Enregistrez-vous!	1-9
Support du produit	1-11
Comment contacter I.R.I.S.	1-12

Chapitre 2: Visite Guidée

Démarrage du logiciel	2-1
Le premier démarrage	2-2
Découvrez l'interface de Readiris	2-3
Débuter avec une première leçon	2-6
Agrandir l'image	2-11
Un, décomposition d'une image numérisée	2-14

Un et demi, trier des fenêtres	2-17
Deux, fenêtrage manuel d'une image	2-20
Trois, enregistrer des modèles de fenêtrage	2-24
Readiris vous emmène autour du monde	2-26
Allez plus loin avec les lexiques d'utilisateur	2-31
Readiris change de langue au besoin	2-34
Lire des documents comprenant plusieurs langues	2-36
Définir les caractéristiques du document	2-38
Readiris devient chaque fois plus intelligent!	2-40
Apprendre	2-42
Ne Pas Apprendre	2-42
Effacer	2-43
Annuler	2-43
Finir	2-43
Interrompre	2-44
Le rôle des dictionnaires de fontes	2-44
Envoyer le résultat dans votre application	2-47
Sauvegarder les résultats dans un fichier texte	2-51
Créer des documents portables... ..	2-55
... Ou les lire!	2-62
Reconnaissance de pages multiples	2-65
Édition de documents multipages	2-75
Commencer avec un nouveau document	2-77
Reconnaissance des zones de texte	2-78
Organisation du texte saisi	2-80
Configuration du numériseur	2-81
Mettez un peu de couleur dans vos textes!	2-83
À différents appareils, différentes résolutions	2-86
Sauvegarder les paramètres par défaut	2-91
Sauvegarde de paramètres spécifiques	2-91
Numériser des documents	2-92
Ajuster les images numérisées	2-96
Laisser l'assistant travailler pour vous	2-101
Readiris recrée la mise en page d'un document	2-102
Des colonnes s'il vous plaît, pas de cadres!	2-107
Formater le texte, deuxième partie	2-110
Exporter du texte plusieurs fois	2-111



Sauvegarder des graphiques à part	2-112
Sauvegarder les couleurs de l'arrière-plan	2-115
Allez plus loin avec vos graphiques!	2-118
Lecture de fax et reconnaissance en différé	2-120
Reconnaissance de documents en lots	2-121
Etablir un dossier surveillé	2-123
Organiser des lots de documents	2-125
Reconnaissance des tables	2-129
Reconnaître des textes manuscrits	2-134
Lire les barres et les espaces	2-138
Reconnaître des cartes de visite	2-140
Numériser des cartes de visite	2-141
Activez le mode de lecture des cartes de visite	2-146
Reconnaissance des cartes de visite	2-150
Obtenir de l'aide en ligne	2-151

DROITS D'AUTEUR

Le logiciel Readiris est conçu et développé par I.R.I.S. Les technologies OCR, ICR, BCR, lecture de codes à barres, Connectionist, AutoFormat et la technologie linguistique par I.R.I.S. I.R.I.S. détient tous les droits d'auteur pour le logiciel Readiris, la technologie OCR, la technologie ICR, la technologie de lecture de codes à barres, la technologie BCR, la technologie linguistique, l'aide en ligne et ce manuel.

AutoFormat, Connectionist, l'IBCR-II, le logo I.R.I.S., Readiris et la technologie linguistique I.R.I.S. sont des marques déposées d'I.R.I.S.

Le parser XML est développé par Apache. Ce produit inclut un logiciel conçu par Apache Software Foundation (www.apache.org).

Acrobat et Reader sont des marques déposées d'Adobe. Excel, Windows et Word sont des marques (déposées) de Microsoft. Intel est une marque déposée d'Intel.

Chapitre 1

INSTALLATION

Ce chapitre décrit les exigences du système et l'installation de Readiris.

EXIGENCES DU SYSTÈME

Voici la configuration minimale requise pour utiliser Readiris:

- un PC Intel 486 ou compatible. Un PC Pentium est recommandé.
- 64 MO de mémoire vive. 128 MO de mémoire vive sont recommandés pour traiter les images en échelle de gris et en couleur.
- 120 MO d'espace disponible sur le disque dur. 105 MO d'espace suffisent si vous laissez les fichiers d'exemple sur le CD-ROM.
- le système d'exploitation Windows XP, Windows ME, Windows 2000, Windows 98 ou Windows NT 4.0.

Readiris Corporate exige un moniteur d'une résolution de 1.024 x 768.

Notez que certains **pilotes de numériseurs** peuvent ne pas fonctionner sous la (ou les) dernière(s) version(s) de Windows! Consultez la documentation fournie avec votre scanner pour savoir quelles sont les plates-formes supportées.

INSTALLATION DE READIRIS

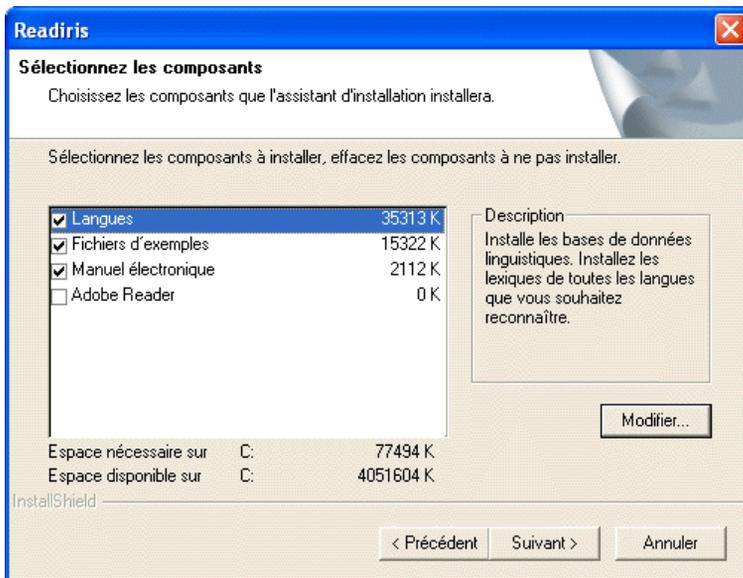
Readiris est exclusivement livré sur un **CD-ROM auto-exécutable**. Pour l'installer, insérez simplement le CD-ROM dans votre lecteur de CD-ROM et attendez que le programme d'installation commence à tourner. Suivez les instructions à l'écran.



Si l'installation ne démarre pas quand le CD-ROM est inséré dans votre lecteur, utilisez le programme MENU.EXE pour installer le programme.

Les utilisateurs de Windows XP, Windows 2000 et Windows NT doivent s'assurer qu'ils ont les **droits d'accès** nécessaires - contactez l'administrateur du système si nécessaire.

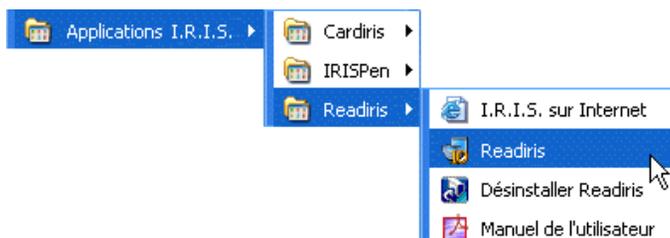
Certaines options d'installation sont offertes. Assurez-vous d'installer les **bases de données linguistiques** de toutes les langues que vous souhaitez lire. Par défaut, tous les lexiques sont installés. Nous vous recommandons d'installer les **images exemples** qui sont utilisées dans le didacticiel de ce manuel.



De façon similaire, installez le programme Adobe Reader requis pour accéder à la documentation sur le programme si cela s'avère nécessaire. Le **manuel**

électronique est par défaut copié sur votre disque dur. Vous pouvez par contre le laisser sur le CD-ROM.

Le sous-menu "Applications I.R.I.S.- Readiris" sous le menu "Programmes" est automatiquement créé à l'installation du programme.



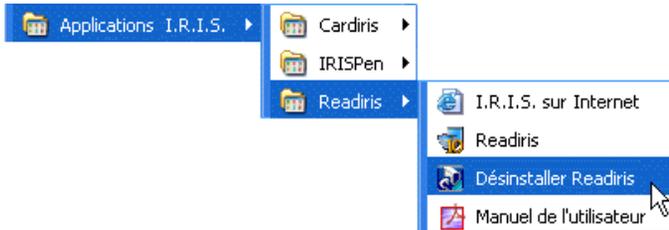
Il en va de même pour le **raccourci** vers Readiris sur le **bureau** de Windows. Vous pourrez ainsi démarrer directement Readiris depuis votre bureau.

DÉSINSTALLATION DE READIRIS

Il n'y a que deux manières correctes de désinstaller Readiris: en utilisant le programme de désinstallation de Readiris et en utilisant l'assistant de (dés)installation de Windows. Nous vous recommandons fermement de *ne pas* désinstaller Readiris ou ses modules en supprimant manuellement les fichiers du logiciel.

Programme de désinstallation de Readiris

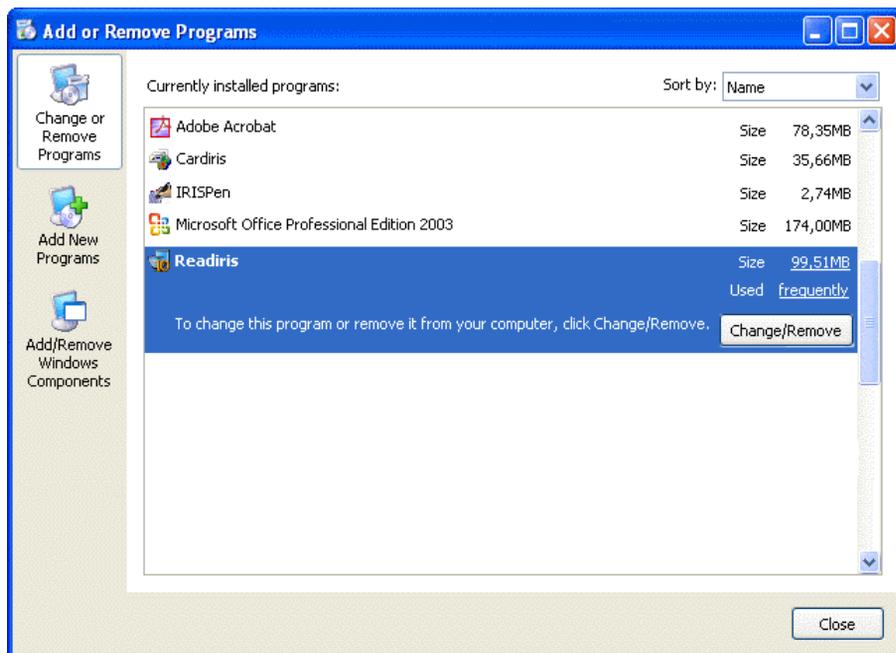
Sélectionnez "Désinstaller Readiris" dans le sous-menu "Applications I.R.I.S.- Readiris" pour lancer le programme d'effacement de Readiris et suivez les instructions à l'écran.



Assistant de (dés)installation de Windows

Exécutez les étapes suivantes pour utiliser l'assistant de (dés)installation de Windows.

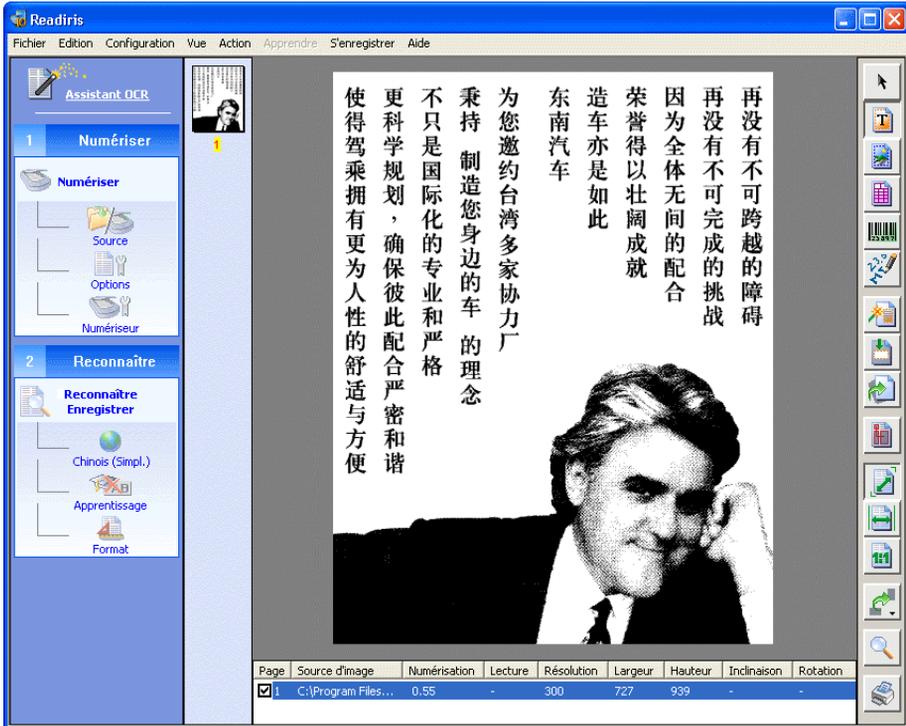
- Cliquez sur "Paramètres" dans le menu "Démarrer" de Windows et passez sur le "Panneau de Configuration".
- Cliquez sur l'icône "Ajout/Suppression de Programmes" dans le panneau de configuration.



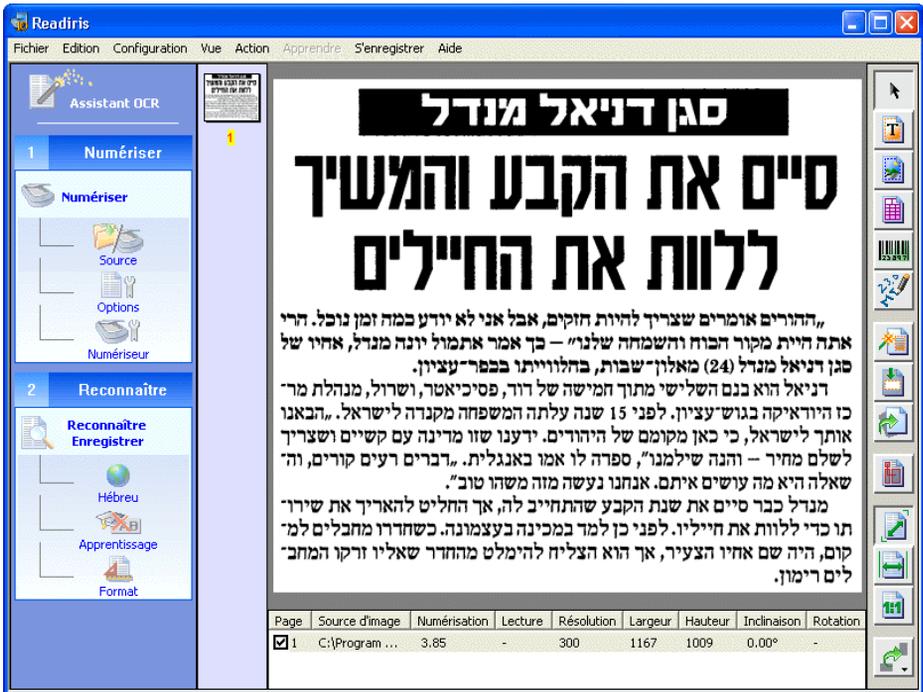
- ❑ Suivez les instructions à l'écran pour supprimer le logiciel Readiris.

INSTALLATION DES OPTIONS LOGICIELLES

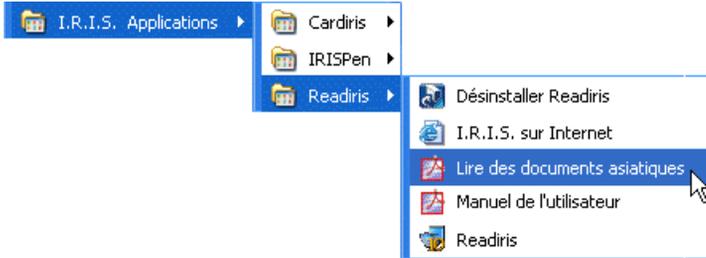
Il y a deux options logicielles disponibles qui complètent Readiris: le “Complément d’OCR asiatique” et le “Complément d’OCR hébreu”. Le “Complément d’OCR asiatique” vous permet de lire le japonais, le chinois traditionnel, le chinois simplifié et le coréen.



Le "Complément d'OCR hébreu" vous permet de reconnaître des documents en hébreu.



Ces logiciels sont également livrés sur un CD-ROM auto-exécutable. En installant l'option “Complément d'OCR asiatique”, vous disposez d'une documentation spécifique traitant de la reconnaissance de documents asiatiques.



FICHIERS INSTALLÉS

Le programme d'installation a créé un dossier où se trouvent les fichiers de Readiris. N'essayez jamais de désinstaller Readiris ou certains de ses modules en supprimant manuellement des fichiers du programme, utilisez pour ce faire le programme de désinstallation ou l'assistant de désinstallation de Windows. Voir plus haut.

Fichier “Lisez-moi” et documentation

README.HTML Fichier de type “Lisez-moi” (en format HTML)

MANUAL.PDF Manuel d'utilisateur (en format Adobe Acrobat)

Formulaire pour les textes manuscrits

TEMPLATE.PDF Formulaire vierge pour textes manuscrits
(version pour l'impression)

TEMPLATE.DOC Formulaire vierge pour textes manuscrits
(version éditable)

Pilotes de numériseur

N'hésitez pas à contacter votre fabricant de numériseur ou son revendeur si les problèmes avec les pilotes de votre numériseur persistent. La plupart des

fabriquants vous permettent de télécharger les dernières versions des pilotes des numériseurs depuis leur site web.

ENREGISTREZ-VOUS!

N'oubliez pas d'enregistrer votre licence Readiris! Ceci nous permettra de vous tenir informé des futurs développements du produit ainsi que des autres produits I.R.I.S. apparentés. Les avantages de l'enregistrement, entre autres le **support gratuit du produit** et les **offres spéciales**, sont exclusivement réservés aux utilisateurs enregistrés.

Vous pouvez enregistrer le produit de plusieurs façons: envoyez la carte d'enregistrement ou télécopiez son équivalent électronique, téléphonez à I.R.I.S. pendant les heures ouvrables ou encore remplissez un formulaire d'enregistrement sur le site web d'I.R.I.S.



Aide de Readiris

Hide Back Forward Home Print Options

Contents Index Search

- Bienvue dans l'aide de Readiris
- Introduction à l'OCR
- Reconnaître des documents
- Reconnaître du texte manuscrit
- Reconnaître des codes à barres
- Comment faire...?
- Informations de référence
- Versions et options du logiciel
- Enregistrement du produit
 - Enregistrer votre licence Readiris**
 - Formulaire d'enregistrement
 - S'enregistrer par courrier électronique
- Support technique
- I.R.I.S.

Enregistrez votre licence Readiris

Pourquoi s'enregistrer?

- L'enregistrement vous permet de rester informé des **développements** futurs de votre **produit** et des produits apparentés d'I.R.I.S.
- L'enregistrement vous donne droit au **support technique** gratuit et à des **offres promotionnelles**.
- Selon la version reçue du logiciel, vous recevrez une **clé logicielle** dont vous pourriez avoir besoin pour continuer d'utiliser votre Readiris après un mois.

Comment s'enregistrer?

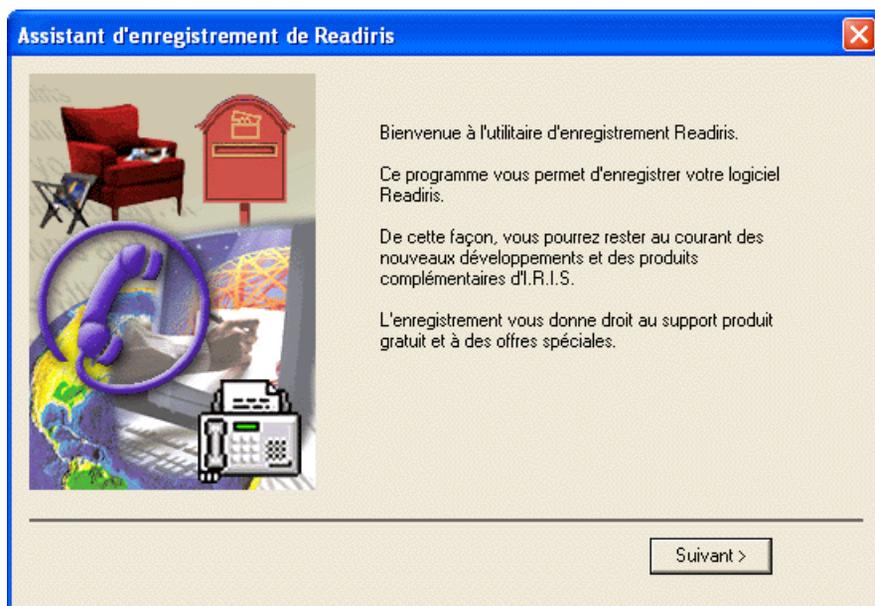
Par courrier

Envoyez votre **carte d'enregistrement**.

Télécopie

Si vous avez perdu votre carte d'enregistrement, ou si la documentation logicielle n'est fournie que sous **forme électronique**, imprimez la [carte d'enregistrement](#) et envoyez-la par courrier ou télécopie à I.R.I.S.

L'**assistant d'enregistrement** que vous trouverez sous le menu "S'enregistrer" du logiciel Readiris vous accompagnera pour accomplir facilement la procédure d'enregistrement.



Selon la version que vous possédez, vous recevrez en retour la **clé logicielle** qui vous sera peut-être nécessaire pour pouvoir continuer à utiliser Readiris après un mois.

SUPPORT DU PRODUIT

La commande "Support du Produit" sous le menu "Aide" de Readiris vous explique comment obtenir de l'aide technique. Décrivez clairement le problème rencontré et ajoutez à votre description toutes les données utiles au sujet de Readiris, votre numériseur et de votre ordinateur.



Aide de Readiris

Hide Back Forward Home Print Options

Contents Index Search

- Bienvne dans l'aide de Readiris
- Introduction à l'OCR
- Reconnaître des documents
- Reconnaître du texte manuscrit
- Reconnaître des codes à barres
- Comment faire...?
- Informations de référence
- Versions et options du logiciel
- Enregistrement du produit
- Support technique
 - Obtenir le support technique**
 - Contacteur I.R.I.S.
 - Obtenir le support technique par I.R.I.S.

Comment accéder au support produit?

Un **support technique gratuit** est offert à tous les **utilisateurs enregistrés**. (Vous [enregistrez](#) vous permet également de bénéficier d'offres promotionnelles.)

Europe

Ligne technique: 32-10-45 13 64 (heures de bureau)
(langues principales)
Télécopie: 32-10-45 34 43

États-Unis

Ligne technique: 1-561-921-0847 / 800-447-4744
(heures de bureau)
Télécopie. 1-561-921-0854

WWW

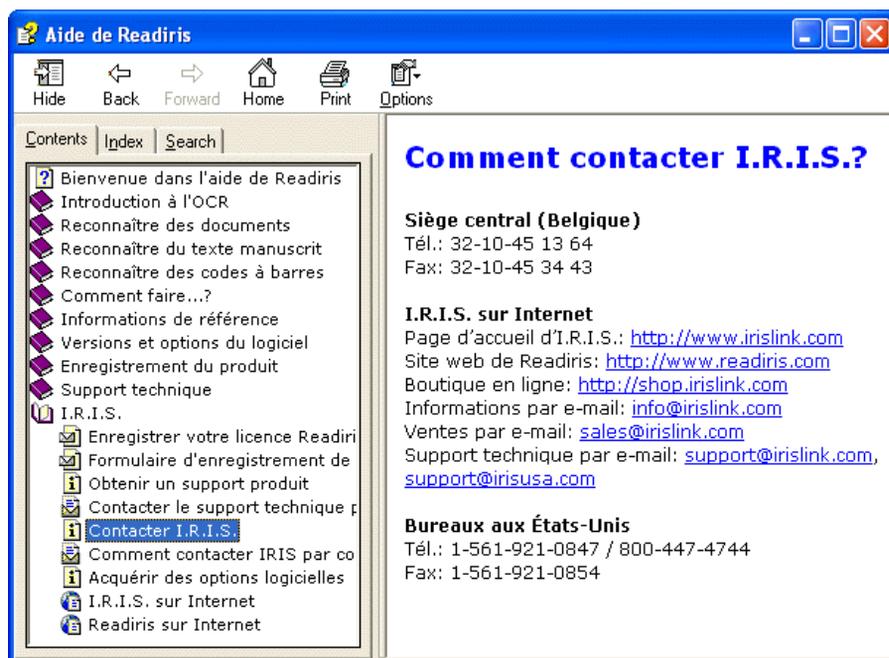
www.irislink.com/support.html (informations sur les problèmes fréquemment rencontrés)
Cliquez [ici](#) pour accéder à ces informations sur les problèmes fréquemment rencontrés.

E-mail

support@irislink.com, support@irisusa.com

COMMENT CONTACTER I.R.I.S.

Vous pouvez toujours contacter I.R.I.S. pour en savoir plus sur les programmes complémentaires; la commande "Contacter I.R.I.S." sous le menu "Aide" de Readiris vous explique de quelles manières vous pouvez contacter I.R.I.S.



Une icône d'application dans le sous-menu "Applications I.R.I.S. - Readiris" sous le menu "Programmes" vous permet d'accéder directement à la **page d'accueil** d'I.R.I.S. De même l'écran de démarrage de Readiris et la commande "I.R.I.S. sur Internet" sous le menu "Aide" de Readiris.



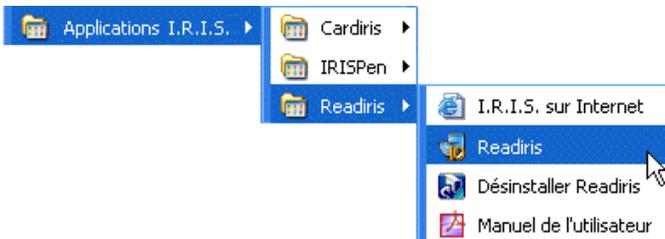
Chapitre 2

VISITE GUIDÉE

Readiris est un produit de pointe pour la reconnaissance optique de caractères (OCR), avec de nombreuses caractéristiques avancées. Nous envisagerons dans ce chapitre les principales applications de Readiris ainsi que de nombreuses astuces pour mieux l'utiliser.

DÉMARRAGE DU LOGICIEL

Cliquez sur l'application Readiris dans le sous-menu "Applications I.R.I.S. - Readiris", ou cliquez sur le raccourci vers Readiris sur votre bureau Windows.



L'écran de démarrage ainsi que la fenêtre de l'application Readiris apparaissent. L'écran de démarrage affiche la version du logiciel ainsi que des messages liés aux droits d'auteur. Il vous permet en outre un accès direct à la **page d'accueil** d'I.R.I.S. - il vous suffit de cliquer sur l'URL pour visiter le site web d'I.R.I.S. Cliquez avec la souris n'importe où ailleurs pour faire disparaître cet écran.

La fenêtre suivante concerne l'assistant d'OCR. Cliquez sur "Annuler" pour le moment.

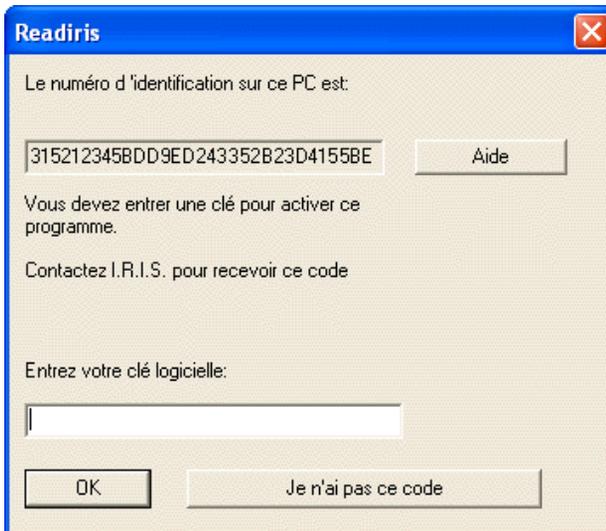


LE PREMIER DÉMARRAGE

En fonction de la version acquise du logiciel, le premier démarrage peut être particulier: on peut vous demander d'enregistrer votre licence.

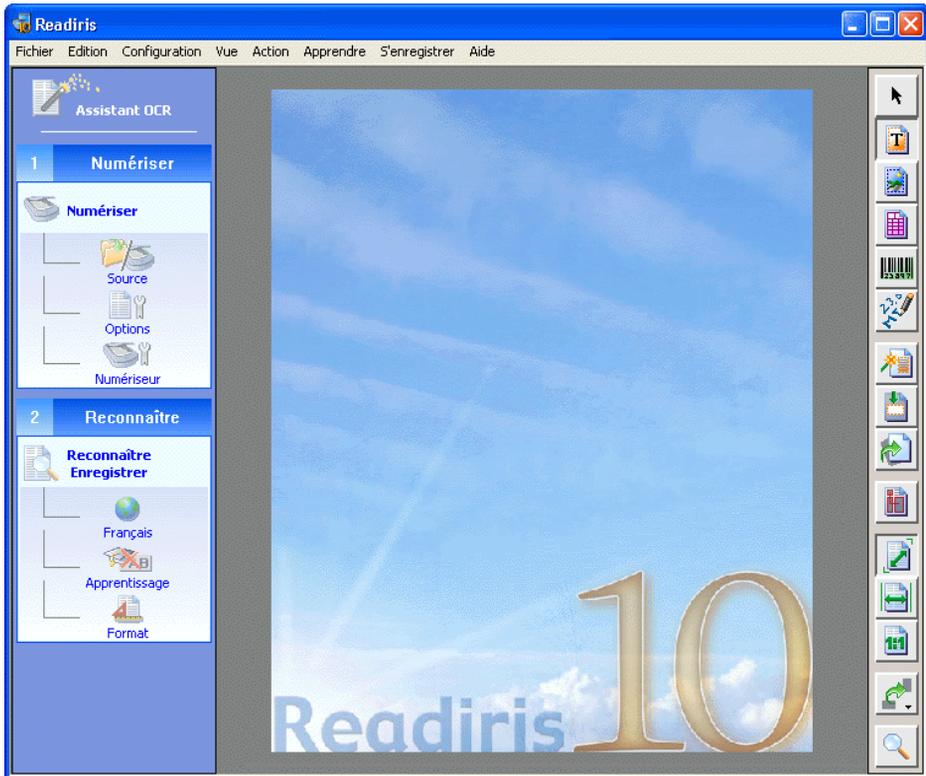
Si c'est le cas, l'utilisation de Readiris est limitée à un mois, et en vous enregistrant, I.R.I.S. vous enverra gratuitement une **clé logicielle** afin de pouvoir continuer à utiliser le logiciel après 30 jours.

Votre **numéro d'identification** est nécessaire à la création de la clé logicielle. Assurez-vous que ce numéro est disponible ou mentionné quand vous enregistrez votre licence.



DÉCOUVREZ L'INTERFACE DE READIRIS

La fenêtre d'application de Readiris ne contient pas seulement les **menus de commande** mais aussi deux barres de boutons qui donnent accès aux commandes fréquemment utilisées. Initialement, certains menus de commandes sont grisés: ils concernent la prévisualisation. Aussi longtemps qu'aucune image n'est ouverte, ils ne sont pas accessibles.



Il en est de même pour la **barre d'outils image** à la droite de la fenêtre d'application: elle contient toutes les commandes utiles à la prévisualisation. La **barre d'outils principale** sur la gauche de l'écran permet un accès rapide à toutes les commandes générales fréquemment utilisées.

Pour savoir quelle commande correspond à quel bouton, tenez le pointeur de votre souris sur lui pendant un moment: la **bulle explicative** de l'outil s'affichera.



La zone centrale de la fenêtre ou la **zone d'image** est l'endroit où les images numérisées sont affichées. Vous pouvez glisser-déposer des fichiers image dans la zone image et sur l'icône Readiris pour les reconnaître.

Dès que des images ont été traitées, une barre d'outils supplémentaire, la **barre d'outils des pages**, s'ajoute à gauche de la zone d'image: elle représente les différentes pages du document et permet l'accès aux commandes de la page en cliquant sur le bouton droit de la souris (menu "Contexte").



De plus, le **panneau des documents** est alors affiché en dessous de l'image numérisée. Ce panneau affiche des **données statistiques** et des informations sur les pages scannées: la source de l'image et la résolution, la durée de la numérisation et de la reconnaissance, etc. (Le panneau des documents est également accompagné de bulles explicatives...)

Page	Source d'image	Numérisation	Lecture	Résolution	Largeur	Hauteur	Inclinaison	Rotation
<input checked="" type="checkbox"/> 1	C:\Program Files\Readiris...	1.03	-	300	2000	2388	-	-
<input checked="" type="checkbox"/> 2	C:\Program Files\Readiris...	0.93	-	300	2000	1888	-	-
<input checked="" type="checkbox"/> 3	C:\Program Files\Readiris...	0.80	-	300	1912	2004	-	-

DÉBUTER AVEC UNE PREMIÈRE LEÇON

La meilleure façon de se familiariser avec le logiciel Readiris est indubitablement de l'utiliser. Un certain nombre d'**images prénumérisées** est fourni avec le programme; elles vous permettent de commencer son utilisation

même s'il n'y a pas de scanner connecté à votre ordinateur. Voyons ce qu'il en est.

Le bouton "Source" sur la barre des boutons principale détermine si vous allez scanner une image ou si vous allez utiliser une image prénumérisée comme image source.

Les images en couleur, niveaux de gris et noir et blanc sont indifféremment supportées. Readiris vous permet d'ouvrir des documents Adobe Acrobat PDF, des télécopies DCX (une version multipage du format Paintbrush), des images DjVu (*.djv, *.djvu), des images JPEG, des images JPEG 2000 (*.j2c, *.jp2), des images PNG, des images TIFF (non-compressées et compressées en LZW, PackBits, Groupe 3 et Groupe 4), des images TIFF multipages, des bitmaps Windows (*.bmp) et des images ZSoft Paintbrush (*.pcx).

Cette possibilité est particulièrement utile pour convertir vos **télécopies** en fichiers texte éditables.

Comme nous allons ouvrir une image préscannée, vous devez sélectionner "Fichiers Image" et non le numériseur comme source d'image au moyen du bouton "Source".



Ensuite, cliquez sur le bouton "Ouvrir". (Quand vous sélectionnez le disque comme source d'images, le bouton "Numériser" est remplacé par le bouton "Ouvrir" et la commande correspondante "Numériser" sous le menu "Action" est remplacé par la commande "Ouvrir".)

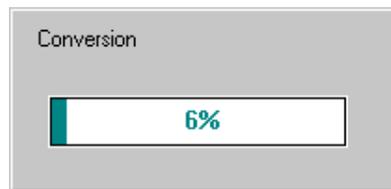
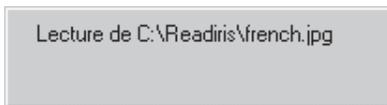




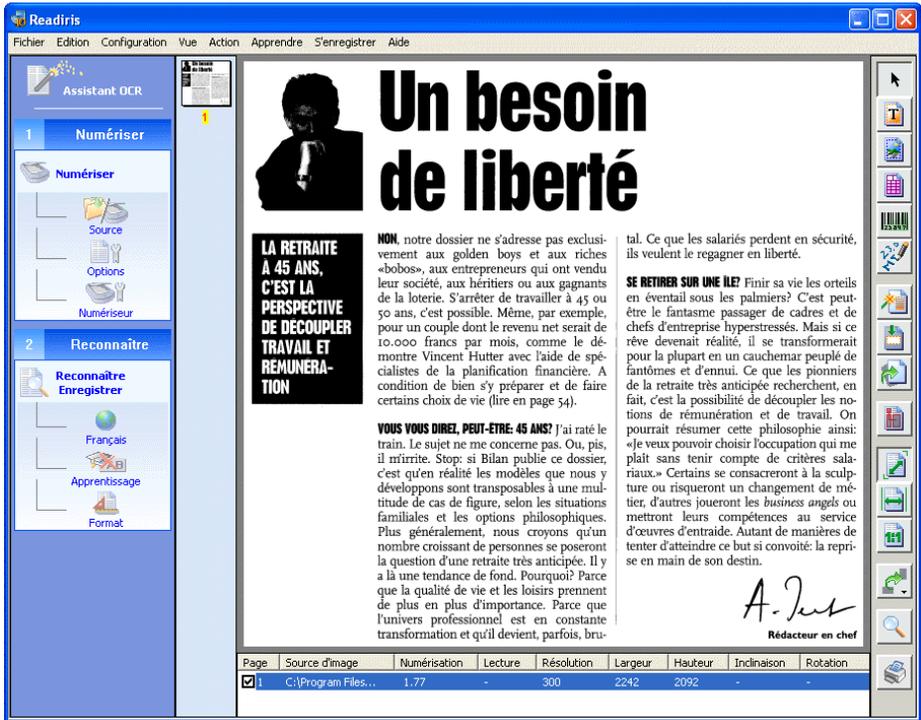
Vous pourriez également sélectionner la commande "Ouvrir" du menu "Fichier" et ouvrir directement une image prénumérisée - ceci fonctionne même lorsque votre numériseur est utilisé comme source d'image.



Vous êtes invité à sélectionner un fichier image. Sélectionnez le fichier FRENCH.JPG dans le dossier Readiris. Comme ce fichier d'exemple est une image couleur, il ne sera pas seulement lu depuis le disque dur: une version noir et blanc est créée pour le procédé d'OCR.



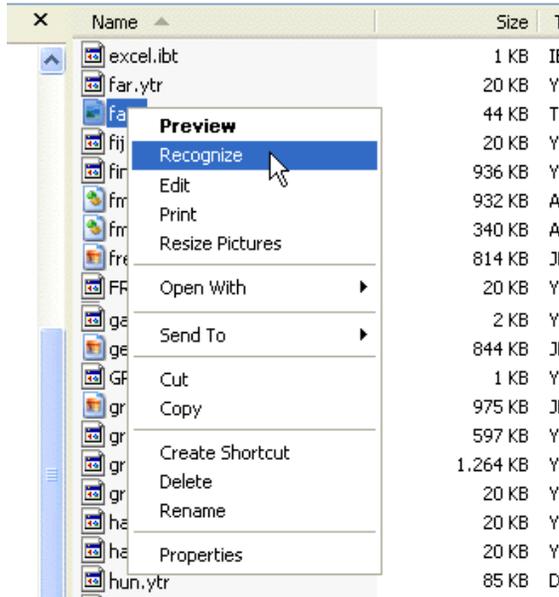
Finalement, l'image s'affiche dans la zone image. La barre d'outils des pages et le panneau des documents indiquent qu'une seule page est chargée dans Readiris.



Une troisième façon d'ouvrir des images prénumérisées est d'utiliser la méthode du "glisser-déposer". Faites glisser des images depuis l'Explorateur Windows vers la zone image de Readiris ou sur l'icône Readiris pour les ouvrir directement.



Enfin, vous pouvez également ouvrir des images à partir de l'Explorateur Windows: **cliquez** avec le bouton **droit** de la souris sur un fichier image et sélectionnez la commande "Reconnaître" du menu "Contexte". (Cette commande n'apparaît que lorsque le type de fichier est supporté.)



Cela ne signifie pas que l'OCR s'exécute sur le champ: afin que l'utilisateur bénéficie de plus de souplesse, Readiris démarre tout simplement et l'image s'ouvre aussitôt.

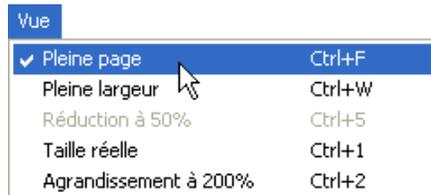
La barre d'outils image à la droite de la fenêtre d'application de Readiris contient toutes les commandes dont vous avez besoin pour la prévisualisation: les outils pour indiquer les zones d'intérêt, pour faire pivoter l'image, pour agrandir l'image, etc.

AGRANDIR L'IMAGE

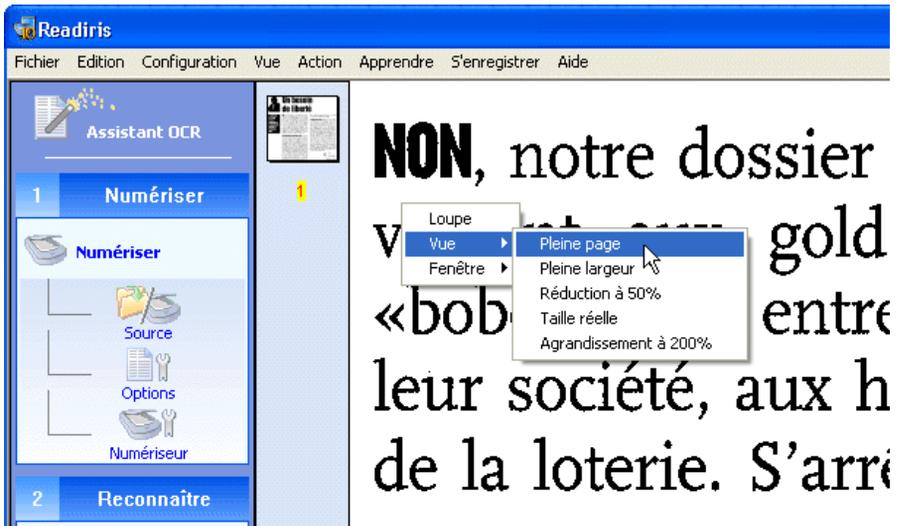
Readiris a plusieurs commandes qui vous permettent d'**agrandir** l'image numérisée, par exemple pour vérifier la qualité de numérisation.



La barre d'outils image contient des boutons qui permettent d'effectuer un zoom en taille réelle, d'ajuster l'image à la largeur de page et d'ajuster l'image entière dans la fenêtre d'aperçu. Le menu "Vue" contient les mêmes commandes et ajoute deux niveaux de zoom supplémentaires: vous pouvez afficher l'image à 50% et 200% de sa taille actuelle. A sa taille actuelle, un point écran correspond à un point image. (Des raccourcis existent pour tous les niveaux de zoom!)



Remarquez aussi que les niveaux d'agrandissement sont accessibles grâce au double clic. Cliquez avec le bouton droit de la souris pour faire apparaître le menu contextuel et sélectionner le niveau de zoom approprié.



De plus, vous pouvez en cliquant *deux* fois avec le bouton droit de la souris sur une région de l'image numérisée en faire directement un agrandissement à sa taille réelle. Refaites l'opération pour revenir à la taille réelle.

Enfin, vous pouvez utiliser la **loupe** pour agrandir les détails des documents numérisés. La loupe est aussi disponible sous le menu "Contexte" lorsque vous cliquez avec le bouton droit de la souris sur l'image.

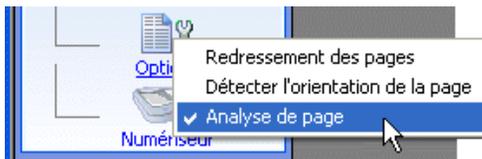




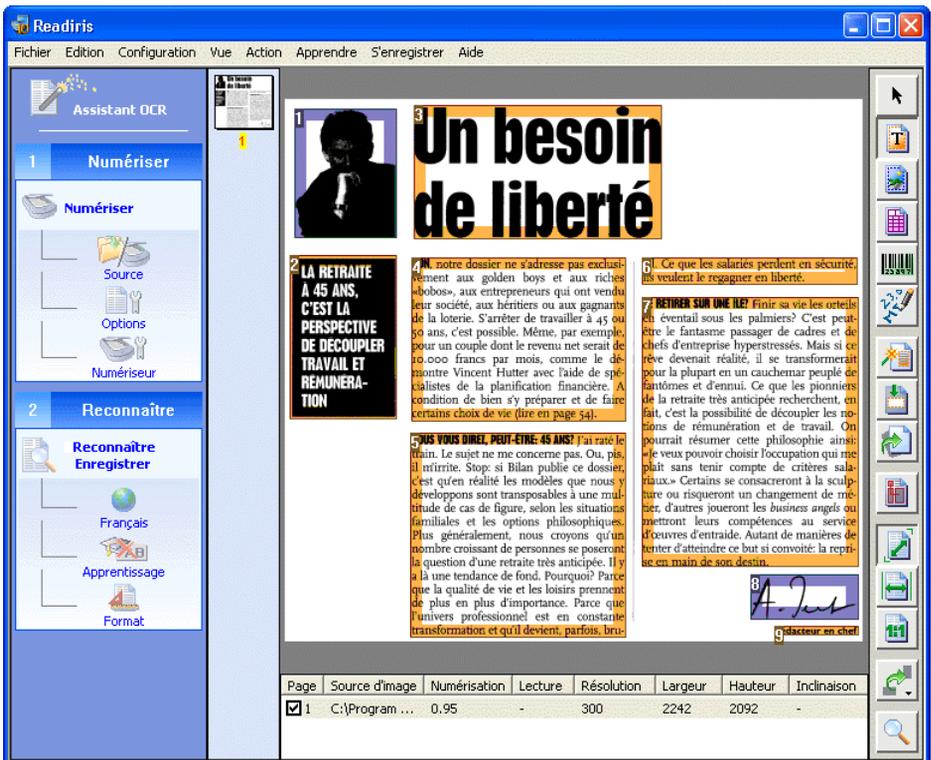
UN, DÉCOMPOSITION D'UNE IMAGE NUMÉRISÉE

Maintenant que l'image est numérisée, vous pouvez définir les parties de textes que vous souhaitez lire en traçant des cadres, également appelées "fenêtres", autour des zones d'intérêt.

En fait, Readiris effectuera automatiquement cette opération quand l'option "Analyse de Page" est activée sous le bouton "Options" de la barre d'outils principale (ou sous le menu "Configuration").



La décomposition automatique de la page est particulièrement utile lorsque des **textes sous formes de colonnes** ou à la mise en page complexe - qui pourrait inclure des graphiques et des tables - sont reconnus.



L'analyse de page utilise trois **types de fenêtres**: les fenêtres textes, les fenêtres graphiques et les fenêtres tables. Readiris distingue les blocs de texte, les tables et les zones graphiques qui contiennent des photos, des illustrations, etc.



(La sauvegarde des graphiques et la reconnaissance des tables seront expliquées plus tard.)

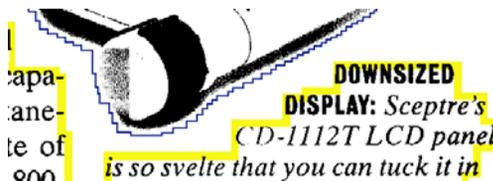
Il y a deux types de fenêtres que vous devez toujours dessiner manuellement: les zones de codes à barres et les zones de texte manuscrit. (Nous parlerons de lecture de codes à barres et de reconnaissance de lettres manuscrites “non cursives” plus tard.)

Un **code couleur** permet de distinguer les types de fenêtre: les fenêtres textes ont un bord orange, les fenêtres graphiques ont un bord pourpre et les fenêtres tables un bord rose. Les fenêtres de codes à barres sont vertes et les fenêtres de texte manuscrit sont bleues.

Le nombre de fenêtres est spécifié à tout moment dans les bulles explicatives des outils des fenêtres.



L’analyse de page est rapide et hautement précise, elle reconnaît des lignes inclinées et elle trace des figures complexes, “irrégulières”.



L’analyse de page détectera même des zones où vous aurez un **texte blanc sur fond noir**. La reconnaissance de ce type d’insertions ne pose pas de problème: l’image apparaît à l’écran telle qu’elle est numérisée, mais Readiris “inverse” l’image pendant la phase de reconnaissance pour reconnaître de tels blocs de texte. (Vous pouvez grâce à votre numériseur générer des images *complètement* inversées pour les pages ayant du texte blanc sur fond noir. Voir plus bas.)

UN ET DEMI, TRIER DES FENÊTRES

Readiris ne détecte pas seulement les différents blocs, mais en plus le logiciel les *trie*: les blocs sont triés du haut vers le bas, de gauche à droite afin de bien traiter les documents multicolonnés.

Vous pouvez évidemment modifier l'**ordre de tri**. Pour ce faire, cliquez sur le bouton "Trier" sur la barre d'outils image. Le curseur prend la forme d'une main qui pointe dès que le mode de tri est activé.



Cliquez à l'intérieur des fenêtres pour les inclure. Les fenêtres sur lesquelles vous *ne* cliquez *pas* sont tout simplement ignorées, exclues de la "lecture". La différence entre les fenêtres sélectionnées et les fenêtres non sélectionnées est claire: les fenêtres sélectionnées ont une couleur plus vive, les fenêtres non sélectionnées ont une couleur plus pâle et ne portent pas de numéro.



Un besoin de liberté

1 LA RETRAITE A 45 ANS, C'EST LA PERSPECTIVE DE DECOUPLER TRAVAIL ET REMUNERATION

2 PLUS VOUS DIREZ, PEUT-ETRE, 45 ANS? J'ai raté le train. Le sujet ne me concerne pas. Ou, pis, il m'irrite. Stop: si Bilan publie ce dossier, c'est qu'en réalité les modèles que nous y développons sont transposables à une multitude de cas de figure, selon les situations familiales et les options philosophiques. Plus généralement, nous croyons qu'un nombre croissant de personnes se poseront la question d'une retraite très anticipée. Il y a là une tendance de fond. Pourquoi? Parce que la qualité de vie et les loisirs prennent de plus en plus d'importance. Parce que l'univers professionnel est en constante transformation et qu'il devient, parfois, bru-

3 tal. Ce que les salariés perdent en sécurité, ils veulent le regagner en liberté.

4 SE RETIRER SUR UNE ILE? Fimir sa vie les orteilés en éventail sous les palmiers? C'est peut-être le fantasme passer de cadres et de chefs d'entreprise hyperstressés. Mais si ce rêve devenait réalité, il se transformerait pour la plupart en un cauchemar peuplé de fantômes et d'ennui. Ce que les pionniers de la retraite très anticipée recherchent, en fait, c'est la possibilité de découpler les notions de rémunération et de travail. On pourrait résumer cette philosophie ainsi: «Je veux pouvoir choisir l'occupation qui me plaît sans tenir compte de critères salariaux». Certains se consacreront à la sculpture ou risqueront un changement de métier, d'autres joueront les *business angels* ou mettront leurs compétences au service d'œuvres d'entraide. Autant de manières de tenter d'atteindre ce but si convoité: la reprise en main de son destin.

A. J. J.
Rédacteur en chef

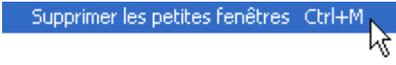
Page	Source d'image	Numérisation	Lecture	Résolution	Largeur	Hauteur	Inclinaison	Rotation
<input checked="" type="checkbox"/> 1	C:\Program Files...	1.89	-	300	2242	2092	-	-

L'analyse de page est activée par défaut. Pour forcer Readiris à décomposer la page activée - ce pourrait être nécessaire si vous avez désactivé l'analyse de page ou supprimé certaines fenêtres par erreur et souhaitez recommencer l'analyse de page, etc. -, cliquez simplement sur le bouton "Analyser la Page" dans la barre d'outils image.



Si vous traitez des documents asiatiques ou hébreux, sélectionnez la langue du document *avant* d'effectuer l'analyse de page car des routines spécifiques propres à la langue sont utilisées. Dans les documents asiatiques, l'espace interligne est généralement plus grand que dans les documents occidentaux, le texte est composé de petites icônes ("idéogrammes") qui pourraient facilement être interprétées comme des zones graphiques dans les documents occidentaux, en outre le texte peut se lire de haut en bas, de droite à gauche. Dans les documents en hébreu, le texte se lit de droite à gauche. Si toutefois vous oubliez de sélectionner la langue adéquate, sélectionnez-la tout de même à posteriori. Readiris fera automatiquement l'analyse de page!

Les images des documents peuvent contenir du bruit (taches "parasites" sur le texte, bord noir autour du texte, etc.). Des petites fenêtres apparaissent alors à l'écran pour indiquer la présence de ce bruit. Si les petites fenêtres ne contiennent aucun texte, vous pouvez les supprimer en cliquant sur la commande "Supprimer les Petites Fenêtres" sous le menu "Edition". Ce faisant, Readiris triera à nouveau les autres fenêtres affichées à l'écran.



Supprimer les petites fenêtres Ctrl+M

December 12, 1990
Dear Friends,
Twenty five years ago this weekend, thousands of people showed up to the polls in my hometown. It was a special election, called by the local Board of Education.
I received your letter that day. Of course, those weren't the exact words (they were "Should Michael Moore be Re-elected to the Board of Education?") but that was pretty much how it felt. As I sat there all day in the same polling place (a school gymnasium), watching the people stream in to cast their ballot, I wondered why I had given up going away to college, just so I could try just to do things some things in the school system. I sat there as those teachers overlooking the voting booths and figured I was just nuts.
Two years prior to this "judgment day," I had become the principal of the school office on the school campus - and the first to give an oral report to public office. A constitutional amendment had just passed, lowering the voting age from 21 to 18. It was Vietnam time, and I guess the adults figured I was going to get killed (and all others for no damn reason). Then they figured I was going to be a war hero and give some talks.
So, as I mentioned to you last year in one of these letters, I was a slightly bigger of someone (this was before the time of school shootings) and I think I convinced the board to let me be the high school principal was to become his boss - which I was still a student. So I ran for the school board on the platform of firing him. Remarkably, I won. Nine months later, the principal was gone.

Une routine similaire opère de façon automatique: la non-détection des zones sur les bords. Quand cette routine est activée (via l'option "Ne Pas Détecter de Fenêtre sur les Bords"), l'analyse de page ignore tout simplement les zones sur les bords de page. Cette option est particulièrement utile parce que certains numériseurs ajoutent des bords noirs autour de l'image numérisée; l'analyse de page détecte normalement des petites fenêtres indiquant la présence de "bruit". Grâce à la routine, les fenêtres inutiles ne s'affichent donc pas! Attention, si les



bords de page contiennent des zones graphiques, photos ou illustrations - par exemple, une illustration en arrière-plan ou qui couvre la page entière, etc. -, Readiris détectera malgré tout correctement ces éléments!

- ✓ Analyse de page
- ✓ Ne pas détecter de fenêtre sur les bords



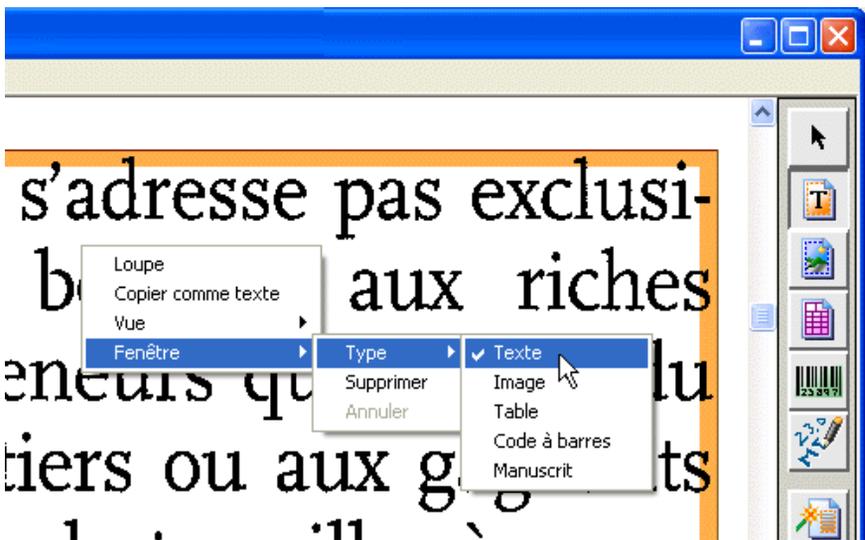
DEUX, FENÊTRAGE MANUEL D'UNE IMAGE

L'analyse de page traite automatiquement le fenêtrage d'une page numérisée, mais vous pouvez dessiner manuellement des zones avec les **outils de fenêtrage** de Readiris.

	Sélectionner une fenêtre
	Dessiner une fenêtre texte: 0
	Dessiner une fenêtre graphique: 0
	Définir une fenêtre table: 0
	Dessiner une fenêtre de codes à barres: 0
	Dessiner une fenêtre de texte manuscrit: 0

(Comme nous avons déjà indiqué, les fenêtres de codes à barres et de texte manuscrit sont toujours dessinées à la main par l'utilisateur: l'analyse de page ne les détecte pas!)

Pour **tracer** un rectangle autour d'une zone d'intérêt, sélectionnez l'outil correspondant dans la barre d'outils image et glissez la souris du coin supérieur gauche vers le coin inférieur droit de la zone. (Les côtés du rectangle ne peuvent être inférieurs à 1 mm car ils ne seraient pas suffisants pour contenir les caractères.)

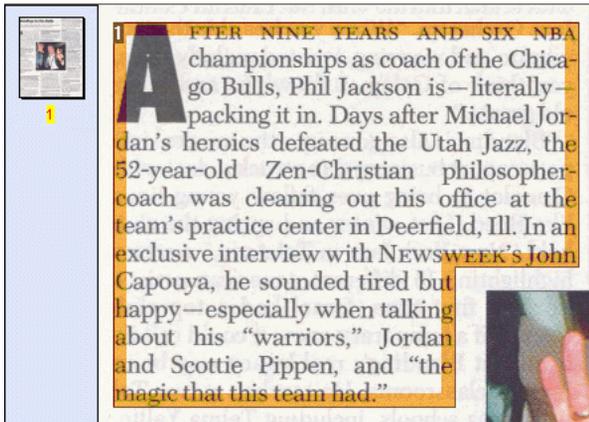


Les fenêtres sont automatiquement triées par ordre de création; des numéros indiquent l'ordre de tri.

Vous pouvez également encadrer des blocs de texte "irréguliers" en dessinant des **fenêtres polygonales** autour d'eux. Les fenêtres non rectangulaires sont créées par l'assemblage de zones rectangulaires: dès que deux rectangles (du même type) entrent en intersection, ils deviennent automatiquement une seule et



même fenêtre! En quelque sorte, vous construisez une maison en ajoutant une pièce à l'autre... (Attention, créer des fenêtres de tables ou de codes à barres polygonaux n'a pas de sens.)



Notez également que le fenêtrage manuel peut être combiné avec le tri des fenêtres: vous pouvez tracer de nouvelles fenêtres pendant que le “mode de tri” est activé. Vous utilisez alors le tri pour inclure certaines fenêtres détectées, et créez manuellement d'autres fenêtres là où l'analyse de page n'a pas obtenu les résultats idéaux. Dès que vous commencez à créer des fenêtres en “mode de tri”, toutes les fenêtres non sélectionnées sont aussitôt effacées!

Pour modifier, déplacer ou supprimer des fenêtres, il vous faut d'abord les **sélectionner**. Pour ce faire, cliquez sur l'icône "Sélectionner une Fenêtre" (la flèche) sur la barre d'outils image et cliquez à l'intérieur de la fenêtre. Des marqueurs rectangulaires apparaissent alors à chaque coin et sur les bords de la fenêtre.

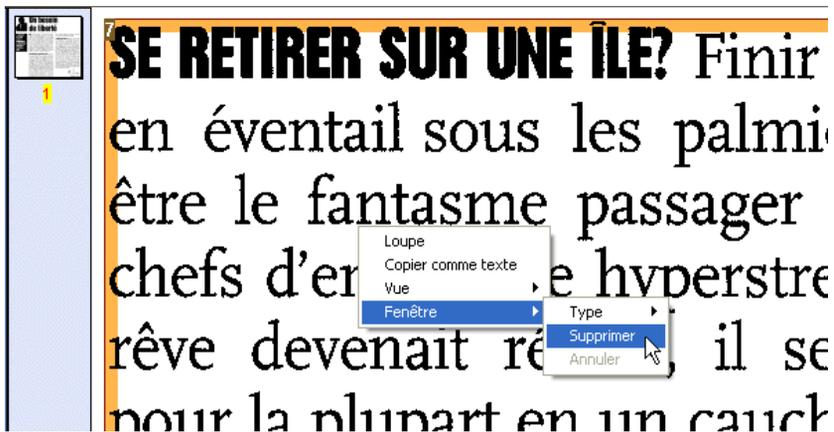


Pour **annuler la sélection** des fenêtres, cliquez ailleurs avec le bouton de la souris. Pour sélectionner des **fenêtres supplémentaires**, maintenez appuyée la touche Maj tout en cliquant sur ces fenêtres supplémentaires. Pour sélectionner une fenêtre et les **fenêtres incluses** (d'un autre type), maintenez enfoncée la touche Ctrl tout en cliquant sur la fenêtre principale.

Voici pour la sélection des fenêtres. Pour **modifier** une fenêtre, sélectionnez-la, placez le curseur de la souris sur un marqueur et faites glisser le bord pour changer la taille de la fenêtre.

Pour **déplacer** une fenêtre, sélectionnez-la et faites-la simplement glisser vers l'endroit désiré.

Pour **supprimer** des fenêtres, sélectionnez-les, cliquez sur celles-ci avec le bouton droit de la souris et sélectionnez la commande "Fenêtre - Supprimer" du menu "Contexte". Cette action supprime aussi bien toutes les fenêtres sélectionnées que la fenêtre sous le curseur de votre souris.



Alternativement, vous pouvez sélectionner les zones et choisir la commande "Couper" ou "Effacer" du menu "Édition". La commande "Couper" place votre fenêtre dans une mémoire tampon interne, la commande "Effacer" la supprime



définitivement. Lorsque vous collez des fenêtres, elles sont réintégréées à leur place originale et vous devez les déplacer vers l'endroit souhaité.

En fait, *toutes* les commandes habituelles du menu "Édition" s'appliquent aux fenêtres: vous pouvez les supprimer, les couper, les copier et les coller! La commande "Annuler" peut également vous servir: si vous avez malencontreusement supprimé, déplacé, redimensionné, etc. certaines fenêtres, "Annuler" annulera votre dernière opération!

Edition	
Annuler	Ctrl+Z
Couper	Ctrl+X
Copier	Ctrl+C
Coller	Ctrl+V
Effacer	Del
Supprimer les petites fenêtres	Ctrl+M
Sélectionner tout	Ctrl+A

Notez aussi que des raccourcis existent pour toutes les commandes! Par exemple: pour supprimer toutes les fenêtres existantes, vous pouvez utiliser la commande "Sélectionner Tout" du menu "Édition" ou son raccourci clavier Ctrl+A et cliquer sur la commande "Effacer" ou son raccourci Supprimer. Vous pouvez maintenant recréer la mise en page désirée. Pour restaurer la mise en page antérieure, choisissez la commande "Annuler" ou le raccourci Ctrl+Z.

TROIS, ENREGISTRER DES MODÈLES DE FENÊTRAGE

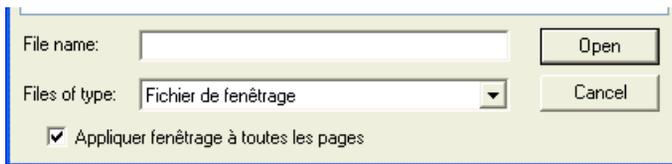
Le masque créé peut être enregistré en **gabarits de fenêtrage** pour une utilisation ultérieure avec la commande "Sauvegarder le Fenêtrage" du menu "Fichier" et chargée en mémoire avec la commande "Charger un Fenêtrage".



Si vous devez reconnaître des documents avec une mise en page similaire, par exemple un rapport de 50 pages où les en-tête et pied de page doivent être

supprimés pour des raisons évidentes, un simple modèle peut servir à aménager les 50 pages.

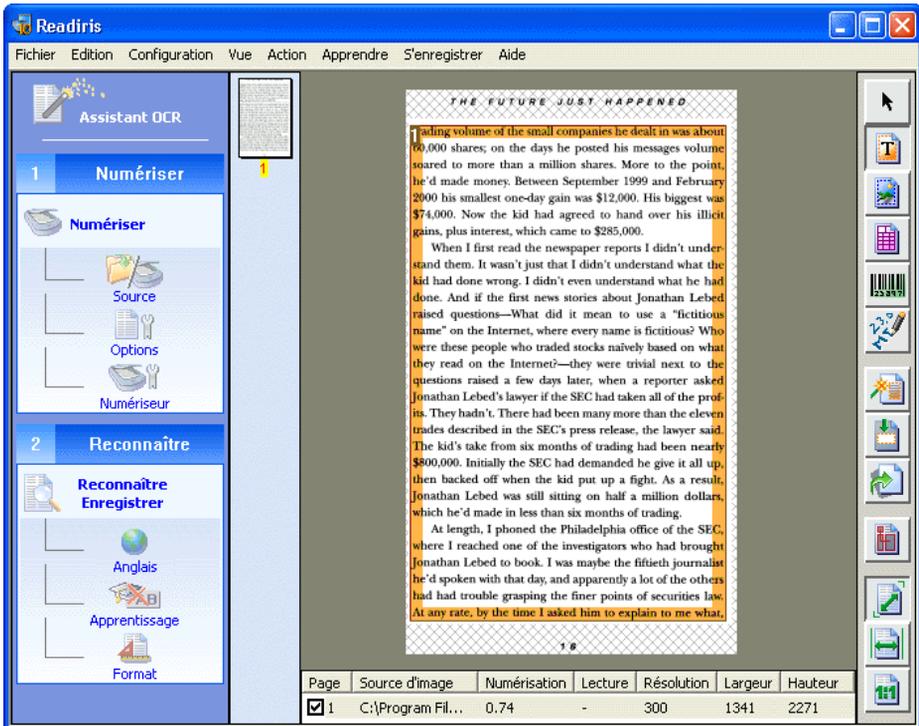
Lorsque vous chargez un fenêtrage fixe, l'analyse de page se désactive automatiquement et le gabarit choisi reste actif jusqu'à ce que vous réactiviez l'analyse de page sur la barre d'outils principale. Si vous chargez un fenêtrage mis en mémoire, activez l'option "Appliquer Fenêtrage à Toutes les Pages" pour appliquer ce fenêtrage directement à toutes les pages du document en cours.)



De plus, vous disposez d'une alternative très intéressante aux gabarits: l'outil de prévisualisation "Ignorer la Zone Extérieure" limite la décomposition de la page à la partie "encadrée" de l'image.



Sélectionnez cet outil et saisissez la portion de l'image que vous voulez traiter. Si vous utilisez un document multipage, vous pouvez exclure la même zone externe de l'analyse de page à chaque page. (Relancez l'analyse de page pour supprimer la "découpe" de l'image, ou pour changer les zones manuellement.)



READIRIS VOUS EMMÈNE AUTOUR DU MONDE

Maintenant que les fenêtres sont correctement définies, vous êtes presque prêt à effectuer la reconnaissance de caractères. “Presque” parce que nous n’avons pas encore vérifié la langue et les caractéristiques du document!

Le paramètre de langue se trouve sur la barre d’outils principale.



Cliquez sur le bouton "Langue" pour modifier la langue du document.



Vous pouvez appuyer sur une touche lettre pour la sélectionner directement: si le français est actuellement sélectionné et que vous souhaitez choisir l'occitan, vous pouvez enfoncer la lettre "O" de votre clavier: l'occitan sera sélectionnée. Quand différentes langues ont la même initiale, appuyez sur la lettre plusieurs fois afin de trouver les options. Par exemple: Readiris lit le polonais et le portugais. En appuyant sur "P" une fois, vous sélectionnez le polonais, en appuyant une seconde



fois sur la touche "P", vous sélectionnez le portugais, et en appuyant une troisième fois sur "P", vous revenez au polonais. (Pour sélectionner une *autre* lettre, par exemple "D", appuyez sur Arrière avant d'appuyer sur la touche "D".)

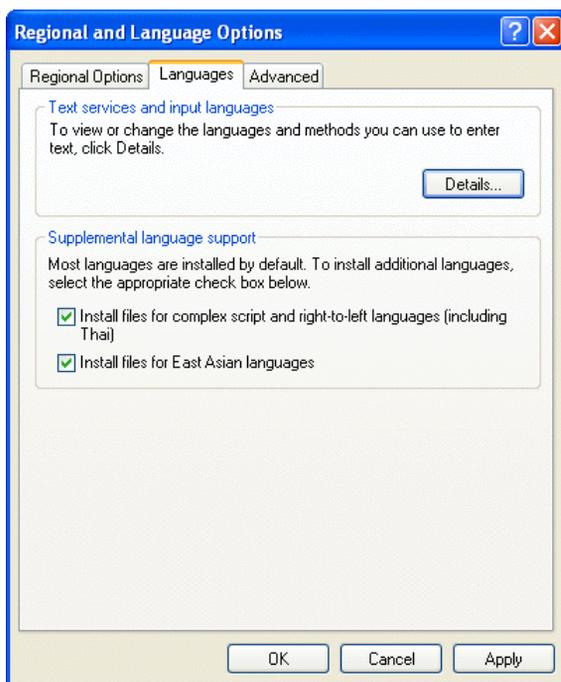
Readiris est de loin limiter au français: il reconnaît jusqu'à 117 **langues**! Toutes les langues européennes et américaines sont supportées, y compris les langues d'Europe centrale, le grec, le turc, les langues cyrilliques ("russes") et les langues de l'Europe baltique.

Un module optionnel permet de lire les **documents asiatiques** et **hébreux**: le module supplémentaire "Complément d'OCR hébreux" propose la reconnaissance de l'hébreux, l'option logicielle "Complément d'OCR asiatique" propose la reconnaissance du japonais, du chinois simplifié, du chinois traditionnel et du coréen. (Le chinois simplifié est utilisé sur le continent chinois et à Singapour, le chinois traditionnel est employé à Hong Kong, Taiwan, Macau et dans les communautés chinoises d'outre-mer.)

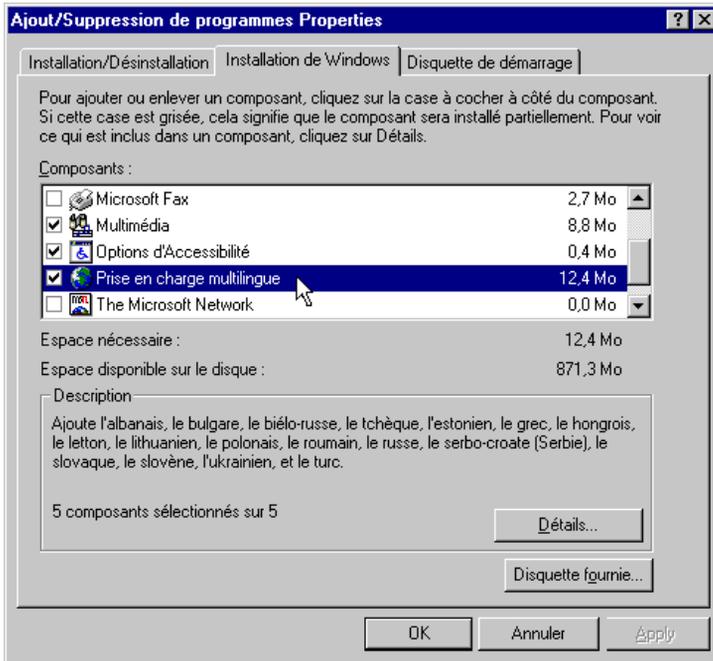
Notez également que l'on distingue les variantes britanniques et américaines - "internationales" devrions-nous dire - de la langue anglaise. Le logiciel permet également de distinguer l'espagnol et le mexicain.

Une configuration de Windows particulière est obligatoire pour afficher des caractères d'Europe centrale, grecs, turcs, cyrilliques et baltes. Vous pourriez avoir à installer le "module" "**Prise en Charge Multilingue**" de Windows avant que votre système Windows ne puisse s'adapter à ces langues.

Sur un système d'exploitation Windows XP, 2000 et Windows NT 4.0, sélectionnez l'icône "Options Régionales (et Langues)" sous le "Panneau de Configuration".



Sur un système d'exploitation Windows ME et 98, sélectionnez l'icône "Ajout/Suppression de Programmes" dans le "Panneau de Configuration" pour vérifier si le module de prise en charge multilingue est installé sur votre PC.



Pour afficher et d'éditer des documents asiatiques et hébreux, vous devez installer une version asiatique ou une version en hébreu du système Windows. Vous pouvez également utiliser Word 2003, Word 2002 ou Word 2000 pour visualiser et modifier ce genre de documents: Office 2003 System, Office XP et Office 2000 ont été spécifiquement conçus pour traiter des documents dans une grande diversité de langues. Référez-vous au **fichier "Lisez-moi"** de Readiris pour plus d'informations à ce sujet.

Sélectionner correctement la langue du document est impératif. Basé sur la sélection d'une langue, le logiciel sait quel **jeu de symboles** il doit reconnaître.

Un support multi-linguistique vous assure une reconnaissance correcte de caractères "exotiques" tels que ß, ñ, ψ et ø.

Deuxièmement, le logiciel utilise de façon extensive ses **bases de données linguistiques** pour valider ses résultats. Supposez que vous ayez à lire le mot "président" où une tache d'encre donne au "r" l'aspect d'un "f". En consultant le lexique de la langue française, Readiris reconnaît de façon autonome le mot "président" et conclut qu'il n'y a aucun sens à la présence du symbole "f". Cette **technique "d'auto-apprentissage"** dépend étroitement du contexte linguistique.

La linguistique offre une aide utile pour résoudre des **cas ambigus** tels qu'un "O" qui pourrait être confondu avec '0'. Un autre cas typique est la lettre "l" et le chiffre '1' qui ont une forme identique dans de nombreuses fontes - pensez aux textes produits sur d'anciennes machines à écrire! Le contexte linguistique aide à déterminer s'il s'agit d'un "l" ou d'un '1'.

L'illustration ci-dessous montre différentes formes de "l" et de '1'. Les formes sur la première ligne ne sont pas ambiguës, les formes sur la deuxième ligne peuvent prêter à confusion, mais la linguistique peut résoudre leur interprétation. Lorsque le contexte ne suffit pas, l'utilisateur intervient.

193 1950s. 1hr
Well, Rossellini

ALLEZ PLUS LOIN AVEC LES LEXIQUES D'UTILISATEUR

Renforcez la linguistique de vos textes en la personnalisant. Aussi puissants que peuvent être les lexiques standards, les utilisateurs de Readiris Corporate ont la possibilité d'améliorer encore la **précision de l'OCR** en chargeant des lexiques d'utilisateur avec la commande "Lexiques d'Utilisateur" sous le menu "Configuration".

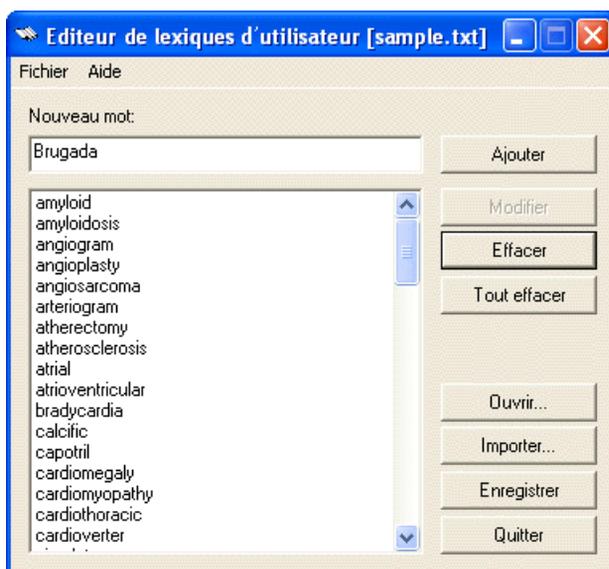


La bulle explicative du bouton "Langue" indique quel lexique d'utilisateur est activé.

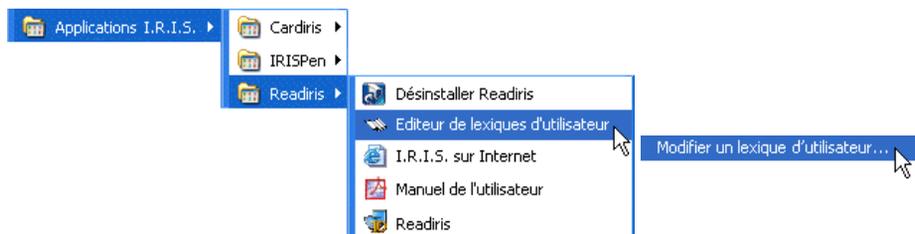


Les lexiques d'utilisateur sont des listes de mots contenant des termes qui n'apparaissent pas dans les lexiques de base, tout usage. Songez par exemple aux **termes** techniques, scientifiques, légaux et **spécifiques aux entreprises!**

Readiris est accompagné d'un Editeur de Lexiques d'Utilisateur, une utilité mise au point pour créer et maintenir de tels **lexiques d'utilisateur**. Cet outil est très facile à employer. Consultez son aide en ligne si vous avez le moindre doute quant à son utilisation.



Vous trouverez cet éditeur dans le sous-menu "I.R.I.S. Applications - Readiris" et sous le menu "Configuration" de Readiris.

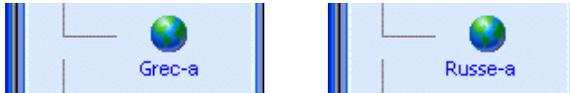




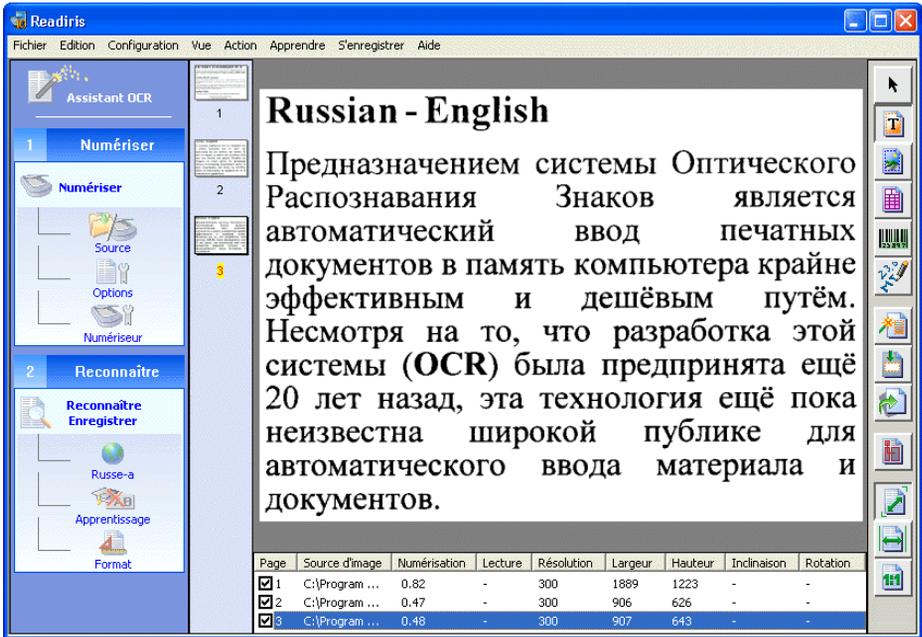
READIRIS CHANGE DE LANGUE AU BESOIN

Les performances de Readiris ne se limite pas à cela: Readiris peut changer de langue en cours de phrase sans l'intervention de l'utilisateur! Quand des mots occidentaux apparaissent dans des documents grecs, cyrilliques, hébreux ou asiatiques - beaucoup de noms propres non modifiables, des noms de marques, etc. sont écrits en utilisant les symboles latins -, Readiris peut s'adapter automatiquement au type d'alphabet. En d'autres termes, il est possible d'activer des **alphabets mélangés** comportant à la fois des caractères grecs, cyrilliques, hébreux ou asiatiques, et des caractères occidentaux.

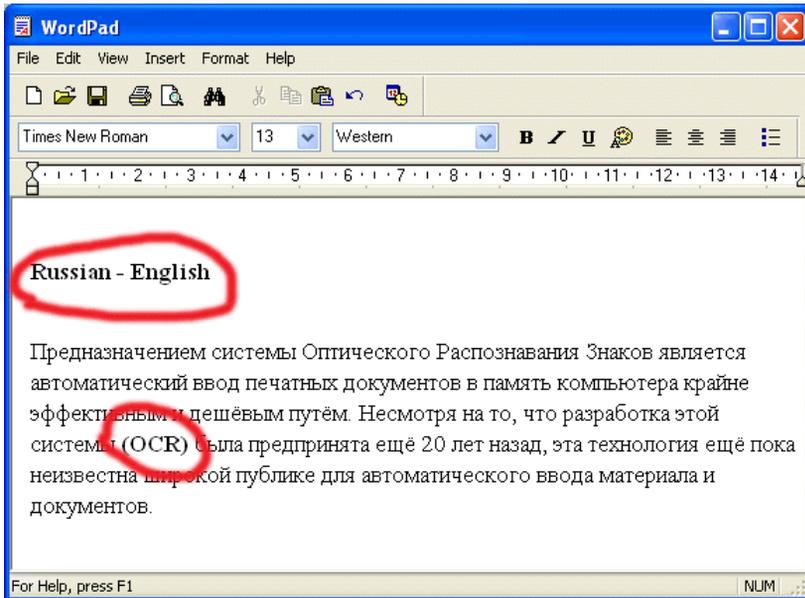
Assurez-vous de sélectionner "Grec-anglais" ou l'option correcte cyrillique - par exemple "Biélorusse-anglais". En d'autre mots: ne vous contentez pas juste de sélectionner "Grec" ou "Biélorusse" en espérant que les caractères occidentaux seront reconnus correctement!



Voici l'exemple d'un texte russe contenant des mots anglais. Vous pouvez tester par vous-même cette page: ouvrez le fichier image ALPHABETS.TIF et lancez la reconnaissance de la page correspondante.



Le résultat final ressemble à ceci lorsque vous l'ouvrez avec votre traitement de texte - vous pourriez avoir à sélectionner la **fonte** cyrillique pour pouvoir afficher le texte russe correctement.



LIRE DES DOCUMENTS COMPRENANT PLUSIEURS LANGUES

Readiris Corporate est plus puissant quand il s'agit de lire des documents couvrant différentes langues. Avec cette version, vous pouvez sélectionner une langue principale et (jusqu'à 4) **langues secondaires** (du même groupe linguistique).



Pour lire un document où la traduction française se trouve le long du texte anglais, vous devez sélectionner le français comme langue principale et l'anglais comme langue secondaire. Mais sachez que vous ne faites pas seulement qu'étendre votre jeu de caractères en y incluant des caractères accentués tels que ç, é et ù. Les deux bases de données linguistiques fournissent également un "retour" linguistique au processus OCR, ce qui permet à Readiris "d'auto-détecter" quelle langue est utilisée à tel ou tel endroit dans votre document!

Vous pouvez sélectionner jusqu'à 5 langues simultanément. Mais seules les langues appartenant au même groupe peuvent être activées ensemble. Les langues couvertes par un autre code de page Windows que celui de la langue principale *ne* peuvent *pas* être activées en même temps. Avant tout, ne choisissez pas une langue qui n'apparaît pas dans votre document: plus le jeu de caractères est grand, plus la reconnaissance est lente et plus le risque d'erreurs est important!



Si vous voulez rapidement savoir quelles langues sont actuellement sélectionnées, la bulle explicative du bouton "Langue" les énumère...



DÉFINIR LES CARACTÉRISTIQUES DU DOCUMENT

Maintenant que la langue est définie, allons voir les autres caractéristiques du document. Vous pouvez peaufiner la reconnaissance en indiquant quelques spécifications du document: le type de fonte et l'espacement de caractère. (Ces commandes ne s'appliquent pas aux documents asiatiques.) Qu'est-ce que cela veut dire?

Commençons par la commande "Type de Police" dans le menu "Configuration". Ce mode distingue les documents "normaux" des **documents imprimés** en qualité **matricielle**. Les symboles matriciels "qualité brouillon" ou "9 pin" sont composés de points isolés, séparés, et des routines de reconnaissance hautement spécialisées sont requises pour les reconnaître.

ape--descended life

La qualité d'impression matricielle "qualité lettre" ou "25 pin" exige l'option normale "Automatique", de même que les **qualités d'impression** imprimées, tapées, impression laser et impression à jet d'encre.

L'option "Automatique" permet à Readiris de détecter automatiquement le type de police utilisée. Laissez dans tous les cas Readiris détecter automatiquement le type de police - à moins que vous ne soyez certain de reconnaître des documents matriciels! (Bien entendu, "Automatique" est la valeur par défaut.)



La bulle explicative du bouton "Reconnaître-Enregistrer" indique le mode de lecture: lorsqu'aucun message n'est ajouté à la bulle explicative, la détection automatique de la qualité d'impression s'applique, lorsque le message "À aiguilles" est ajouté à la bulle, le mode matriciel est activé.



L'**espacement de caractères** peut être placé avec la commande "Distance des Caractères" sous le menu "Configuration".



Avec les fontes *fixes* ou "mono-espace", tous les symboles de la fonte ont la même largeur. Un "l" prend autant d'espace horizontal sur une ligne qu'un "w", comme c'est le cas dans cette phrase. Pensez aux documents produits par une machine à écrire, où le chariot se déplace selon une distance fixe pour chaque symbole tapé.

Un espacement *proportionnel* signifie que la largeur d'un caractère dépend de sa forme. Les symboles comme "m" et "w" sont plus larges, occupent un espace horizontal plus large sur une ligne que les caractères "minces" "l" ou "j". Pratiquement tous les livres, magazines et journaux sont imprimés avec un espacement proportionnel.

La solution la plus simple est de toujours laisser cette option activée par défaut sur "Automatique": Readiris détectera la largeur de caractère automatiquement.



READIRIS DEVIENT CHAQUE FOIS PLUS INTELLIGENT!

Quand la langue du document est choisie et ses caractéristiques définies, activez l'apprentissage interactif et cliquez sur le bouton "Reconnaître-Enregistrer".



La progression de l'OCR est indiquée sur l'écran. Vous pouvez cliquer le bouton "Arrêter" pour interrompre la reconnaissance du texte.



En fin de lecture, Readiris entre en phase interactive d'apprentissage si l'apprentissage est activé grâce au bouton "Apprentissage" sur la barre d'outils principale.

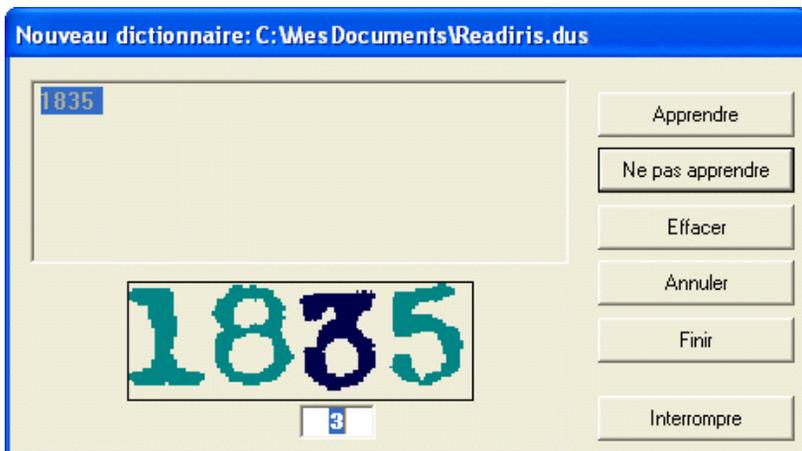
(L'apprentissage interactif ne s'applique pas aux documents asiatiques: l'apprentissage n'a pas de sens pour ces langues qui connaissent des milliers de symboles différents - en supposant que vous êtes capable d'introduire les idéogrammes, ce qui n'est pas chose aisée sur un clavier occidental!)

L'**apprentissage de fontes** peut sensiblement mettre en valeur l'exactitude du système d'identification. Quand l'utilisateur essaye de lire des symboles distordus

et déformés comme on en trouve dans des documents réels ou des fontes stylisées que Readiris n'identifie pas de façon optimale, l'apprentissage peut surmonter cet "échec" provisoire.

L'apprentissage par l'utilisateur est également utilisé pour entraîner le système sur les **symboles spéciaux** que Readiris ne peut pas identifier, comme des symboles mathématiques et scientifiques et des dingbats. Quelques exemples: Readiris peut être formé pour identifier le symbole " π " comme "pi" ou le caractère spécial "☎" en tant que "Tél.". (Cependant, la liste de symboles identifiés ne peut pas être étendue avec les symboles " π " et "☎"!)

Le texte reconnu s'affiche progressivement et le système s'arrête sur les caractères douteux, ou - si vous traitez les caractères joints ("ligatures") - sur les chaînes de caractères douteuses. Elles sont toujours présentées dans leur contexte et les caractères douteux sont mis en surbrillance. Les caractères non reconnus sont représentés par un tilde (le symbole "~").



La première chose à faire est de vérifier si vous avez activé le dictionnaire de fontes et le mode de dictionnaire corrects - ceux-ci sont toujours indiqués dans le titre de la fenêtre d'apprentissage. Si ce n'est pas le cas, cliquez sur le bouton



"Interrompre" - l'image du document est à nouveau affichée avec son fenêtrage tel qu'il avait été crée - activez le bon dictionnaire ou mode de dictionnaire et relancez l'OCR. (Le fonctionnement des dictionnaires de fontes sera expliqué sous peu.)

Si nécessaire, entrez un caractère (ou une chaîne de caractères) pour une forme incorrecte ou inconnue et cliquez sur un des boutons suivants.

Apprendre

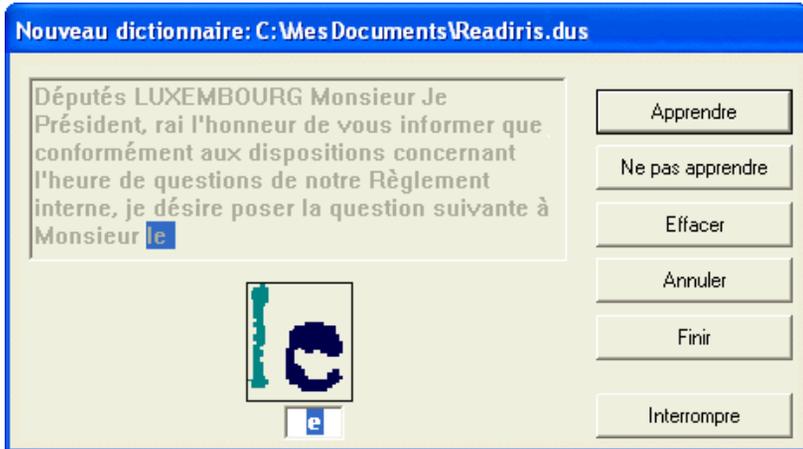
Vous êtes d'accord avec la solution proposée ou vous la corrigez. Le programme sauvegarde le caractère douteux dans le dictionnaire de fontes comme étant "certain", final. La reconnaissance suivante n'exigera plus votre intervention, la forme est considérée apprise une fois pour toutes.

Dans l'exemple ci-dessus, le système s'arrête sur un caractères abîmé, et nous cliquons sur "Apprendre" pour accepter une forme qui ne peut pas être confondue avec d'autres caractères.

Ne Pas Apprendre

Vous êtes d'accord avec la solution proposée ou vous la corrigez. La différence avec le bouton "Apprendre" est que le symbole appris obtient le statut "incertain" dans le dictionnaire. Pour une future identification, le système proposera la solution "apprise" mais exigera toujours une confirmation.

Ce bouton est utilisé pour les symboles qui pourraient être confondus avec d'autres: un "e" déformé qui pourrait être confondu avec un "c", un "t" endommagé qui ressemblerait étroitement à un "r", etc.



Le "e" ci-dessus est sérieusement endommagé - en fait il est proche du symbole "c" -, et vous devriez cliquer "Ne Pas Apprendre" pour ne pas le confondre avec le symbole "c".

Effacer

On élimine la forme affichée du résultat. Ce bouton est utilisé pour ignorer le "bruit" sur les documents - taches de café, etc. qui pourraient être reconnues comme points, virgule, etc. -, et pour effacer tout autre symbole non désiré.

Annuler

Vous retournez en arrière pour corriger les erreurs. Vous pouvez annuler les 32 dernières opérations.

Finir

L'apprentissage est interrompu mais l'OCR continue en mode automatique. Toutes les décisions du système sont acceptées sans validation de l'utilisateur.



Cliquez sur ce bouton quand vous voyez que la reconnaissance est précise et n'exige pas de correction approfondie.

Interrompre

Ne confondez pas le bouton "Finir" avec le bouton "Interrompre". Ce dernier arrête le processus alors qu'avec le bouton "Finir", le texte est créé, il ne subit simplement plus de correction approfondie!

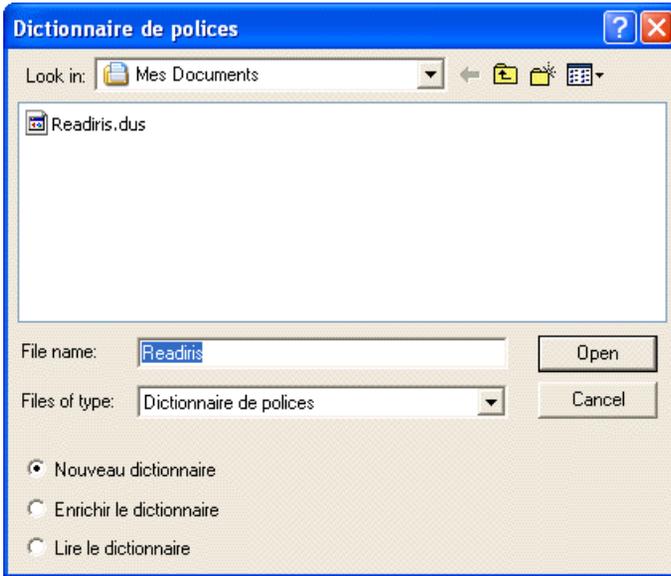
LE RÔLE DES DICTIONNAIRES DE FONTES

Les résultats de chaque session de formation sont temporairement maintenus dans la mémoire mais ils peuvent et doivent être enregistrés dans des fichiers appelés les "dictionnaires" pour une future utilisation.

(Ne confondez pas dictionnaires de polices et lexiques! Les dictionnaires de fontes contiennent des formes de caractères apprises durant la phase interactive de l'OCR, les lexiques sont des bases de données linguistiques qui assistent la reconnaissance.)

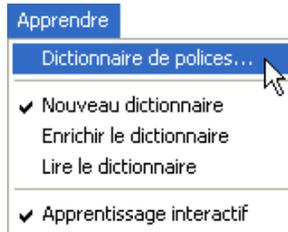
Ces dictionnaires de polices doivent être chargés en mémoire quand vous voulez reconnaître des documents semblables afin d'utiliser l'intelligence supplémentaire qu'ils contiennent. De cette façon, Readiris tient compte de l'intelligence enregistrée dans ces bibliothèques de polices. Vous pourriez dire que Readiris devient plus intelligent à chaque utilisation!

Comment cela se fait-il? L'enrichissement des dictionnaires de fontes est contrôlé par le menu "Apprendre": vous devez choisir un dictionnaire avec la commande "Dictionnaire de Polices" et déterminer son mode de fonctionnement.



Les **dictionnaires de polices** sont limités à 500 formes, et nous vous recommandons de créer des dictionnaires séparés pour des applications spécifiques, par exemple par type de document. Les dictionnaires ont par défaut l'extension *.dus. L'apprentissage et l'enrichissement n'ont plus d'effet quand le dictionnaire est rempli: les résultats de l'apprentissage ne sont plus mémorisés et stockés dans un dictionnaire.

Vous pouvez sélectionner le mode dictionnaire sous la commande "Dictionnaire de Polices" ou directement sous le menu "Apprendre". Trois modes de dictionnaires sont disponibles: nouveau, enrichir et lire.



En choisissant "Nouveau Dictionnaire", vous indiquez que les résultats d'apprentissage seront sauvegardés dans un *nouveau* dictionnaire. (Si vous choisissez un dictionnaire existant, son contenu sera effacé.)

Le mode d'enrichissement indique que les résultats d'apprentissage seront sauvegardés dans un dictionnaire *existant*: la reconnaissance se sert de l'intelligence déjà contenue dans le dictionnaire, et vous lui ajoutez les nouvelles formes de fonte. En termes simples, cette option vous permet de construire un dictionnaire de fontes en plusieurs étapes.

(Quand vous définissez le nom de fichier d'un nouveau dictionnaire et activez le mode d'ajout, un dictionnaire vide est créé et vous le complétez.)

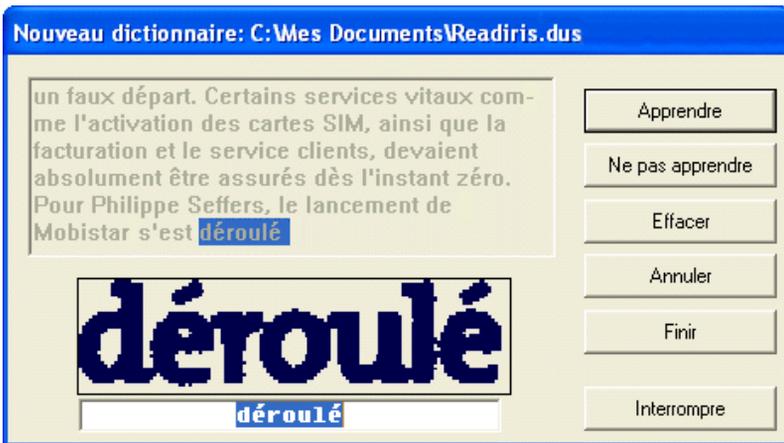
Avec la dernière option, "Lire le Dictionnaire", le dictionnaire fonctionne en mode inaltérable: vous vous servez du dictionnaire *sans* lui ajouter de nouvelles formes de fonte.

Choisissez le nouveau mode quand une page simple est reconnue. Pour lire beaucoup de pages du même type - des pages avec une qualité d'impression et des fontes similaires - choisissez le mode nouveau pour la première page, le mode ajout pour les pages suivantes et le mode lecture pour le reste du document.

Sachez que votre bouton d'apprentissage vous indiquera à tout moment grâce à sa bulle explicative quel est le dictionnaire de fontes actif et dans quel mode ce dictionnaire fonctionne.



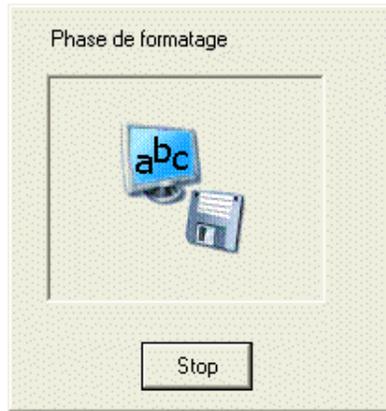
Quand vous êtes en mode d'apprentissage interactif, le dictionnaire et son mode opératoire sont indiqués en titre de fenêtre; en cas d'erreur, vous pouvez cliquer sur "Interrompre" et recommencer.



ENVOYER LE RÉSULTAT DANS VOTRE APPLICATION

L'apprentissage interactif vient en conclusion à la reconnaissance de caractères. Comme Microsoft Word fonctionne comme application cible par défaut, votre traitement de texte est démarré automatiquement (si nécessaire) à la fin de la reconnaissance, et le texte reconnu est inséré.

Il est possible de voir une barre de progression apparaître à l'écran lors du traitement du texte reconnu. (Que cette barre apparaisse ou non à l'écran dépend de la taille du document et de la complexité du traitement à réaliser.)

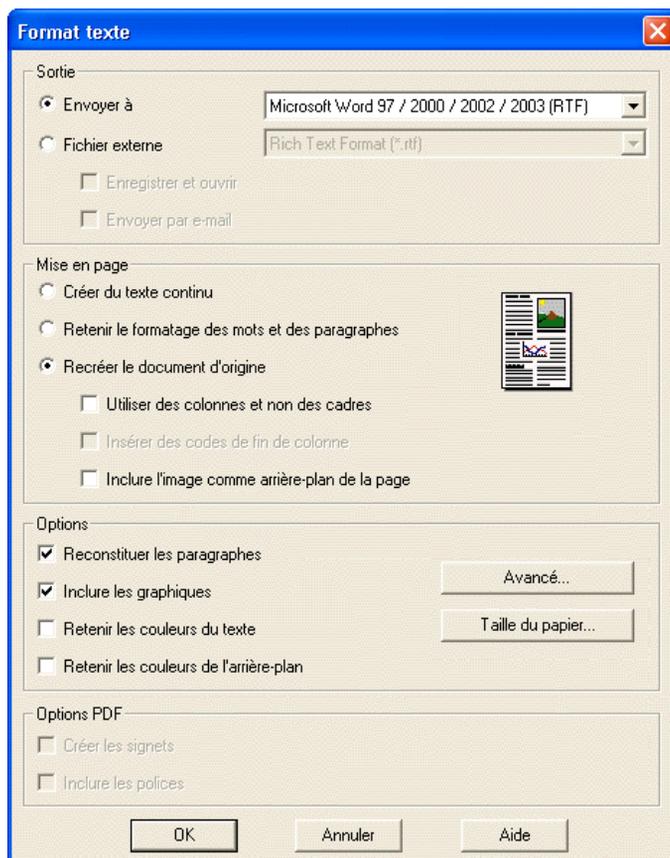


L'image numérisée est à nouveau affichée avec son fenêtrage. L'image est disponible pour la poursuite du traitement, elle reste affichée jusqu'à ce que vous numérisiez une autre page.

Vous venez en effet de convertir un document papier en un fichier ordinateur éditable, et ce 40 fois plus vite qu'en dactylographie manuelle! Vous pouvez maintenant comparer votre texte à l'image qui se trouve dans votre fenêtre Readiris.

En fait, Readiris vous offre trois méthodes pour enregistrer le résultat OCR: envoyer le document reconnu dans une application cible, sauvegarder les fichiers dans un fichier externe et placer les résultats dans le presse-papiers Windows.

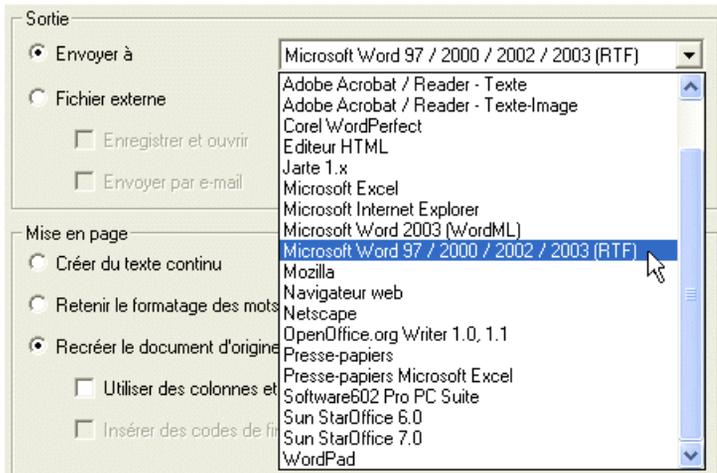
La **cible sortie** est sélectionnée avec le bouton "Format" de la barre d'outils (ou la commande "Format de Texte" du menu "Configuration").



L'option "Envoyer à" offre un lien OCR direct entre votre numériseur et vos applications Windows: vous **envoyez** les documents numérisés directement dans



vosre traitement de texte, dans votre feuille de calcul ou votre navigateur web, dans Adobe Reader, etc.!



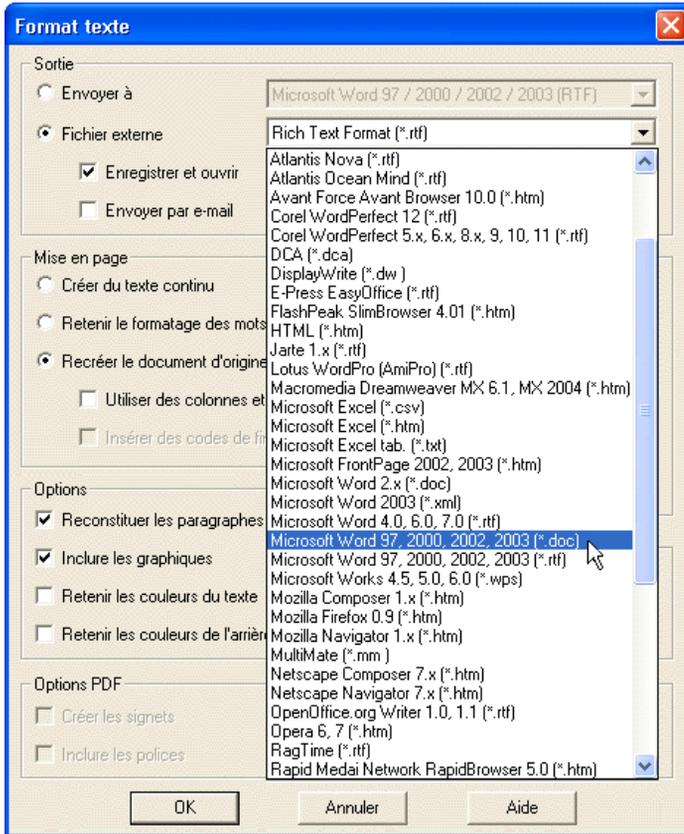
À la fin de la reconnaissance, l'application cible démarre automatiquement et le document reconnu est inséré dans un nouveau fichier texte ou feuille de calcul.



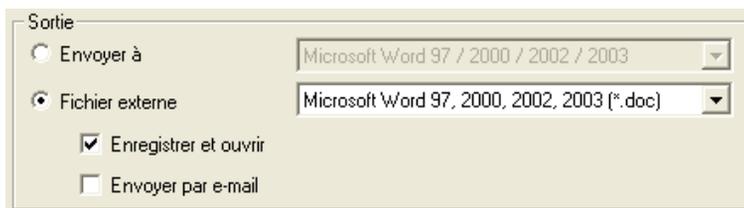
Cependant, n'oubliez pas que l'option "Envoyer à" permet également de copier le document reconnu dans le **presse-papiers** Windows. Il n'y a donc pas de besoin impérieux d'exporter le résultat... ou de l'enregistrer dans un fichier externe!

SAUVEGARDER LES RÉSULTATS DANS UN FICHER TEXTE

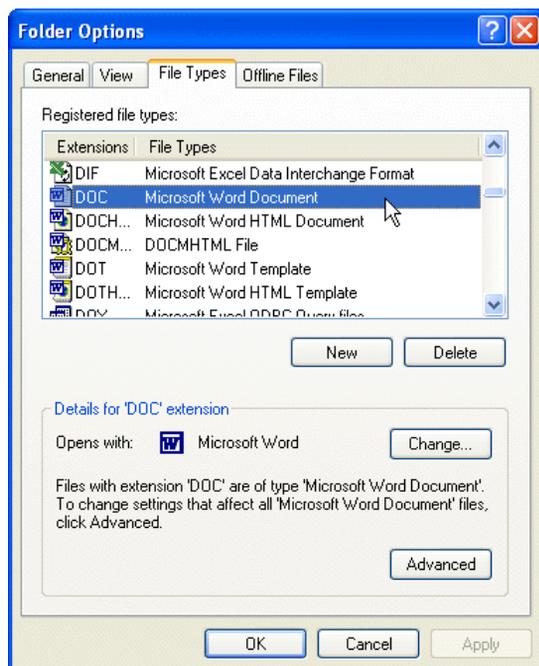
Vous pouvez en effet placer le résultat du processus OCR dans un fichier "externe"! Ici encore, Readiris supporte un grand nombre de formats de fichiers (fichiers de tous les traitements de texte, des feuilles de calcul, des applications web courantes, etc.) (Readiris supporte entre autres le format WordML qui est le nouveau format de texte de Microsoft Office 2003!)



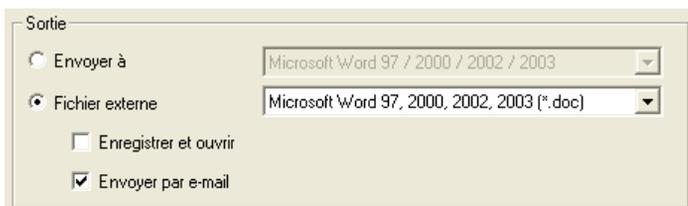
L'option "Enregistrer et Ouvrir" est en grande partie identique à la fonction "Envoyer": vous ouvrez le document reconnu une fois qu'il est sauvé.

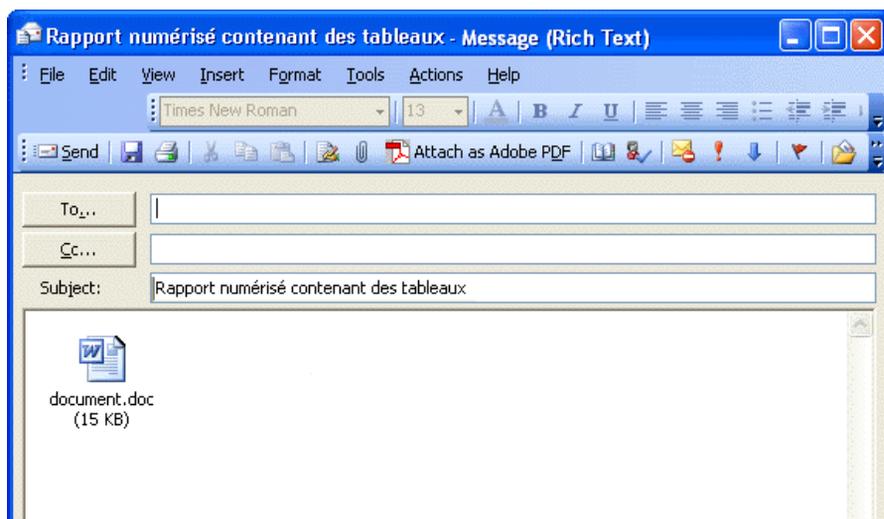


Cependant, la méthode employée pour envoyer vers une sortie cible est différente. Cette fois, le **type de fichier Windows** détermine quelle application sera lancée. C'est comme si vous double-cliquez sur le fichier dans l'Explorateur Windows... (Avec l'option "Envoyer à", Readiris envoie le fichier directement vers l'application cible.)



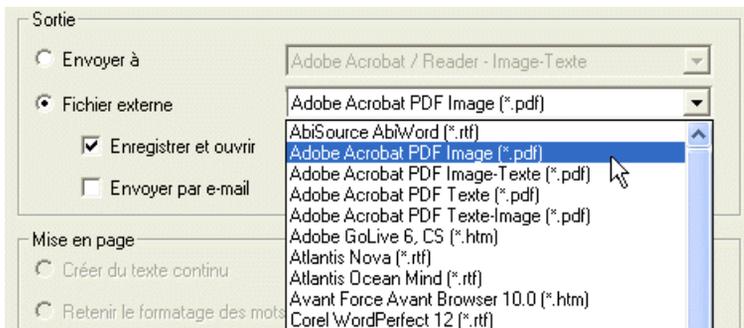
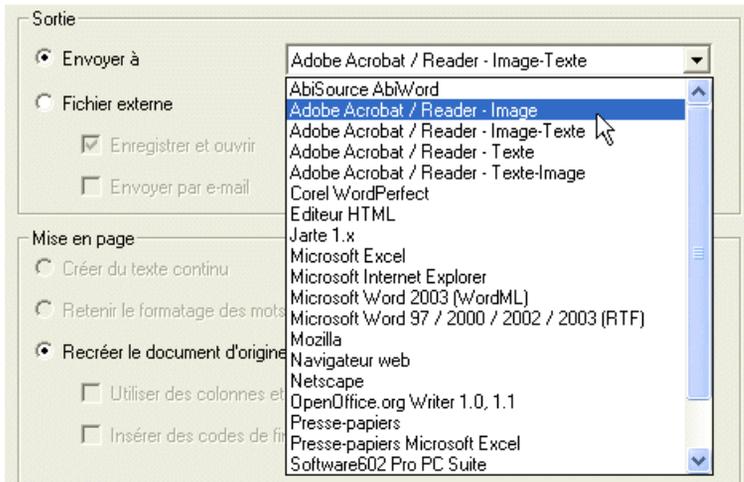
L'option "Envoyer par E-mail" crée un nouveau message **mail** et insère le document reconnu comme pièce jointe. Connaissez-vous une manière plus rapide pour envoyer un document papier...?





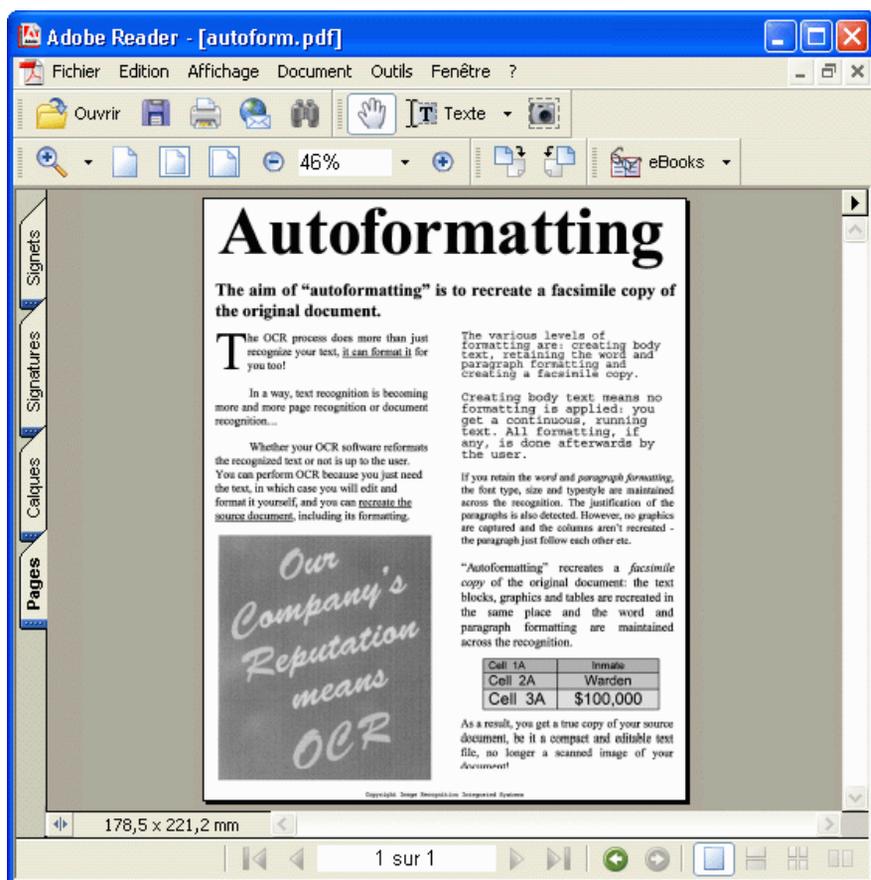
CRÉER DES DOCUMENTS PORTABLES...

Nous allons aborder plus en profondeur un des formats: **Adobe Acrobat PDF**. Readiris vous permet de générer des documents PDF basés sur le texte et sur l'image.



Quelle est la différence entre ces formats? Lorsque vous sélectionnez le format "PDF Texte", Readiris crée un fichier PDF contenant le résultat sous forme de texte. (Il peut y avoir des graphiques mais seulement lorsque des zones graphiques sont présentes sur la page - photographiques, illustrations, etc.) En réalité, l'image de la page *n* est *pas* contenue dans le fichier PDF. Ce dernier est constitué

d'une seule couche. Le format "PDF Image", quant à lui, ne contient lui non plus qu'une seule couche. Cette couche contient en fait l'image numérisée, et non pas le résultat de la reconnaissance.



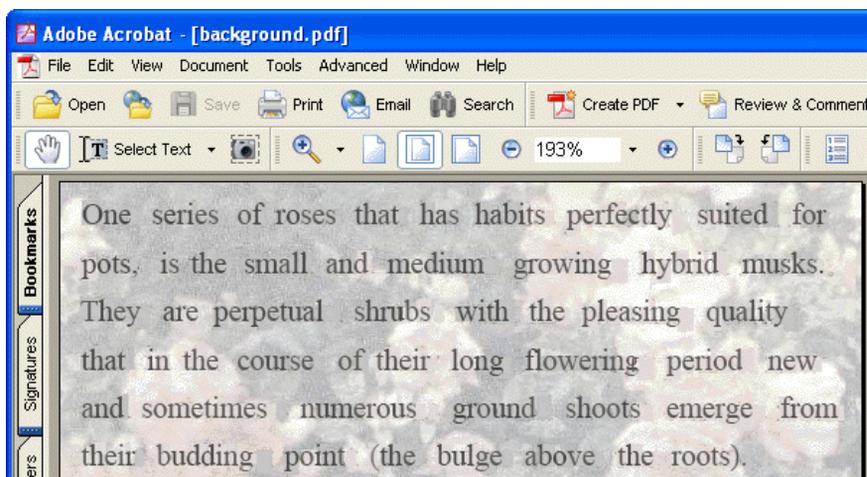
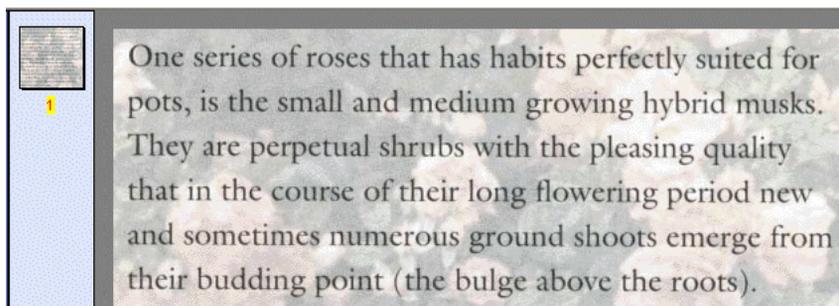


Les formats "PDF Texte-Image" et "PDF Image-Texte" génèrent des résultats différents: Readiris crée un fichier PDF pouvant être recherché et contenant le texte reconnu *et* l'image de la page. Dans le cas des fichiers de type "texte-image", le texte est placé *au-dessus* de l'image de la page dans un fichier PDF composé de deux couches; dans le cas des fichiers de type "image-texte", le texte est placé *en-dessous* de l'image de la page. Pour bien comprendre ces concepts, utilisez la commande "Rechercher" d'Adobe Reader.



Les fichiers PDF de type "texte-image" sont en fait assez sophistiqués: les pixels du texte reconnu sont effacés pour assurer la création d'un document lisible! Afficher le texte reconnu en noir sur des bitmaps de caractères noirs vous donnerait du texte avec une ombre lourde... Servez-vous de l'image d'exemple

BACKGROUND.JPG pour tester ce phénomène. (Readiris Corporate propose la même fonctionnalité pour d'autres formats...)



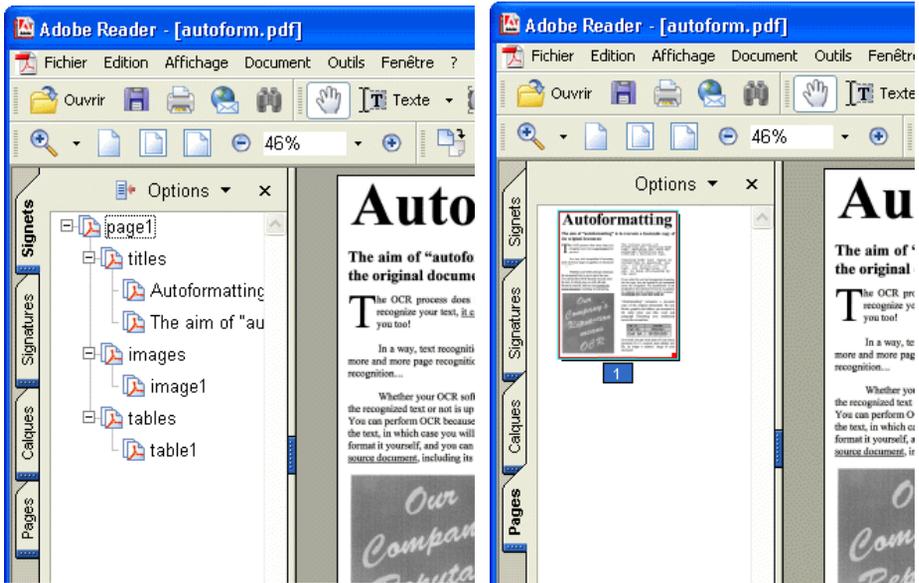
Tous les fichiers PDF basés sur le texte encodent les URLs de sites web comme des liens visibles: cliquez sur un lien et vous visitez directement le site web mentionné!



Cliquez sur le bouton "Format" pour découvrir quelques options concernant le format PDF Acrobat: "Créer les Signets" et "Inclure les Polices".

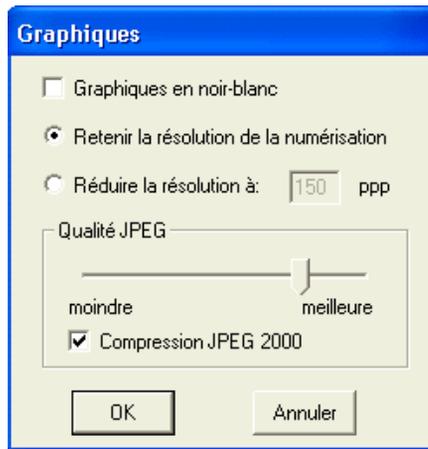


L'option "Créer les Signets" assure qu'un **signet** sera créé pour chaque élément du document - les graphiques tout comme les blocs de texte et les tables. Pour les zones de texte, Readiris applique un algorithme intelligent qui engendre un titre, un "sommaire" pour chaque zone; les tables et les graphiques sont simplement numérotés. (Un autre élément de navigation des documents PDF, les **vignettes** de la page, peut être réalisé dynamiquement par votre logiciel Adobe Reader!)



L'option "Inclure les Polices" inclut les fontes dans les fichiers PDF. L'inclusion des polices empêche la substitution des fontes lorsque les lecteurs affichent et impriment le document reconnu. Cette option s'assure que les lecteurs - quelque soit leur configuration d'ordinateur - visualisent le texte dans les polices originales. Cependant, l'inclusion des polices augmentent la taille du fichier des documents reconnus (légèrement)!

Une autre option du format PDF est proposée comme option avancée de la fenêtre "Format de Texte". Quand vous produisez des fichiers PDF, vous pouvez sélectionner la méthode de compression - JPEG ou JPEG 2000 - pour les graphiques sauvegardés à l'intérieur des fichiers PDF. (JPEG 2000 est la nouvelle version plus compacte du standard JPEG.)



... OU LES LIRE!

Découvrons une autre possibilité de Readiris... Alors que Readiris offre un support total du format Adobe Acrobat PDF, vous ne disposez pas seulement de la possibilité de générer des fichiers PDF, vous pouvez également les *lire*!

Pouvoir **donner un nouveau but aux documents PDF** est certainement une des applications principales de Readiris. Et ce pour plusieurs raisons. Tout d'abord car c'est un moyen de convertir des images en texte: ouvrez des documents PDF basés sur l'image, exécutez la reconnaissance et enregistrez le résultat OCR sous un document texte (dans n'importe quel format texte supporté). Or, les fichiers texte sont éditables, tandis que les fichiers image ne le sont pas.

Ensuite, vous pouvez convertir des fichiers PDF basés sur l'image en documents PDF basés sur du texte. Vous exécutez la reconnaissance sur des fichiers

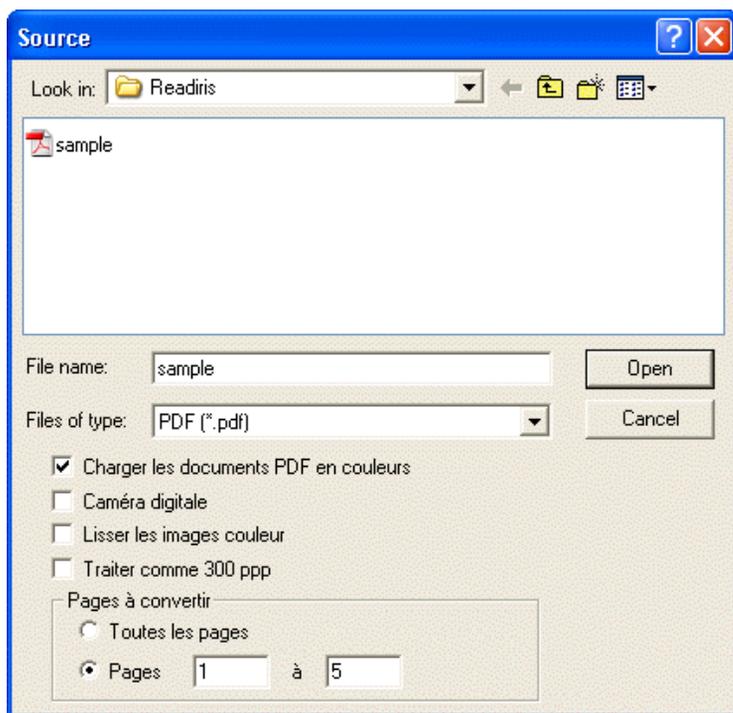
PDF “image uniquement” et sauvegardez les résultats OCR... comme des textes PDF! Vous pouvez rechercher et éditer des fichiers PDF basés sur du texte mais pas des fichiers PDF “image uniquement”.

Enfin, convertir des fichiers PDF est une manière de “déverrouiller” le contenu d’un fichier PDF. Vous pouvez reconnaître des documents PDF “en lecture seule”, où le texte est inaccessible en temps normal. Avec les fichiers PDF non protégés, le contenu peut être récupéré (copié et sauvé dans un fichier Texte). Avec les fichiers “en lecture seule”, le contenu ne peut pas être extrait. De tels fichiers peuvent seulement être affichés et imprimés!

Deux nuances importantes: même si Readiris franchit toutes les autres barrières de protection, il n’ouvre pas les documents PDF protégés par mot de passe! (Plus précisément, les mots de passe “maîtres” qui établissent les permissions d’un fichier PDF ne gênent pas Readiris, mais les mots de passe “utilisateur” dont vous avez besoin pour ouvrir un fichier PDF sont une entrave pour Readiris.) Par ailleurs, Readiris ne convertit pas les documents PDF qui contiennent des images compressées au format JPEG 2000.

Procédez comme d’habitude: chargez les fichiers PDF dans la mémoire comme vous ouvrez des images prénumérisées - des télécopies, des instantanés pris avec votre caméra numérique, etc. Cliquez sur le bouton "Arrêter" ou la touche Ech pour arrêter le processus de chargement entre deux pages.

Vous trouverez une option spécifique pour les fichiers PDF. Vous pouvez les ouvrir comme documents couleur et noir et blanc. Cette option est offerte car ouvrir des documents couleur est beaucoup plus lent!



En outre, vous pouvez indiquer quelles pages vous voulez convertir. Si vous souhaitez reprendre seulement un chapitre d'une longue publication au format PDF, rien ne vous oblige à charger le livre entier dans Readiris... Indiquez les **pages à convertir** et vous gagnerez beaucoup de temps! (Il en va de même pour les images TIFF multipages.)

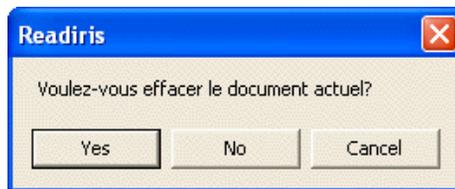
RECONNAISSANCE DE PAGES MULTIPLES

Après l'OCR, l'image numérisée sera ré-affichée avec les zones créées en vue d'être disponible pour un traitement ultérieur.

Vous pouvez maintenant ouvrir le texte reconnu avec votre traitement de texte ou votre éditeur de texte, l'importer dans le logiciel de publication (PAO) ou dans n'importe quelle application texte. Allez-y plus loin et comparez-le avec l'image située dans votre fenêtre de Readiris.

Mais comment sauvegarder le texte pour les pages supplémentaires? En d'autres mots: comment traiter un document de plusieurs pages? C'est en fait très simple: continuez à reconnaître des pages et enregistrez dans le même fichier! (Soyez sûr qu'il n'est pas déjà ouvert, car cela vous empêcherait de vous y enregistrer!) Ensuite, n'oubliez pas de mettre le dictionnaire de fontes en mode d'enrichissement pour vous permettre de poursuivre le traitement des fontes confortablement.

Dès que vous numérisez des pages (ou ouvrez des fichiers image) dans un document, vous devrez décider si vous désirez démarrer un nouveau document ou compléter le document actuel.



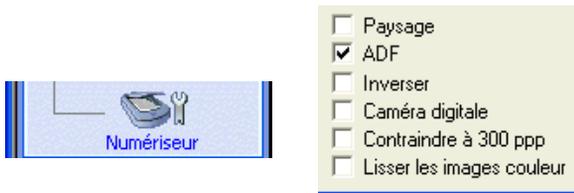
Répondez "non" pour ajouter des pages au document actuel, "oui" pour créer un nouveau document. Cette réponse a le même effet que la commande "Nouveau Document" dans le menu "Fichier".





Mais il existe un moyen plus efficace pour reconnaître d'affilée plusieurs pages que de les numériser et les OCRiser l'une après l'autre: il suffit de traiter directement des **documents multipages**!

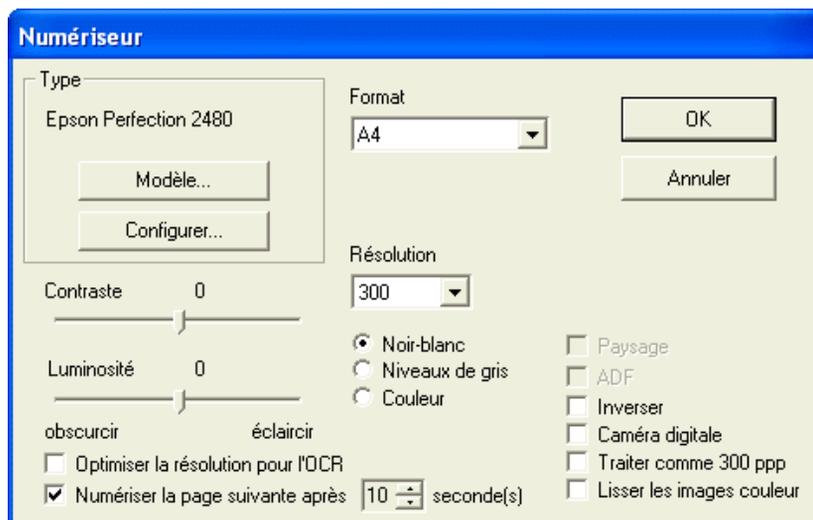
Pour numériser un document composé de plusieurs pages en une seule opération, activez l'alimentation automatique de votre scanner grâce à l'option "ADF" sous le bouton "Numériseur".



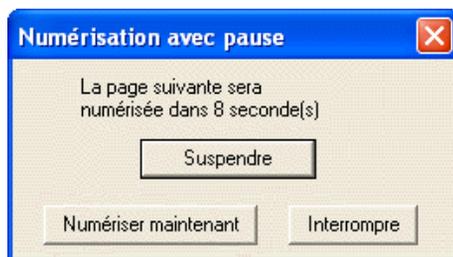
Placez les pages de votre document dans l'alimentation automatique et démarrez la numérisation: toutes les pages sont numérisées jusqu'à ce que l'alimentation se soit vidée.

Si votre numériseur ne dispose pas d'une alimentation automatique, vous pouvez utiliser la **numérisation avec pause** pour acquérir des documents multipages avec efficacité. Le numériseur scanne automatiquement une autre page après un délai déterminé par l'utilisateur. La pause vous laisse le temps de remplacer la page qui se trouve sur la fenêtre de votre numériseur à plat.

Indiquez dans les paramètres de numérisation quel temps de pause vous souhaitez déterminer pour pouvoir placer chaque nouvelle page dans votre numériseur: cliquez sur le bouton "Numériseur" et choisissez une valeur pour l'option "Numériser la Page Suivante après x Seconde(s)".



Cliquez sur le bouton "Numériser" pour démarrer l'acquisition des images. Cliquez sur "Interrompre" dans la fenêtre de numérisation avec pause si vous voulez mettre fin à la numérisation automatique.

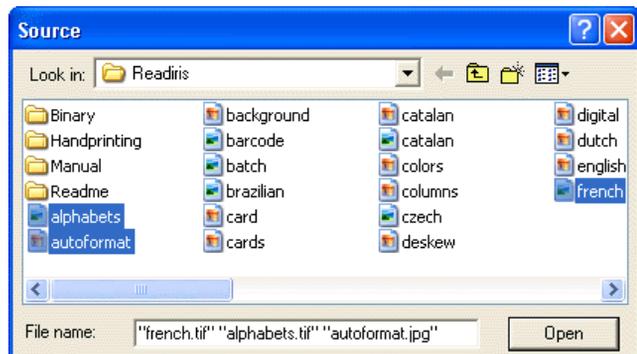


Readiris est très flexible: vous pouvez réduire l'intervalle de temps afin d'acquérir une image plus rapidement, ou encore "geler" la pause si le téléphone vous interrompt! Cliquez sur le bouton "Suspendre" dans la fenêtre de numérisation

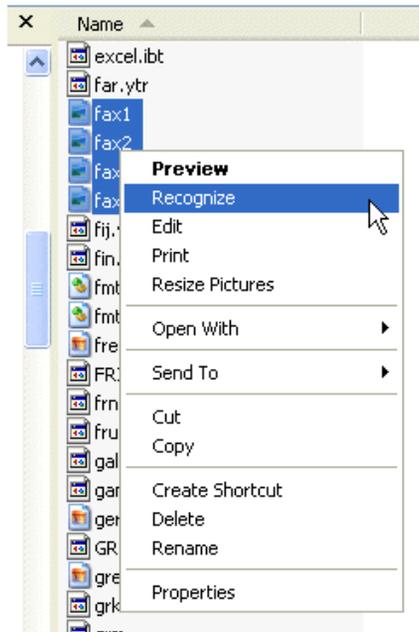


avec pause pour geler l'intervalle de numérisation; cliquez sur "Poursuivre" quand vous êtes prêt à continuer. Cliquez sur "Numériser Maintenant" dans cette fenêtre pour réduire le temps de la pause.

Vous pouvez également *ouvrir* plusieurs images prénumérisées. Pour charger plusieurs images en mémoire, sélectionnez la première image et maintenez la touche Ctrl enfoncée pendant que vous sélectionnez les images supplémentaires. Pour charger une rangée continue d'images, sélectionnez la première image et maintenez la touche Maj enfoncée tout en sélectionnant la dernière image.



Vous pouvez obtenir aisément le même résultat depuis l'Explorateur Windows: sélectionnez plusieurs fichiers image, cliquez dessus avec le bouton droit de la souris et sélectionnez la commande "Reconnaître" dans le menu "Contexte". Vous pouvez répéter cette opération: toutes les images envoyées vers Readiris enrichissent le document actuel jusqu'à ce que vous cliquiez sur la commande "Nouveau Document".



Vous pouvez même *glisser-déposer* plusieurs images préscannées depuis l'Explorateur Windows vers la fenêtre de Readiris! Le même principe est applicable: toutes les images que vous faites glisser vers la fenêtre de Readiris sont ajoutées au document actuel jusqu'à ce que vous cliquiez sur la commande "Nouveau Document".

Readiris trie automatiquement les images - l'image 001.tif précède l'image 002.tif qui précède 003.tif, etc.

Le **panneau des documents** affiche des **données statistiques** et donne des informations sur toute page acquise - la source de l'image et sa résolution, la durée de la numérisation et de la lecture, etc. Vous pouvez cacher ou afficher le panneau des documents avec l'option correspondante du menu "Vue".



✓ Panneau des documents

Le panneau des documents vous indique également quelle est l'image qui est chargée en mémoire. Attention: si une image multipage a été ouverte, il n'y a évidemment qu'un seul fichier pour toutes les images! Pendant la numérisation, le panneau des documents ne mentionne que le modèle du numériseur.

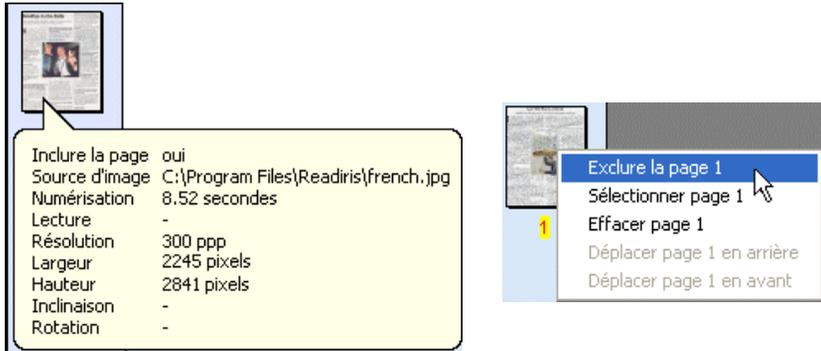
Page	Source d'image	Numérisation	Lecture	Résolution	Largeur	Hauteur	Inclinaison	Rotation
<input checked="" type="checkbox"/> 1	C:\Program Files\Readiris...	1.03	-	300	2000	2388	-	-
<input checked="" type="checkbox"/> 2	C:\Program Files\Readiris...	0.93	-	300	2000	1888	-	-
<input checked="" type="checkbox"/> 3	C:\Program Files\Readiris...	0.80	-	300	1912	2004	-	-

Vous pouvez modifier la taille des colonnes à l'aide de votre curseur. (Vous ne pouvez pas modifier l'ordre des colonnes.) Si par manque de place, une colonne n'affiche pas des données complètes, survolez la ligne d'information à l'aide de votre curseur pour obtenir une information complète.

Source d'image	Numérisation
C:\Program Files\...	2.81

Page	Source d'image	Numérisation
<input checked="" type="checkbox"/> 1	C:\Program Files\Readiris\	french.tif

Si vous sélectionnez la commande "Informations" du menu "Fichier", Readiris affichera le même type d'information pour toutes les pages. Vous pouvez aussi afficher ces mêmes informations pour chaque page individuellement en maintenant votre curseur sur la vignette de la page voulue dans la **barre d'outils des pages**. Cette barre d'outils s'affiche automatiquement dans la partie gauche de l'écran dès l'instant où des pages sont en cours de traitement. Un simple clic droit sur la vignette d'une page permet par ailleurs d'accéder aux actions à réaliser sur cette page.



La page active est mise en évidence dans le panneau des documents et la barre d'outils des pages et est mentionnée dans la barre de titre de Readiris.

Pour effacer une page, sélectionnez-la dans le panneau des documents et appuyez sur la touche Del! (Ou sélectionnez la page dans la barre d'outils des pages, cliquez sur le bouton droit de la souris et sélectionnez la commande "Effacer Page" du menu "Contexte".)

Si vous souhaitez obtenir un aperçu de votre document, vous pouvez rapidement **imprimer** les **images** scannées via le bouton "Imprimer" de la barre d'outils image (ou avec la commande "Imprimer les Images" sous le menu "Fichier").

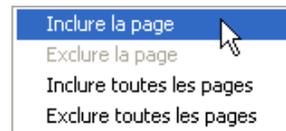


Vous pouvez imprimer la page actuelle ou toutes les pages. Sélectionnez le nombre de pages ou de vignettes que vous désirez imprimer sur une page.

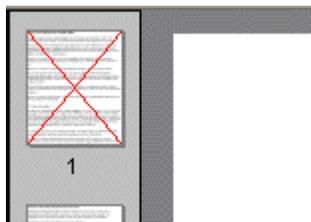


Mais vous ne devez pas nécessairement imprimer toutes les pages: le panneau des documents (ainsi que les commandes correspondantes du menu "Edition" et les commandes contextuelles de la barre d'outils des pages) permet d'exclure (temporairement) des pages du processus d'impression. Pour exclure une page du processus d'impression (et de reconnaissance), cliquez sur le numéro de cette page dans le panneau des documents. Pour l'inclure à nouveau, cliquez une deuxième fois. Pour appliquer ces commandes directement sur *toutes* les pages, utilisez les commandes du menu "Edition".

Page	Source d'image	Numérisation
<input type="checkbox"/>	1	C:\Program Files\Readiris... 0.94
<input checked="" type="checkbox"/>	2	C:\Program Files\Readiris... 0.71
<input checked="" type="checkbox"/>	3	C:\Program Files\Readiris... 0.73



Les vignettes des pages exclues sont rayées. Rappelez vous que l'impression de la page actuelle fonctionne toujours, même si celle-ci est désactivée en ce moment!



Chargez l'image d'exemple MULTIPAGE.TIF et démarrez la reconnaissance. Les différentes pages s'affichent les unes après les autres; la barre de titre de Readiris indique le numéro de la page.



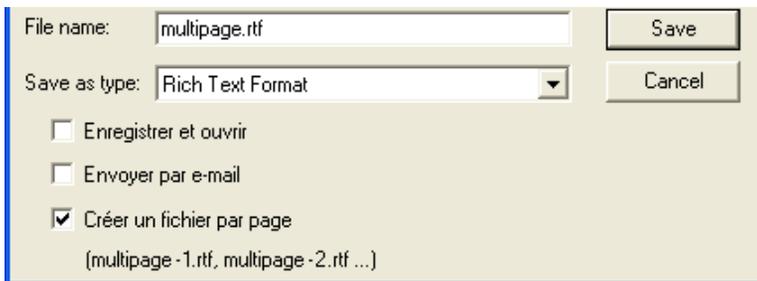
Page	Source d'image	Numérisation	Lecture	Résolution	Largeur	Hauteur	Inclinaison	
<input checked="" type="checkbox"/>	2	C:\Program Files\...	0.96	0.21	300	2000	1888	-
<input checked="" type="checkbox"/>	3	C:\Program Files\...	0.83	0.21	300	1912	2004	-
<input checked="" type="checkbox"/>	4	C:\Program Files\...	0.85	1.35	300	2004	2000	-

Si l'apprentissage interactif est activé, vous exécuterez les phases de reconnaissance et d'apprentissage page par page. Le mode dictionnaire "Nouveau" s'appliquera à la première page et le mode "Enrichir" aux pages suivantes.

Quand vous cliquez sur le bouton "Finir", toutes les décisions seront acceptées par le système sans validation de l'utilisateur. En d'autres termes, l'apprentissage interactif est interrompu pour *toutes* les pages; l'OCR se poursuit en mode automatique pour ce document.

Le résultat de la reconnaissance de documents multipages est enregistré dans un seul fichier de sortie. Lorsque le résultat de la reconnaissance est envoyé dans une application cible, des pages multiples sont créées à l'intérieur d'un seul document.

C'est en tout cas ce qui l'on peut réaliser si l'option "Créer un Fichier par Page" est désactivée au moment de sauvegarder le document reconnu. Cette option permet de faire en sorte que chaque page d'un document multipage soit sauvegardée dans un fichier séparé. Si l'utilisateur donne le nom de fichier texte.doc, les fichiers seront nommés texte-1.doc, texte-2.doc, etc. (Cette option n'est pas disponible quand vous envoyez les résultats de l'OCR vers une application cible, elle n'est utilisable que lors de la création de fichiers externes.)



ÉDITION DE DOCUMENTS MULTIPAGES

L'utilisateur peut modifier des documents multipages, principalement afin de corriger des erreurs dues à la numérisation: il peut ainsi effacer des pages du document et déplacer des pages à d'autres endroits du document.

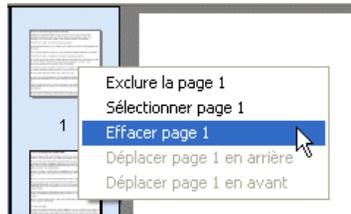
Voyons d'abord le déplacement. Pour *accéder à une page*, cliquez sur la page dans le panneau des documents ou sur sa vignette sur la barre d'outils des pages. (Ou placez le curseur sur sa vignette, cliquez dessus avec le bouton droit de la souris pour faire apparaître le menu "Contexte" et utilisez la commande "Sélectionner Page".) Pour aller à la page précédente, vous pouvez utiliser le



raccourci PageUp et pour aller à la page suivante, appuyez sur PageDn. Appuyer sur Home pour aller à la première page et sur End pour vous rendre à la dernière page. Alternativement, vous pouvez utiliser les commandes correspondantes du menu "Vue".

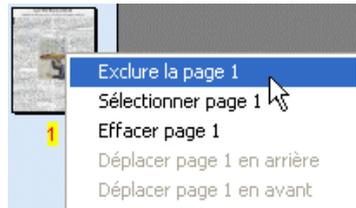
Première page	Home
Page précédente	PageUp
Page suivante	PageDown
Dernière page	End

Abordons à présent la question de l'édition des documents. Pour *supprimer une page* d'un document, placez le curseur sur son imagette, opérez un clic droit et sélectionnez la commande "Effacer Page" du menu "Contexte".



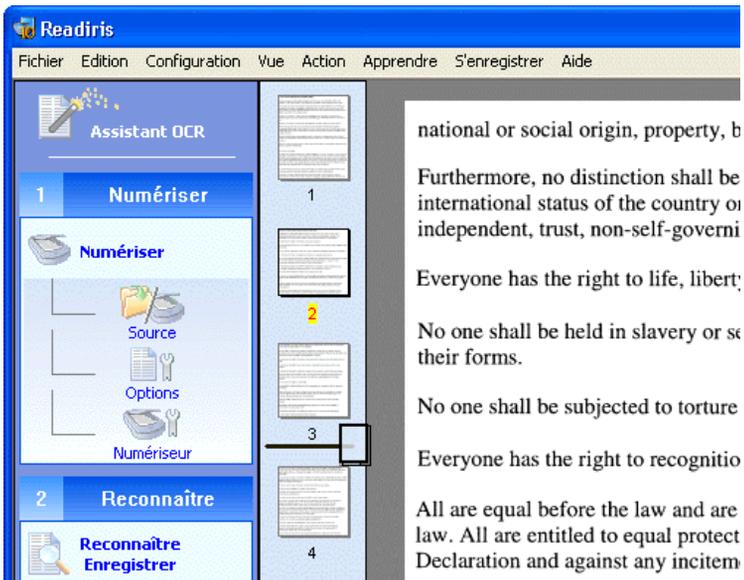
Nous vous rappelons que vous pouvez temporairement exclure les pages, sans les supprimer, du processus de reconnaissance (et de l'impression des images): le panneau des documents, la barre d'outils de la page (et le menu "Edition") offrent les commandes nécessaires.

Page	Source d'image	Numérisation
<input type="checkbox"/> 1	C:\Program Files\...	1.09
<input checked="" type="checkbox"/> 2	C:\Program Files\...	1.21



Pour *déplacer* une page *vers l'avant* du document, utilisez la commande "Déplacer Page en Avant" et pour *déplacer* une page *vers l'arrière*, utilisez la commande "Déplacer Page en Arrière".

Pour *déplacer une page* vers une location différente dans le document, glissez l'icône vers cette nouvelle localisation.



COMMENCER AVEC UN NOUVEAU DOCUMENT

Vous pouvez utiliser la commande "Nouveau Document" sous le menu "Fichier" pour fermer le document actuel.

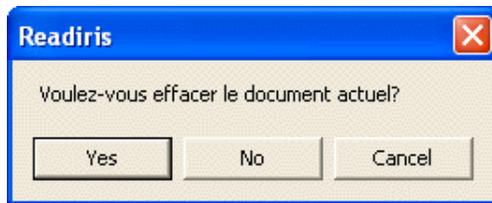




Cette commande "efface l'ardoise". Tout document chargé en mémoire - contenant une seule ou plusieurs pages - sera effacé. Vous pouvez donc créer un nouveau document dès à présent.

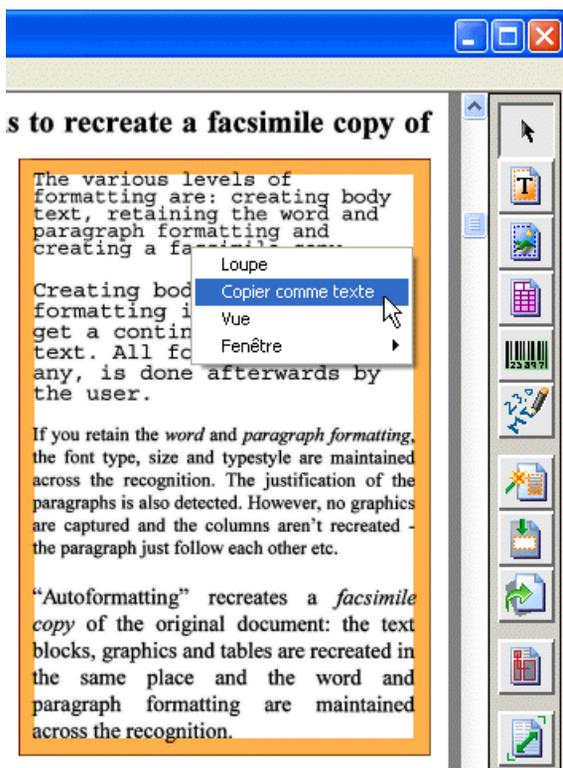
Mais vous pouvez aussi créer un nouveau document depuis le document actif. Tant que l'OCR n'a pas été exécuté, le système suppose que vous voulez ajouter des pages au document actuel. Vous pouvez par exemple numériser toutes les pages depuis l'alimentation automatique du scanner, remplir à nouveau l'alimentation et recommencer. Toutes les pages numérisées composeront un seul et même document. Ou alors vous pouvez numériser un certain nombre de pages et y ajouter des fichiers image, par exemple des télécopies. Ces pages forment encore un document unique, tout ce que vous devez faire c'est changer entretemps l'image source au moyen du bouton "Source".

Lorsque l'OCR *était* déjà exécuté auparavant et que vous recommencez la numérisation (ou le chargement d'images), vous êtes invité à démarrer un nouveau document ou compléter le document actuel.



RECONNAISSANCE DES ZONES DE TEXTE

Nous savons maintenant comment reconnaître des pages et comment traiter des documents multipages. Mais est-ce possible de reconnaître plus d'une page avec un confort identique? C'est possible! Cliquez sur le bouton droit de la souris et sélectionnez la commande "Copier comme Texte" du menu "Contexte": la fenêtre de texte sous la souris est reconnue et envoyée vers le presse-papiers.



La configuration du système actuelle- langue, type de police, etc. - s'applique. Le résultat de la reconnaissance est placé dans le presse-papiers comme texte "continu", non-formaté.



ORGANISATION DU TEXTE SAISI

Sauvegarder ou exporter le texte signifie plus que sélectionner un mode de sortie ou donner un nom de fichier au texte reconnu. Vous sélectionnez également le format du fichier texte et déterminez l'apparence du texte reconnu. En bref, vous devez décider ce que vous allez faire du texte avant de lancer l'exécution.

Certaines options du bouton "Format" vous permettent de changer l'apparence du texte saisi.

Le **flux du texte** du document saisi est directement influencé par l'option "Reconstituer les Paragraphes".



Laissez cette option activée afin que Readiris détecte les paragraphes: Readiris appliquera alors la **frappe en continu** comme dans les traitements de texte, sinon, un retour à la ligne est ajouté à la fin de chaque ligne et les mots avec un trait d'union restent tels quels. La détection des paragraphes est activée par défaut.

Voici un exemple: quand les trois premières lignes d'une colonne sont "Le nouveau prési-", "dent faisait signe du balcon." et "Sa femme l'a rejoint.", la détection de paragraphe vous donne le résultat suivant "Le nouveau **président** faisait signe du balcon. **Sa** femme l'a rejoint." Les deux parties du mot "président" ont été "recollées" et un espace est ajouté à la fin de la première phrase, créant ainsi du texte qui "coule" naturellement.

Si la détection de paragraphe *n'a pas* été activée, la mise en page originale sera conservée, avec un retour à la ligne ajouté à la fin de chaque ligne.

Cette option *n'est pas* disponible lorsque le format PDF est sélectionné: les fichiers PDF Adobe Acrobat sauvegardent toujours le texte ligne par ligne.

(Le bouton "Format" dispose de quelques options de formatage dont nous parlerons plus loin.)

CONFIGURATION DU NUMÉRISEUR

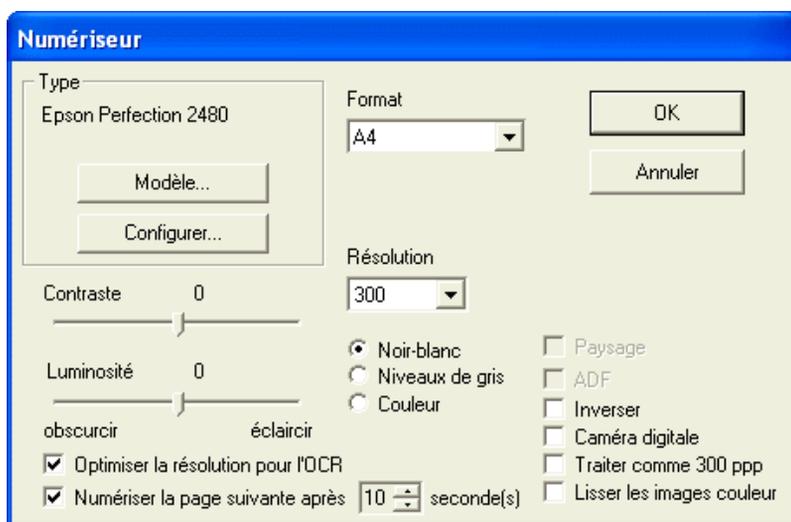
Configurons maintenant notre numériseur sous Readiris. Il est entendu que le numériseur et ses pilotes sont correctement installés.

Si votre licence Readiris faisait partie des logiciels fournis avec le scanner, il se peut que cette étape ne soit pas nécessaire, dans la mesure où le numériseur peut déjà se trouver configuré sous Readiris.

Cliquez sur le bouton "Numériseur" de la barre d'outils principale.



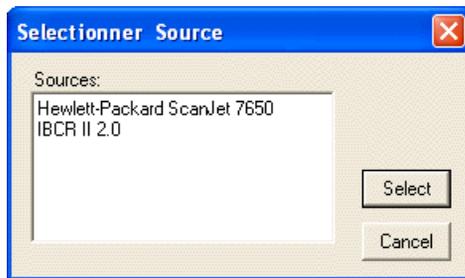
Cliquez sur le bouton "Modèle" pour déterminer le **modèle** de votre numériseur.





Quand vous sélectionnez l'option "<Image>" comme "numériseur", les images préscannées fonctionneront dans toutes les circonstances comme source d'image - sans même devoir sélectionner le disque comme source d'image en utilisant le bouton "Source" sur la barre d'outils principale.

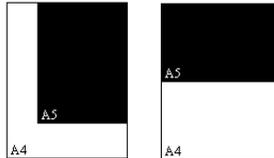
Le bouton "Configurer" n'est opérationnel que si votre numériseur l'autorise. Il donne accès à certains paramètres pointus de numérisation; avec les scanners Twain, cliquer sur le bouton "Configurer" vous permet de choisir la source Twain. (Vous pouvez également utiliser la commande "Sélectionner Source" sous le menu "Fichier".)



Une fois le scanner sélectionné, la même fenêtre peut vous permettre d'établir la résolution de numérisation, le format de page, l'orientation, la luminosité et le contraste, et peut vous permettre d'indiquer si vous allez utiliser l'alimentation automatique du numériseur. Avec les numériseurs compatibles Twain, tous les paramètres de numérisation sont souvent définis dans l'interface Twain.

Définissez la **brillance** et, si disponible, le **contraste**.

En activant l'option "Paysage", vous indiquerez que le texte est orienté en largeur (documents "paysage") plutôt qu'en hauteur (documents "portraits"). L'orientation de page s'applique plutôt à des pages de format réduit: sur un scanner à plat A4, vous pouvez numériser des pages A5 (à moitié moins grandes) en mode paysage ou portrait, mais vous ne pouvez numériser toute la surface d'un A4 que dans un sens!



Enfin, l'option "Inverser" vous permet de générer des **images "inversées"** pour le mode de numérisation en noir et blanc - vous pouvez activer cette option pour les pages dont tout le texte est écrit en blanc sur fond noir.

Nous avons déjà documenté la numérisation avec pause. Il s'agit d'une façon très efficace d'acquérir des documents multipages dans le cas où vous disposez d'un numériseur à plat non équipé d'une alimentation automatique.

METTEZ UN PEU DE COULEUR DANS VOS TEXTES!

Readiris supporte les images en noir et blanc, niveaux de gris et couleur de la même manière, vous est donc libre de choisir le **mode couleur** le mieux adapté à vos besoins. Pour inclure des graphiques en mode trait dans les documents reconnus, numérisez-les en noir et blanc. Pour inclure des photos en noir et blanc, effectuez une numérisation en niveaux de gris et pour inclure des photos en couleur, numérisez en couleur.

Mais pourquoi réduire les couleurs des images lors de la numérisation? Cela va sans dire que les images en niveaux de gris et couleur seront plus lentes à acquérir et exigent une mémoire vive plus importante que les images en noir et blanc.

Numériser en niveaux de gris et en couleur n'est pas seulement utile pour sauvegarder les graphiques avec une qualité suffisante. Dans certains cas ce sera aussi intéressant ou nécessaire pour obtenir de bons résultats d'OCR! Lorsqu'un texte est imprimé sur un fond coloré, la numérisation en couleur peut donner des variations de tons qui manqueraient dans des images en noir et blanc. Lorsqu'il n'y a qu'un contraste limité entre le texte et l'arrière-plan, celui-ci peut



occasionner du “bruit” qui rendra plus difficile, voire impossible, la reconnaissance!

Pensez par exemple à du texte noir imprimé sur un fond sombre: si vous numérisez un tel document en noir et blanc, vous ne pourrez vous “débarrasser” de l’arrière-plan sans perdre l’information textuelle, quoique vous fassiez pour ajuster la luminosité du numériseur...

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.

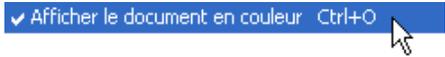
YASUMITSU SHIGETA, 35, has invested in more than 70 Web or mobile Net-based ventures in Japan and the U.S., including Tumbleweed Communications and Phone.com. Shigeta is also developing new businesses that take advantage of the growth of the Internet and mobile communications.

Readiris crée une version noir et blanc pour chaque image en niveaux de gris et couleur. Grâce à ses routines intelligentes, même des documents récalcitrants sont résolus - voici comment une image “difficile” est binarisée!

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.

Pour visualiser une image numérisée en noir et blanc, activez l'option "Afficher le Document en Couleur" sous le menu "Vue".



✓ Afficher le document en couleur Ctrl+O

En fait, vous n'afficherez pas d'images en noir et blanc sur l'écran de votre ordinateur - même si vous numérisez des images bitonales! Readiris optimise les images pour obtenir un maximum de lisibilité à l'écran: la technique d'affichage spécialisée en haute résolution d'I.R.I.S. convertit les images noir et blanc en niveaux de gris.

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

Les images en niveaux de gris et en couleur sont lissées.

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.

Par conséquent, il n'est pas nécessaire d'agrandir les images, même sur des ordinateurs portatifs avec écran à cristaux liquides ou sur des ordinateurs de bureau avec un écran de 15" de basse qualité.

Affichez une image à sa taille réelle (ou zoomez davantage) si vous voulez visualiser l'image “pure”, telle qu'elle a été numérisée.

À DIFFÉRENTS APPAREILS, DIFFÉRENTES RÉOLUTIONS

Quel que soit votre mode de la numérisation, utilisez une **résolution** de 300 ppp pour des applications normales. Utilisez la résolution supérieure de 400 ppp

pour de petits caractères (inférieurs à 10 points) et quand le document est vraiment détérioré.

Readiris lit les caractères dont la **taille** va de 6 à 72 points (0,21 à 2,54 cm.).

6 points

72 points

Readiris reconnaît aussi les "**lettrines**", c'est-à-dire les caractères larges qui couvrent plusieurs lignes. (Attention, ces lettrines ne peuvent pas dépasser 72 points!) Même les lettrines inversées sont reconnues...

Readiris lit les "lettrines" (qui recouvrent plusieurs lignes) et les attribue à la première ligne du paragraphe.

Le Festival de Wallonie est dans sa phase ultime et finale avec l'ouverture de ses dernières sections.

Cette semaine, c'est la branche

Étant donné qu'il faut une résolution comprise entre 300 et 400 ppp pour assurer une reconnaissance optimale, Readiris vous avertit lorsque vous lui soumettez des images dont la résolution est inférieure à 200 ppp ou supérieure à 800 ppp. Remarquez aussi que la résolution de telles images apparaît en rouge dans le panneau des documents.

Page	Source d'image	Numérisation	Lecture	Résolution	Largeur	Hauteur
<input checked="" type="checkbox"/> 1	C:\Documents and..	1.24	-	96	400	447

Readiris peut corriger pour vous les documents numérisés avec beaucoup de détails! Pour ce faire, activez l'option "Optimiser la Résolution pour l'OCR" dans la configuration du numériseur. A chaque fois que la résolution de l'image de vos numérisations dépasse 600 ppp, la résolution est réduite pour le traitement OCR.



<input checked="" type="checkbox"/> Optimiser la résolution pour l'OCR
<input type="checkbox"/> Numériser la page suivante après <input type="text" value="10"/> seconde(s)

Il existe d'autres manières d'éviter cet avertissement: en lisant des **télécopies** - qui ont une résolution de 100 ou 200 ppp -, lorsque vous générez des images avec un appareil-photo numérique - alors la résolution est inconnue - et lorsque vous ouvrez des images où l'en-tête du fichier contient une mauvaise résolution. Pour traiter de telles images sans problème, activez l'option "Traiter comme 300 ppp". Ce paramètre s'applique à la numérisation directe et l'ouverture d'images prénumérisées.

<input type="checkbox"/> Inverser
<input type="checkbox"/> Caméra digitale
<input checked="" type="checkbox"/> Contraindre à 300 ppp
<input type="checkbox"/> Lisser les images couleur

<input type="checkbox"/> Charger les documents PDF en couleurs
<input type="checkbox"/> Caméra digitale
<input type="checkbox"/> Lisser les images couleur
<input checked="" type="checkbox"/> Traiter comme 300 ppp

Lorsque vous saisissez des images avec une **caméra digitale** au lieu d'un numériseur, il est impératif d'activer une option spéciale. Ce paramètre s'applique encore une fois à la numérisation directe et aux images prénumérisées.

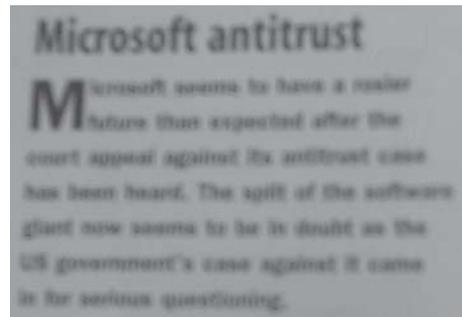
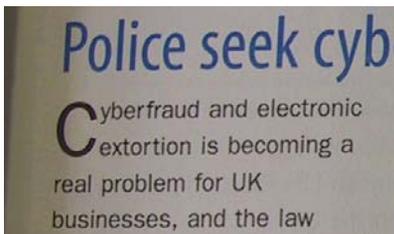
<input type="checkbox"/> Inverser
<input checked="" type="checkbox"/> Caméra digitale
<input checked="" type="checkbox"/> Contraindre à 300 ppp
<input type="checkbox"/> Lisser les images couleur

<input type="checkbox"/> Charger les documents PDF en couleurs
<input checked="" type="checkbox"/> Caméra digitale
<input type="checkbox"/> Lisser les images couleur
<input checked="" type="checkbox"/> Traiter comme 300 ppp

Ce faisant, vous améliorez l'image avant d'entamer le processus de reconnaissance. Il existe plusieurs difficultés propres aux appareils-photos numériques: ils engendrent des images de basse résolution - même lorsque vous maintenez l'appareil très proche du document - et la résolution de l'image est dans tous les cas inconnue.

Voici des "subtilités" dont il faut être conscient pour pouvoir reconnaître correctement les images saisies avec un appareil-photo numérique.

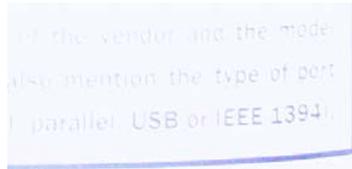
Tout d'abord, sélectionnez la résolution la plus haute possible pour l'image. Créez par exemple des images d'une taille de 2.048 x 1.536 lorsque des images de 1.024 x 768 ou 680 x 480 sont également supportées. En second lieu, activez le mode "macro" de votre caméra pour prendre des gros plans - ce qui est toujours le cas lorsque vous photographiez des documents. (Ce mode a été mis au point pour saisir des fleurs, des insectes, etc.) Sans quoi les images seront floues et illisibles.



Limitez-vous à de petites, voire même à aucune, compressions: une compression trop importante réduirait la netteté du texte saisi. Effectuez un zoom manuel pour encadrer votre document - certaines caméras sont livrées avec un logiciel "d'assemblage" ("photo stitching") de photographies mais ne l'utilisez pas pour saisir un document.

Tenez la caméra directement au-dessus du document pour éviter de saisir le document sous un angle. Essayez cependant d'éviter que l'ombre de l'appareil ou de votre main ne soit projetée sur le document! Essayez de produire des images stables. Pensez à fixer votre caméra sur un pied si cela s'avère nécessaire.

Désactivez le flash lorsque vous photographiez du papier brillant, pour éviter que l'image ne soit trop pâle. De manière générale, mieux vaut adapter la brillance et le contraste à l'environnement - lumière du jour, lumière électrique, néon, etc. (Il y a moyen de calibrer certaines caméras en filmant un document blanc.)

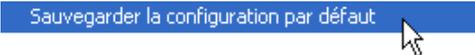


Faites un essai: ouvrez l'image DIGITAL.JPG dans le dossier Readiris et exécutez la reconnaissance.

Page	Source d'image	Numérisation	Lecture	Résolution	Largeur	Hauteur	Inclinaison	Rotation
<input checked="" type="checkbox"/>	1	C:\Program Files...	1.26	-	300	1519	635	-

SAUVEGARDER LES PARAMÈTRES PAR DÉFAUT

Définissez correctement tous les paramètres par défaut et cliquez sur la commande "Sauvegarder la Configuration par Défaut" dans le menu "Fichier" pour une prochaine utilisation.

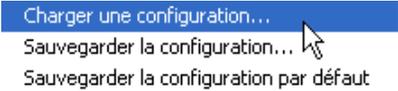


Sauvegarder la configuration par défaut

Les fichiers de paramètres contiennent plus que la **configuration** du numériseur: ils déterminent également si vous allez faire de l'apprentissage interactif, quelle langue est employée dans les documents, quel mode de sortie est utilisé - par exemple du texte envoyé dans WordPad -, etc. En bref, *tous* les paramètres opérationnels de Readiris sont stockés dans les fichiers de paramètres.

SAUVEGARDE DE PARAMÈTRES SPÉCIFIQUES

Les paramètres par défaut seront activés à chaque démarrage, mais vous pouvez sauvegarder des paramètres spécifiques afin de n'avoir pas à redéfinir les paramètres opérationnels. Les commandes "Sauvegarder la Configuration" et "Charger une Configuration" sous le menu "Fichier" s'en chargent.



Charger une configuration...

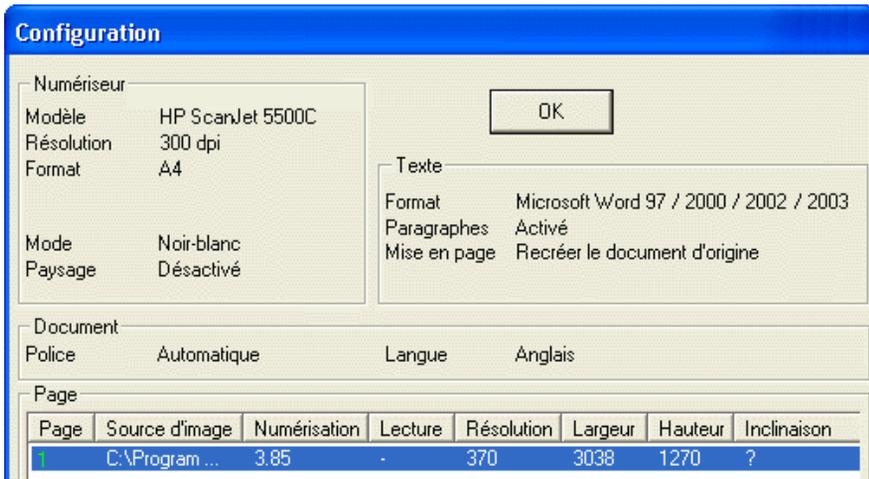
Sauvegarder la configuration...

Sauvegarder la configuration par défaut

Voici un exemple: si vous devez régulièrement faire de l'OCR sur des documents en français avec une mise en page spécifique, nous vous recommandons de créer un fichier de paramètres spécifique pour ce type de documents. Vous choisirez alors "Français" comme langue du document, chargerez un gabarit de fenêtres pour n'avoir pas à réappliquer le même fenêtrage chaque fois, désactiver l'apprentissage mais activer un dictionnaire de fontes en mode lecture étant donné que des polices identiques sont utilisées à chaque fois, etc.



Si vous n'êtes pas certain des paramètres en cours, vous ne devez pas "plonger" dans tous les menus et commandes. Utilisez la commande "Informations" du menu "Fichier" pour savoir ce qu'il en est.



De plus, cette commande affiche les informations que vous trouvez également sur le panneau des documents pour toutes les pages.

NUMÉRISER DES DOCUMENTS

Maintenant que notre scanner est configuré, nous pouvons débiter la numérisation de documents. Voici quelques éléments bons à savoir.

Tout d'abord, faites attention à la ligne de courbe. Bien que l'analyse de page et la reconnaissance acceptent des inclinaisons, il peut s'avérer difficile de fenêtrer et reconnaître une page correctement si la courbe est trop importante. Une inclinaison limitée (moins de 0,5°) sera ignorée car la précision de l'OCR n'en sera pas affectée.

L'option "Redressement des Pages" sous le menu "Options" (et sous le menu "Configuration") détermine si les pages qui ont été numérisées sous un angle seront **redressées**, rendues droites automatiquement. Une courbe limitée sera ignorée. Cette option est désactivée par défaut.

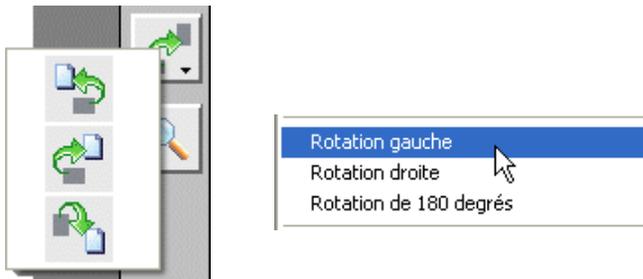


Si vous avez oublié d'activer cette option, utilisez le bouton "Redresser la Page" de la barre d'outils de l'image (et la commande "Redresser la Page" sous le menu "Action") pour rendre droites les pages qui ont été numérisées sous un angle.



Le redressement prendra quelques secondes: l'image est analysée afin de détecter l'angle de l'inclinaison - s'il y en a -, l'image couleur ou en niveaux de gris *et* sa version noir et blanc sont redressées et l'analyse de page est ré-exécutée.

Vous pouvez aussi avoir besoin d'ajuster l'orientation de la page. Utilisez les outils de **rotation** sur la barre d'outils de l'image. (Les commandes correspondantes se trouvent dans le menu "Vue".) Trois directions de rotation sont proposées: à gauche, à droite et 180°. La rotation prend aussi quelques secondes étant donné que l'image elle-même est activée et non simplement son affichage à l'écran.

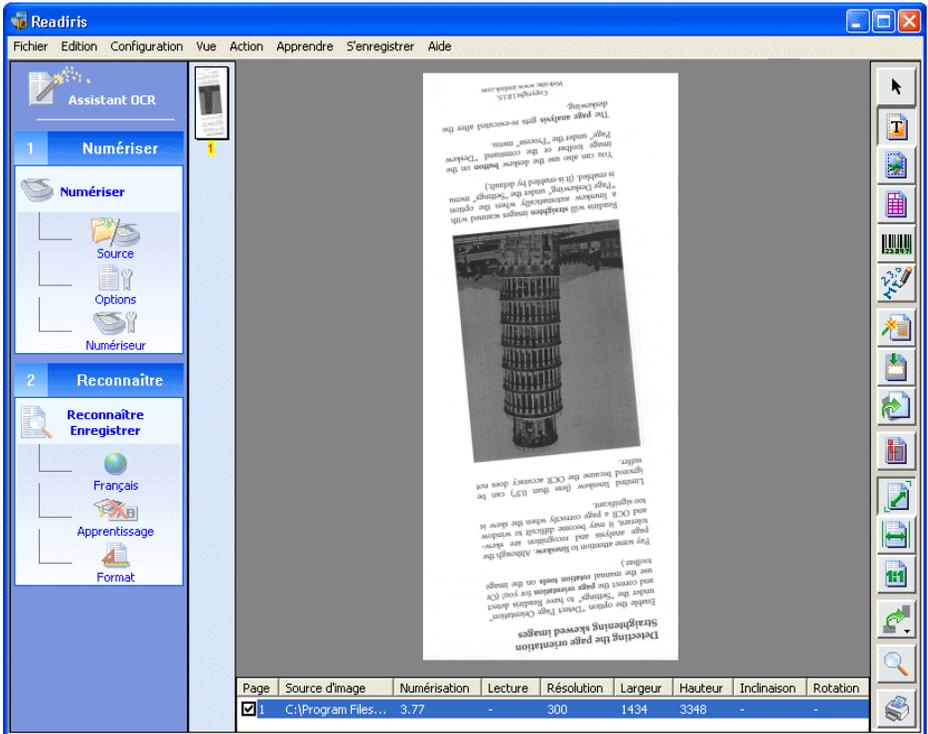




Cependant, Readiris peut corriger pour vous les pages mal orientées. Activez l'option "Détecter l'Orientation de la Page" sous le menu "Options" (ou sous le menu "Configuration") et Readiris corrigera l'orientation de la page si nécessaire.



Il y a moyen de faire bon usage de l'image DESKEW.JPG dans le dossier Readiris si vous voulez essayer cette fonction. Activez l'option "Redressement des Pages" et "Détecter l'Orientation de la Page" avant d'ouvrir l'image et laissez Readiris redresser la tour de Pise, telle que nous l'aimons.



Notez d'ailleurs que le panneau des documents indique quelle inclinaison a été corrigée et quels rotations ont été exécutées!

Largeur	Hauteur	Inclinaison	Rotation
173	114	5.69°	pivoté de 180 degrés



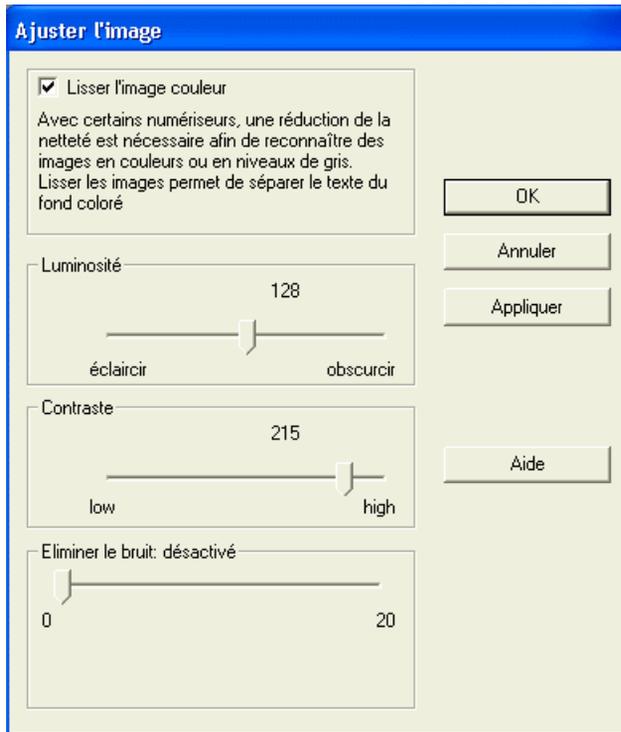
AJUSTER LES IMAGES NUMÉRISÉES

Comme indiqué auparavant, de puissantes routines intelligentes convertissent automatiquement les images en couleur et niveaux de gris en noir et blanc. Si cela est nécessaire, l'utilisateur peut optimiser l'image pour le processus d'OCR. Pour ce faire sélectionnez la commande "Ajuster l'Image" sous le menu "Action".



Ajuster l'image... Ctrl+J

Grâce à cette commande, la version en noir et blanc s'affiche automatiquement. (C'est comme si vous aviez désactivé l'option "Afficher le Document en Couleur") Interviennent ici quelques concepts un peu compliqués, que nous allons approfondir en détail.

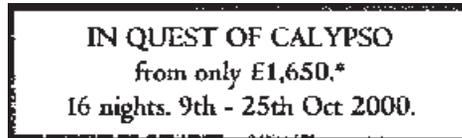
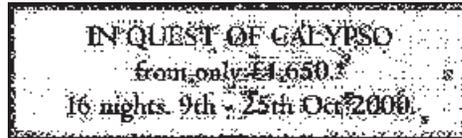
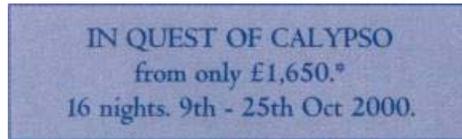


L'option "Lisser l'Image Couleur" rend les images en couleur et niveaux de gris plus homogènes en "aplanissant", lissant des différences d'intensité limitées. Il en résulte un contraste plus prononcé entre le premier plan - le texte - et l'arrière-plan - une illustration, la couleur, etc.

Une fonctionnalité de **prétraitement** peut sembler hautement technologique et difficile à comprendre, mais elle a pourtant son rôle à jouer: pour certains modèles de numériseurs, cette réduction de netteté est nécessaire pour reconnaître



des images en couleur ou en échelle de gris. Lisser est parfois la seule manière de séparer le texte de son arrière-plan coloré! Ci-dessous, vous pouvez voir un exemple d'une image qui serait tout simplement illisible sans le "lissage" d'image.



La fonction "Lisser l'Image Couleur" peut aussi être activée quand vous chargez des images préscannées dans la mémoire!



Abordons à présent la question de la **luminosité**. La luminosité règle la brillance générale de l'image: si l'on obscurcit ou éclaircit l'image, le changement de la brillance s'applique à tous les pixels. La luminosité permet d'enlever l'arrière-plan d'une page. Imaginons deux cas de figure. Dans le premier cas, tous les éléments sur l'image sont foncés. Il faut donc éclaircir l'image pour éliminer l'arrière-plan de la page. Le premier plan - le texte - doit rester suffisamment

foncé pour être détecté par la binarisation. Dans le deuxième cas, l'image est tellement claire que le texte disparaît lors de la binarisation! Dans ce cas, il est nécessaire d'obscurcir l'image pour rendre le texte lisible.

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

wyjscia kazdego
brawia, ze nasze



wyjscia kazdego
brawia, ze nasze

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

wyjscia kazdego
brawia, ze nasze

Le **contraste** influence le contraste local entre les parties claires et foncées de l'image. (Le texte est généralement plus foncé que l'arrière-plan - à moins que l'on ne traite des textes inversés...) Modifiez le contraste si vous voulez vous assurer de bien faire ressortir les formes des caractères par rapport à l'arrière-plan (en couleur). Voici un exemple d'origine où le contraste doit être augmenté. Le contraste par défaut donne des caractères cassés.

**A Look at International
Planning the Future**

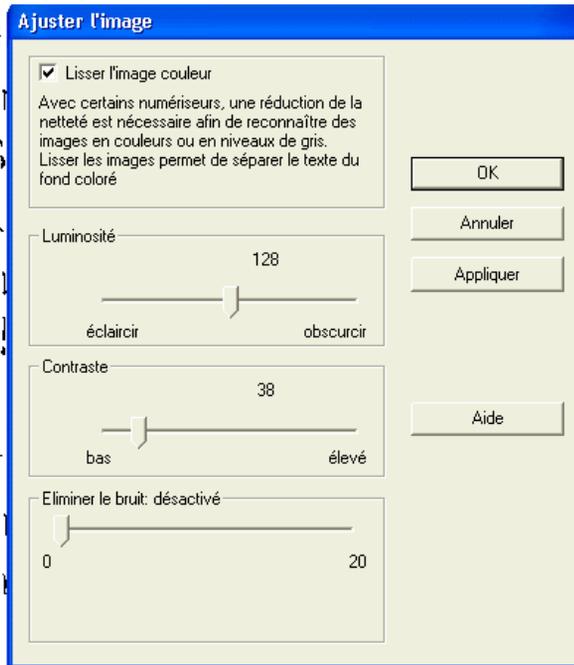
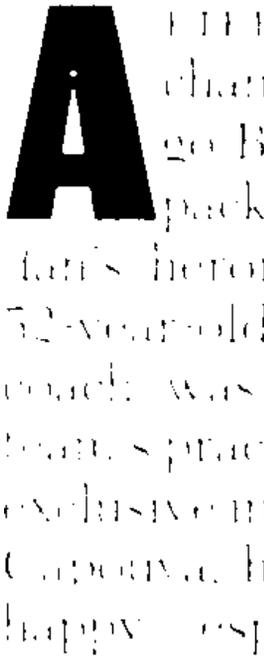
**A Look at International
Planning the Future**

**A Look at International
Planning the Future**

Remarquez cependant qu'aucun ajustement de l'image n'est effectué jusqu'à ce que vous cliquiez sur le bouton "Appliquer"! En cliquant sur "OK", vous exécutez



l'ajustement *et* fermez la fenêtre. Voici un exemple où nous avons excessivement éclairci l'image en noir et blanc - bien qu'à l'évidence sans penser à la précision de l'OCR!

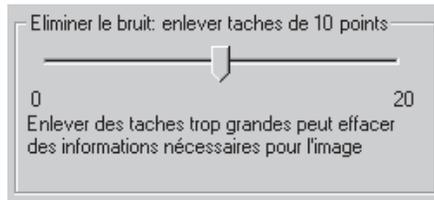


Ces options concernent les images en couleur et niveaux de gris, la dernière, "Eliminer le Bruit", concerne exclusivement les images en noir et blanc. Eliminer le "bruit" signifie que les pixels "parasites" seront éliminés des images en noir et blanc.

The Olympic Games

The Olympic Games

Assurez-vous que vous n'effacez pas des taches trop grandes, sinon vous pourriez perdre les points sur les "i", des parties de lettres matricielles, etc.!



Le meilleur moyen d'optimiser les images pour le processus d'OCR est le suivant: placez la fenêtre d'ajustement là où elle ne vous empêche pas d'évaluer le résultat de ce que vous faites. Adaptez les paramètres - en cliquant sur "Appliquer" à chaque reprise - jusqu'à ce que l'image soit claire et nette.

LAISSER L'ASSISTANT TRAVAILLER POUR VOUS

Commençons maintenant la saisie des documents. Plutôt que de parcourir tous les paramètres et options, nous allons nous servir de l'**assistant OCR**, une manière très confortable de reconnaître des pages.

Cliquez sur le bouton "Assistant OCR" dans la barre d'outils principale (ou sélectionnez la commande "Assistant OCR" dans le menu "Action").



L'assistant vous mène à travers le processus d'OCR de manière confortable: vous répondez à quelques questions simples pour obtenir des résultats rapides avec Readiris.



En fait, cet assistant démarre à chaque fois que vous lancez Readiris; afin d'éviter ce phénomène, désactivez l'option "Activer l'Assistant au Démarrage" dans le premier écran de l'assistant (ou utilisez l'option équivalente sous le menu "Configuration").

READIRIS RECRÉE LA MISE EN PAGE D'UN DOCUMENT

L'assistant OCR rend le processus de reconnaissance très autonome, mais "autonome" ne doit *pas* être confondu avec de "l'autoformatage"! Autoformatage signifie que Readiris recrée une **copie facsimile** du document numérisé: le formatage des mots, des paragraphes et de la page de votre document original sont appliqués.

Des polices similaires (serif et sans serif, proportionnel et fixes, normal et condensé) sont utilisés comme dans le document source; la taille des points, le

style de caractères (gras, italique, souligné, exposant et indice) sont maintenus au cours de la reconnaissance. Les listes à puces et les listes numérotées sont également recrées. La tabulation et l'alignement (à gauche, centré, à droite et justifié) de chaque bloc de texte sont recrées. Toutes adresses e-mail et URLs des pages web sont détectées et recrées comme hyperliens dans la sortie. L'emplacement des colonnes, des parties de texte et des graphiques imitent le document original.

Readiris vous permet donc d'archiver une véritable copie de votre document original, mais il s'agit cette fois d'un texte éditable ou d'un fichier texte compact et non pas d'une image numérisée!

Tout ceci implique qu'un tri des fenêtres exécuté par l'utilisateur ne s'applique que *partiellement* quand "l'autoformatage" est utilisé: vous pouvez inclure et exclure certaines zones, mais quel que soit le tri fait sur les résultats de l'analyse de page, il est ignoré!

Voici un exemple du mode de fonctionnement. Pour vous familiariser avec cette caractéristique, ouvrez l'image AUTOFORMAT.JPG dans votre dossier Readiris.



The screenshot shows the Readiris software interface. The main window displays a document titled "Autoformatting" with the following content:

Autoformatting

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

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

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

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

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

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

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

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

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

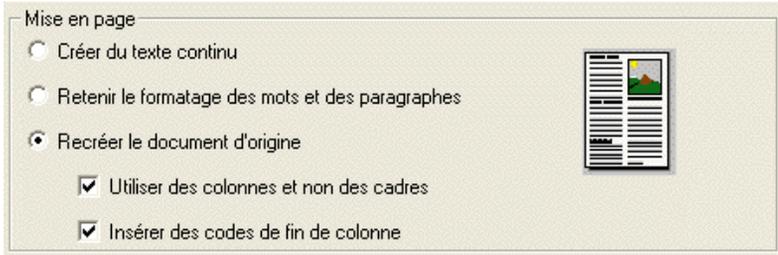
As a result, you get a true copy of your source document, be it a compact and editable text file, no longer a scanned image of your document!

Copyright: Image Recognition Integrated System

The interface also shows a sidebar with two main sections: "1 Numériser" (Digitize) and "2 Reconnaître" (Recognize). The "Format" button is visible in the "Reconnaître" section. At the bottom, a table shows the processing parameters for the document:

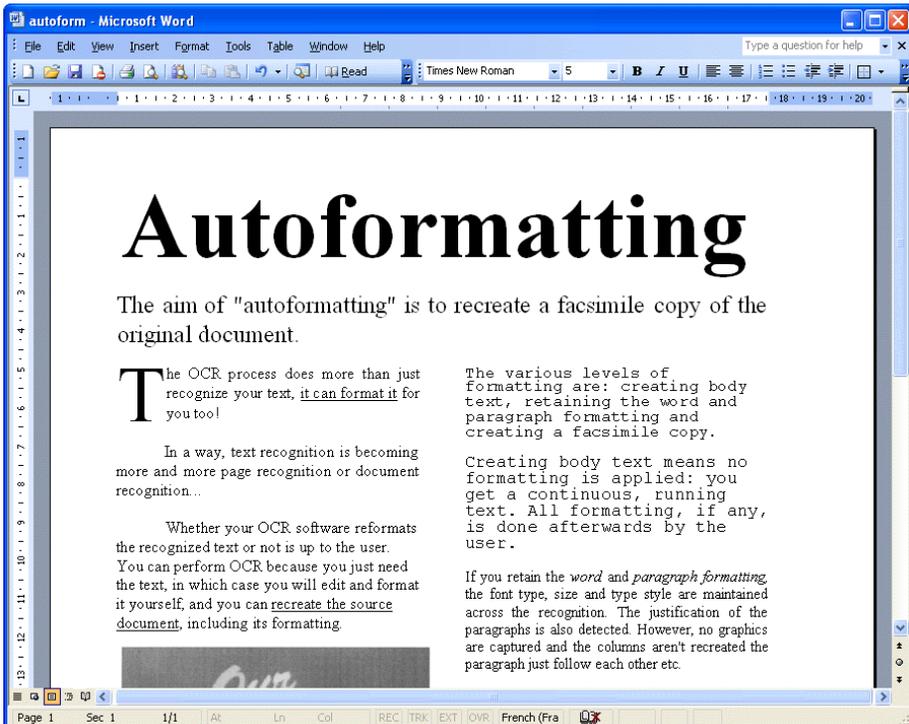
Page	Source d'image	Numérisation	Lecture	Résolution	Largeur	Hauteur	Inclinaison	Rotation
1	C:\Program Files...	4,28	-	300	2110	2615	-	-

Cliquez sur le bouton "Format" dans la barre d'outils principale et envoyez le résultat de la lecture dans Microsoft Word ou sélectionnez le format RTF (Rich Text Format) ou Word (*.doc) comme format à utiliser. Deuxièmement, activez "Recréer le Document d'Origine" comme option de mise en page. (L'option "Reconstituer les Paragraphes" est activé par défaut afin d'appliquer la frappe continue à l'intérieur des paragraphes.)

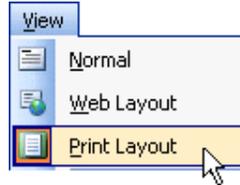


La reconstruction de la mise en page est disponible en fonction du mode de sortie sélectionné. Certains formats “pauvres” générant du texte “plat” tels que Texte (ANSI), Texte MS-DOS (ASCII), etc. *ne* supportent *pas* des codes de formats avancés et par conséquent ne peuvent pas offrir d’autoformatage. D’autre part, le format Adobe Acrobat PDF a été développé pour copier l’aspect de vos documents: les documents PDF emploient par nature l’autoformatage.

Lorsque le texte reconnu est ouvert par un traitement de texte, il ressemble à ceci sans la moindre intervention de l’utilisateur.



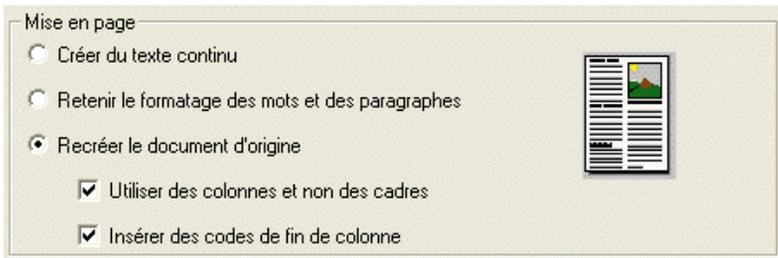
Pour bien voir le résultat, vous devez activer le mode “WYSIWIG” de votre traitement de texte, plus souvent appelé “mode de mise en page”. Cependant, si vous envoyez le document reconnu directement dans Microsoft Word, la vue “mise en page” sera activée automatiquement!



En résumé, Readiris ne reconnaît pas seulement vos textes, il est aussi capable de les formater pour vous. La technologie OCR n'effectue plus seulement de la reconnaissance de texte, elle permet également de tendre vers la **reconnaissance de documents!**

DES COLONNES S'IL VOUS PLAÎT, PAS DE CADRES!

L'option de formatage "Utiliser des Colonnes et Non des Cadres" détermine *comment* "l'autoformatage" se fera: les blocs de texte, les tables et les graphiques peuvent être placés en cadres ou en **colonnes** éditables.



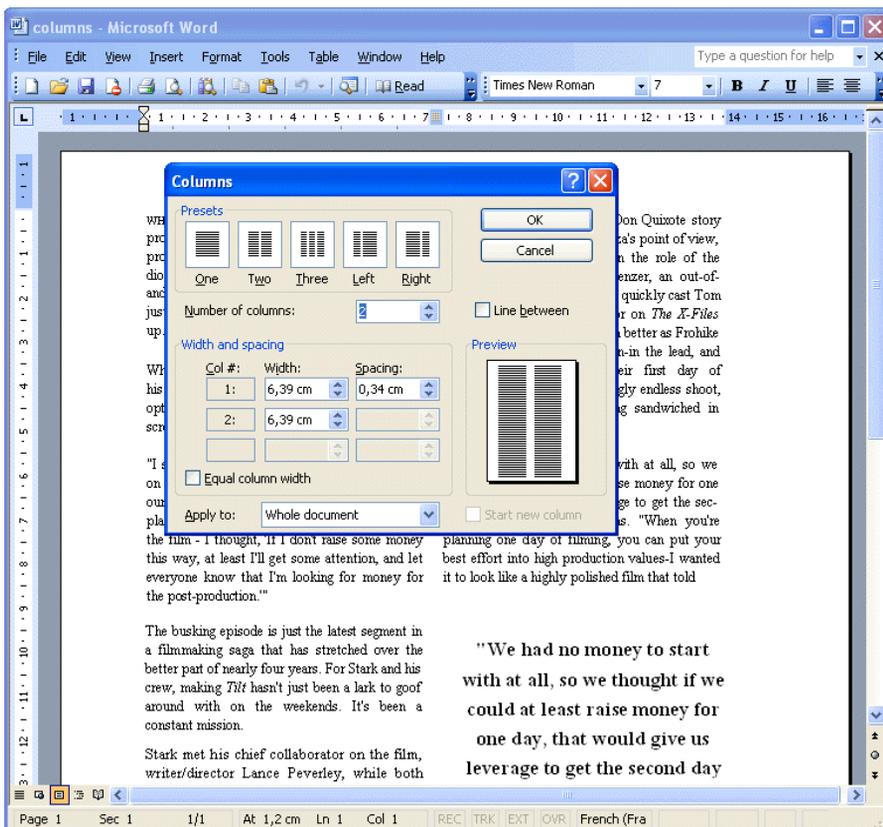
Les "cadres" sont des "réservoirs" séparés contenant du texte, ils permettent de placer plusieurs blocs de texte, des graphiques et des tables sur une page. Avec les colonnes, le texte "coule" naturellement d'une colonne à l'autre et devient bien plus facile à éditer.

Maintenant, nous assumons que des vraies colonnes apparaissent dans le document numérisé: quand le système est incapable de détecter les colonnes dans le



document source, ce mode de formatage utilise de toutes façons les cadres comme solution de repli!

Si vous voulez faire un essai, vous pouvez utiliser l'image COLUMNS.TIF dans le dossier Readiris.



L'option "Insérer des Codes de Fin de Colonnes" améliore la recréation de colonnes: celle-ci détermine si vous insérez des codes de fin de colonnes à l'extrémité de chaque colonne ou pas.

Avec des codes de fin de colonne, quel que soit le texte que vous éditez, ajoutez ou enlevez, le texte reste à l'intérieur de la colonne; le texte ne coule jamais automatiquement sur une colonne. Tout le texte qui suit un code de fin de colonne est déplacé au début de la colonne suivante!

Activez cette option si vous désirez maintenir des codes de fin de colonnes à l'endroit où celles-ci ont été détectées dans le document reconnu - quelle que soit la modification de texte réalisée après la reconnaissance. Dans les journaux et les magazines, les diverses colonnes d'une page correspondent souvent à différents articles. Mettre le flux de texte d'une colonne à l'autre "en cachette" n'est pas une bonne idée!

Désactivez cette option lorsque vous avez du texte continu en colonnes: vous vous assurerez ainsi du flux naturel du texte d'une colonne à l'autre.

Il y a un aspect du document pour lequel vous pouvez décider de *ne pas* recréer exactement la source: la taille des pages. Cette fonctionnalité est intéressante si vous numérisez par exemple des feuilles au format Lettre mais sauvegardez les résultats en A4 car les feuilles seront ainsi plus faciles à imprimer - c'est-à-dire vous ne devez pas insérer les feuilles manuellement dans l'imprimante! Ou vous êtes avocat et que vous les sauvegardez tout simplement vos documents A4 numérisées au format Légal.

Readiris vous permet donc de choisir les tailles de papier voulues pour les documents de sortie. Cliquez sur le bouton "Taille de Papier" dans la fenêtre "Format de Texte" pour accéder aux options correspondantes.



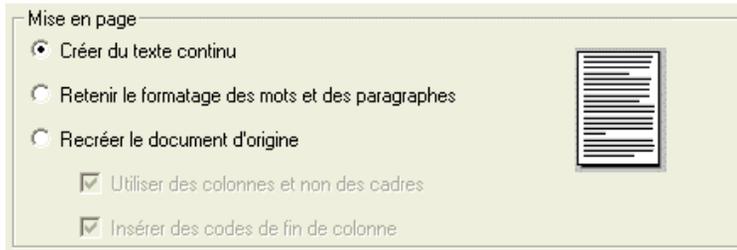
Sélectionnez les formats applicables et non permis: les formats applicables pourront être utilisés, les formats non permis ne le seront jamais. Triez les formats applicables: Readiris parcourra les formats de papier dans l'ordre indiqué et appliquera le premier format de page qui sera assez grand pour contenir le document numérisé. Le bouton "Par Défaut" permet d'appliquer à nouveau la configuration par défaut. (La configuration par défaut dépend de votre configuration Windows!)

Notez que la possibilité de définir le format de sortie ne concerne pas les fichiers HTML: ce format de texte développé pour l'Internet ne supporte pas les tailles de papier! Elle n'intervient pas non plus pour les fichiers PDF; ces derniers utilisent un format personnalisé et recréent avec fidélité le document source.

FORMATER LE TEXTE, DEUXIÈME PARTIE

Les autres options de mise en page sont "Créer du Texte Continu" et "Retenir le Formatage des Mots et des Paragraphes".

Comme l'illustre l'icône à la droite de la fenêtre, créer un **texte continu** signifie que vous créez un texte "courant" non formaté. Le texte est saisi mais sa mise en forme est totalement ignorée. Utilisez cette option lorsque vous avez besoin de resaisir un texte mais pas sa mise en page.



Le texte continu est également ce que vous obtenez lorsque vous reconnaissez rapidement une zone de texte en cliquant le bouton droit sur celle-ci et en sélectionnant le commande "Copier du Texte": lorsque la reconnaissance est faite, vous collerez le texte continu dans votre application de texte.

L'option "Retenir le Formatage des Mots et des Paragraphes" permet une "demi-mesure": le **formatage de mots** - type de fonte, la taille des points et le style de caractères (gras, italique et souligné) sont maintenus au cours de la reconnaissance. Il en est de même pour le **formatage des paragraphes** - la tabulation et l'alignement.

Ne confondez pas cette option de formatage avec l'option "d'autoformatage" complète: cette option place simplement les paragraphes les uns à la suite des autres, elle ne créera pas de colonnes ou ne copiera pas l'emplacement relatif des blocs.

EXPORTER DU TEXTE PLUSIEURS FOIS

En fait, vous pouvez exporter les résultats de l'OCR plusieurs fois sans répéter la reconnaissance! Changez le format du texte et les options de formatage sous le bouton "Format" et cliquez à nouveau sur le bouton "Reconnaitre-Enregistrer".



Aucune reconnaissance n'est exécutée cette fois - à moins que vous définissiez de nouvelles fenêtres ou modifiez les fenêtres existantes! Sinon, Readiris reformate les résultats de l'OCR et les sauvegarde dans la nouvelle mise en forme de texte ou les envoie à l'application cible que vous avez choisie.

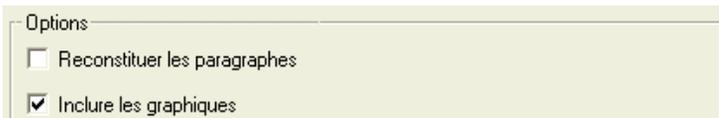


Il en va de même pour n'importe quel autre élément que vous changez : lorsque vous ajoutez une page au document, seulement cette page sera reconnue. Lorsque vous créez une nouvelle zone de texte sur n'importe quelle page, seulement cette zone sera reconnue avant que les résultats soient exportés.

Vous pourrez par exemple reconnaître un document de 10 pages et sauvegarder celui-ci dans un fichier Word. Ensuite, vous numérisez rapidement la synthèse qui se trouvait sur la page de couverture et vous l'envoyez par e-mail à un collègue impatient. Finalement, vous numérisez l'annexe - un tableau - et vous sauvegardez tous les résultats dans un fichier HTML qui sera publié sur le site web de votre entreprise.

SAUVEGARDER DES GRAPHIQUES À PART

Dans notre exemple, le graphique était inclus dans le document reconnu: ceci est dû à l'option de formatage "Inclure les Graphiques". Qu'il soit possible de sauver des graphiques à l'intérieur du texte dépend aussi du mode de sortie. Les formats de texte "pauvre" tel que Texte (ANSI), etc. ne stockent pas de graphiques!



De plus, avec Readiris, vous pouvez sauvegarder des graphiques sans effectuer de reconnaissance de texte. Comme Readiris génère des images en noir et blanc,

niveaux de gris et couleur, vous pouvez saisir des graphiques en mode trait et des photos.

Comment? En dessinant une fenêtre graphique autour de l'illustration, le dessin, etc. dont vous avez besoin. Ceci se fait de la même manière que pour dessiner des fenêtres de texte et table; sélectionnez simplement l'outil "Fenêtre Graphique" cette fois-ci.



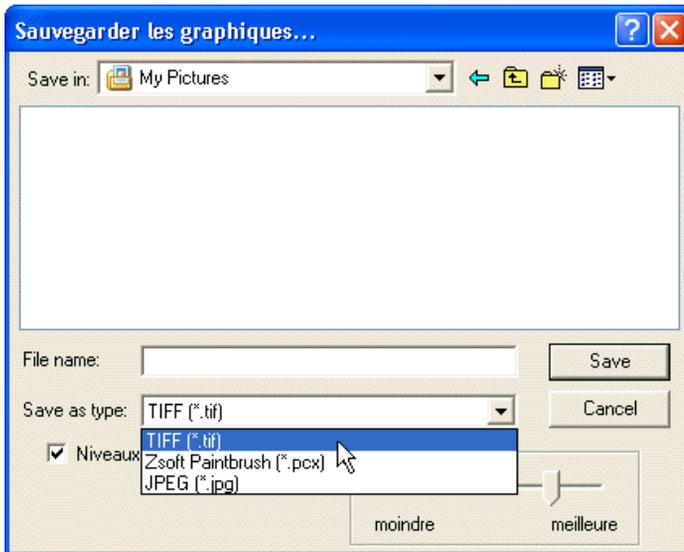
Sélectionnez maintenant la commande "Sauvegarder les Graphiques" sous le menu "Fichier".

A screenshot of a blue menu item with the text "Sauvegarder les graphiques..." and a mouse cursor pointing to it.

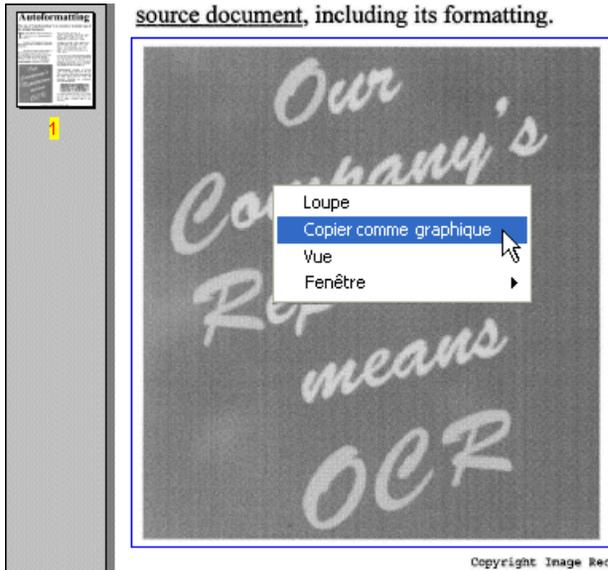
Sauvegarder les graphiques...

Vous êtes invité à donner un nom de fichier. Déterminez quel format de fichier graphique vous utiliserez. Choisissez un format supporté par votre logiciel de retouche photo. Les formats JPEG, TIFF et ZSoft Paintbrush (*.pcx) sont supportés. Readiris Corporate offre aussi le format compact JPEG 2000 (*.j2c).

Activez l'option "Niveaux de Gris/Couleur" pour sauver le graphique comme un graphique couleur ou gris. Quand vous sauvegardez un graphique en noir et blanc au format TIFF, la compression Groupe 4 est utilisée. Quand vous sauvegardez un graphique en niveaux de gris ou en couleurs au format TIFF, la compression JPEG est utilisée et vous pouvez déterminer la qualité JPEG.



Pour envoyer un graphique vers le presse-papiers plutôt que de sauvegarder un fichier image, cliquez avec le bouton droit de votre souris sur une fenêtre graphique et sélectionnez la commande "Copier comme Graphique": la zone graphique sous le pointeur de la souris est prête à être collée!



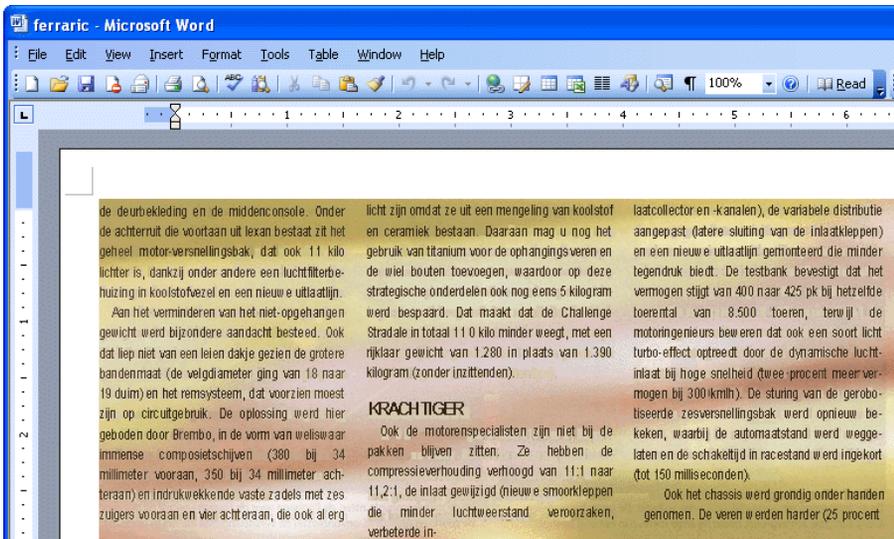
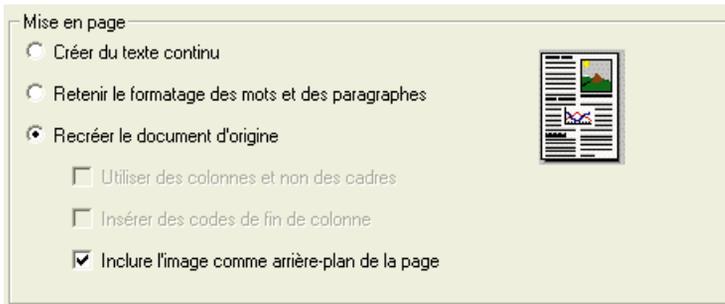
SAUVEGARDER LES COULEURS DE L'ARRIÈRE-PLAN

Avec Readiris Corporate, vous pouvez même sauvegarder l'arrière-plan des documents reconnus! L'option "Inclure l'Image comme Arrière-plan de la Page" permet d'insérer l'image scannée de l'arrière-plan en-dessous du texte reconnu. Cette fonctionnalité existe pour les formats HTML, RTF, Word (*.doc) et WordML (*.xml). (Quand vous générez des fichiers PDF, vous obtenez le même résultat avec le format "PDF Texte-Image".)

Il va de soi que cette option augmente de façon considérable la taille du fichier de vos documents reconnus. Après tout, cette option ajoute l'image numérisée avec tous ses détails au document de sortie. *Tous* ses détails? Pas vraiment: comme c'est le cas pour la sortie PDF "texte-image", les pixels du texte reconnu



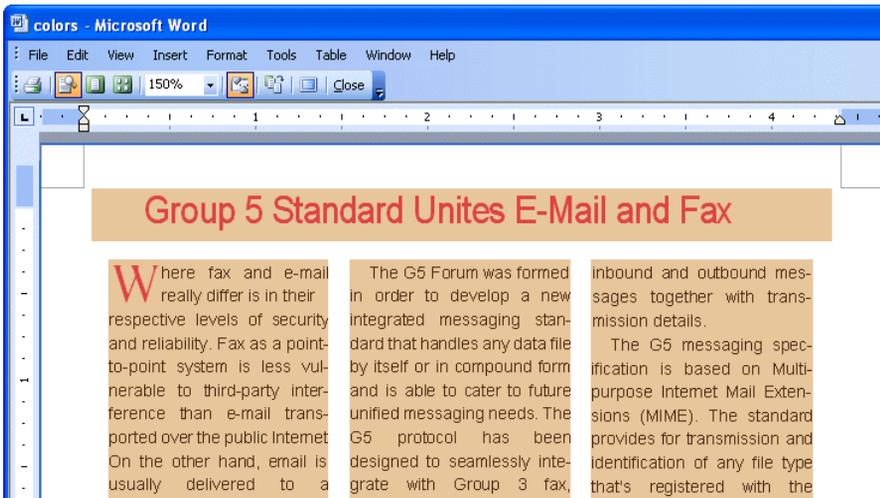
sont effacés pour assurer la lisibilité du document. Afficher du texte en noir au-dessus de bitmaps de caractères noirs vous donnerait un texte avec une ombre lourde... (L'image BACKGROUND.JPG offre un bon exemple.)



Readiris offre une alternative moins “spectaculaire” et plus compacte: l’option "Retenir les Couleurs de l’Arrière-Plan" retient les blocs de couleurs à travers la reconnaissance. (Cette option vous oblige à garder les couleurs du texte.)



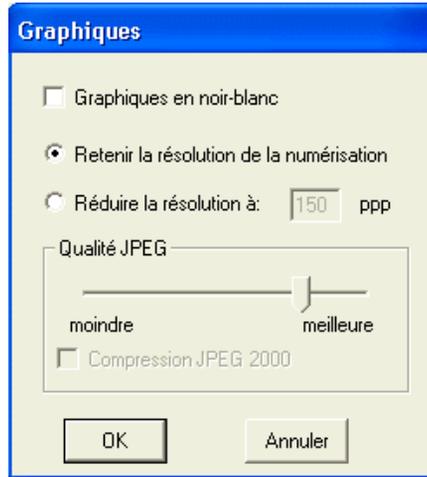
Vous obtenez une couleur uniforme par paragraphe - si le document source le permet bien sûr. Cette option également concerne uniquement les documents HTML, Word et WordML. Les détails d’une photo couvrant tout l’arrière-plan d’un document *ne* seront *pas* retenus, seules les couleurs uniformes des cadres de texte le seront. (Reconnaissez l’image d’exemple COLORS.JPG si vous voulez essayer...)



ALLEZ PLUS LOIN AVEC VOS GRAPHIQUES!

Readiris Corporate offre d'autres options avancées pour vos graphiques. Vous les trouverez sous le bouton "Avancé" de la commande "Format Texte".

Ces paramètres sont applicables à *tous* les graphiques - les zones graphiques incluses dans des documents reconnus et l'image de page que vous placez au-dessus du texte dans un fichier Adobe Acrobat PDF du type "image-texte"!



Déterminez le mode de couleur: enregistrez vos graphiques comme images en échelle de gris ou noir et blanc. Sélectionnez la résolution des images dans les documents reconnus: maintenez la résolution de numérisation ou réduisez-la. (Attention: vous ne pouvez pas *augmenter* la résolution en procédant de la sorte!) Lorsque les documents reconnus sont publiés sur un site web comme des fichiers HTML, vous voudrez sans doute adapter la résolution des graphiques à la résolution écran. En effet, les graphiques des fichiers HTML sont réduits par défaut à 72 ppp. (Vous pouvez remplir manuellement une valeur plus élevée.) Enfin, sélectionnez la qualité JPEG. (Les images JPEG sont utilisées pour placer des graphiques couleur ou niveaux de gris dans des documents PDF, Word, RTF, etc.)

(Nous avons déjà traité l'option "Compression JPEG 2000": elle applique la compression JPEG 2000 à tout graphique et toute image contenu dans un document PDF.)

Ces options vous permettent d'influencer substantiellement la **taille de fichier** de vos documents! Donnons un exemple: si vous générez des fichiers de type

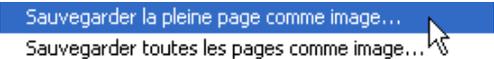


PDF "image-texte" avec des graphiques monochromes, vous stockez les images dans des fichiers TIFF comprimés Groupe 4. Si vous sauvegardez ces mêmes fichiers sous la forme d'images couleur, vous stockez par défaut des fichiers JPEG de haute qualité (0,8).

LECTURE DE FAX ET RECONNAISSANCE EN DIFFÉRÉ

Sauvegarder des images comme fichiers image offre une autre possibilité: vous pouvez sauvegarder la page *entière* et effectuer l'**OCR en différé**. C'est ce que nous avons fait avec les images prés캔ées au cours de ce didacticiel.

Il vous suffit de numériser le document. Sélectionnez la commande "Sauvegarder la Pleine Page Comme Image" du menu "Fichier". Vous serez à nouveau invité à enregistrer l'image entière dans un fichier PDF, JPEG, JPEG 2000 (*.j2c), TIFF ou ZSoft Paintbrush (*.pcx).



Sauvegarder la pleine page comme image...
Sauvegarder toutes les pages comme image...

Sélectionnez la commande "Sauvegarder Toutes les Pages comme Image" pour enregistrer un document multipage. Deux formats de fichier sont disponible ici: PDF et TIFF multipage.

Vous pouvez maintenant sélectionner le disque comme source d'images et ouvrir le fichier image avec le bouton "Ouvrir" (ou à l'aide de la commande similaire sous le menu "Action"). (Si vous utilisez la commande "Ouvrir" dans le menu "Fichier", vous ne devrez pas actualiser la source d'image.)

Comme les images couleur, niveaux de gris et noir et blanc sont indifféremment supportées, Readiris ouvre des documents Adobe Acrobat PDF, des télécopies DCX (une version multipage du format Paintbrush), des images DjVu (*.djv, *.djvu), des images JPEG, des images JPEG 2000 (*.j2c, *.jp2), des images TIFF (non-compressées et compressées en LZW, PackBits, Groupe 3, Groupe 4 et JPEG), des images TIFF multipages, des bitmaps Windows (*.bmp) et ZSoft Paintbrush (*.pcx).

Cette fonctionnalité est particulièrement utile pour convertir vos **télécopies** en textes éditables! Readiris utilise des routines intelligentes supplémentaires quand il doit lire des télécopies: le logiciel détecte les résolutions de fax standards- 100 x 200 ppp ("qualité normale"), 200 x 200 ppp ("qualité fine") et 200 x 400 ppp ("qualité superfine") - et "prétraite" ces images automatiquement pour assurer des résultats OCR optimaux.

Ceci dit, il est toujours recommandé de demander à vos correspondants de vous envoyer des fax avec qualité "fine" - ils donneront de meilleurs résultats OCR.

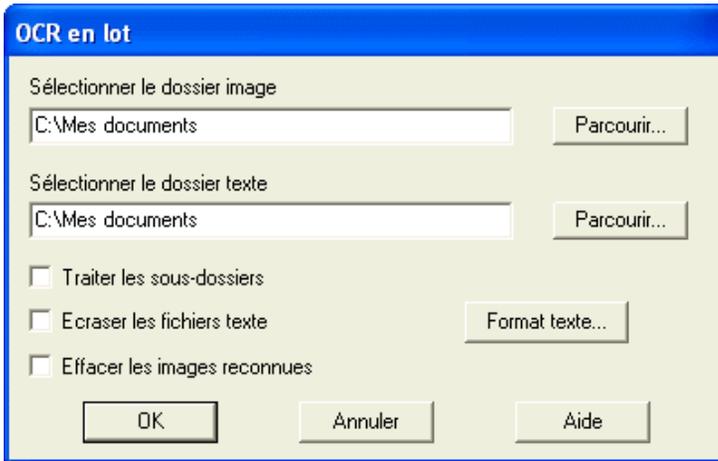
N'oubliez pas que vous pouvez cliquer avec le bouton droit de la souris sur les images dans l'Explorateur de Windows et sélectionner la commande "Reconnaître" dans le menu "Contexte" pour ouvrir des images! De même vous pouvez "glisser-déposer" des images pour les ouvrir: faites glisser des fichiers image de l'Explorateur de Windows vers la zone image ou l'icône de Readiris et elle sont immédiatement ouverte.

RECONNAISSANCE DE DOCUMENTS EN LOTS

Readiris Corporate est beaucoup plus puissant pour reconnaître des images prénumérisées: vous pouvez reconnaître automatiquement des lots entiers de documents et vous pouvez établir un dossier surveillé. Etudions cela en détails!

L'**OCR en lot** exécute la reconnaissance de toutes les images prénumérisées dans un dossier spécifique. Vous pouvez par exemple numériser vos documents pendant la journée et lire pendant la nuit... Scannez tous les documents à reconnaître; quand vous avez fini, démarrez l'OCR en lot et laissez votre PC parcourir son traitement. Lorsque vous reviendrez le lendemain, tous vos documents seront reconnus!

Les documents reconnus ont le même nom de fichier que les fichiers d'image. L'extension du fichier dépend évidemment du format de sortie sélectionné: lorsque que le format Word est sélectionné, le fichier image 001.TIF est converti en 001.DOC.



Sélectionnez le dossier image et le dossier texte. Le dossier texte peut être différent du dossier image - mais cela n'est pas obligatoire! Lorsque le dossier image est identique au dossier de sortie, vous trouverez les documents de texte à côté des images numérisées (à moins que vous activiez l'option "Effacer les Images Reconnues")!

L'option "Traiter les Sous-Dossiers" détermine si les sous-dossiers du dossier image seront également traités. Ceci a du sens lorsque les documents scannés sont placés dans des sous-dossiers spécifiques. Il se peut que vous ayez par exemple un dossier nommé "18-2-2005" qui indique la date et des sous-dossiers "1", "2", etc. qui contiennent les documents mêmes.

Si cette option est activée, elle déclenche le traitement de *tous* les sous-dossiers. Vous ne pouvez pas limiter la reconnaissance à un (des) sous-dossier(s) spécifique(s). Lorsque le dossier texte diffère du dossier image, les sous-dossiers sont recréés, "reflétés" dans le dossier de sortie!

Cliquez sur "OK" pour exécuter. Aucun effort supplémentaire n'est nécessaire pour convertir les documents: le traitement de reconnaissance est entièrement automatique. (L'apprentissage interactif ne s'applique pas.)

De plus, vous êtes limité aux formats de texte "externes". Les résultats de l'OCR sont sauvegardés sur le disque. Envoyer les résultats de lecture directement vers un application cible, les envoyer par e-mail ou les ouvrir automatiquement après la reconnaissance simplement n'a pas de sens lorsque l'OCR est exécuté sur un PC non surveillé.

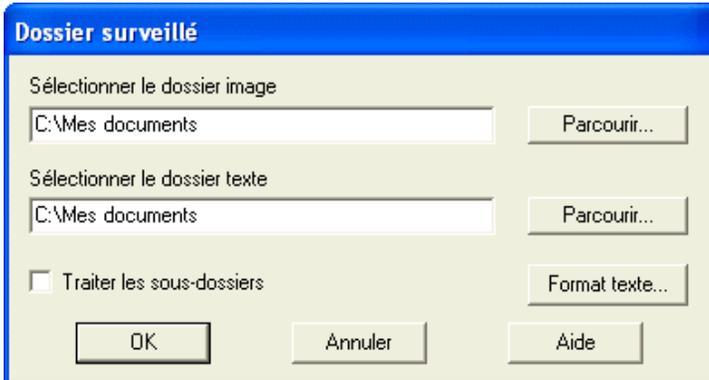
L'OCR en lot fonctionne avec des options avancées. L'option "Ecraser les Fichiers Texte" détermine si le traitement de l'OCR peut remplacer les résultats de la reconnaissance précédente. Désactivez cette option lorsque vous ajoutez des nouveaux fichiers image à un dossier qui contient des images qui ont été auparavant traitées. (Sinon, vos fichiers image seront reconnus une deuxième fois lorsque vous traitez le dossier.)

L'option "Effacer les Images Reconnues" détermine si les fichiers image seront supprimés après la reconnaissance. En d'autres mots, Readiris peut nettoyer votre dossier image à votre place!

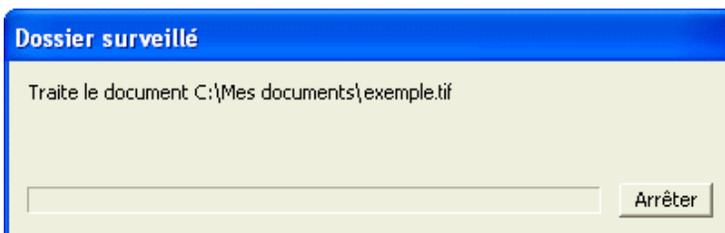
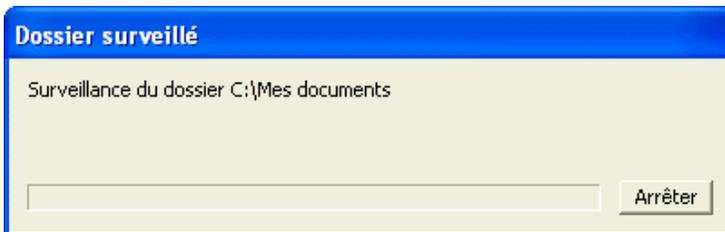
Activez cette option lorsque vous stockez - archivez les documents reconnus, en effaçant les fichiers image "temporairement". Désactivez cette option lorsque vous sauvegardez à la fois les numérisations et les documents reconnus. (Nous répétons que Readiris génère les documents PDF de type "image-texte": ce format de sortie sauvegarde les images numérisées *et* le document reconnu dans un fichier unique!)

ÉTABLIR UN DOSSIER SURVEILLÉ

L'utilisation d'un "dossier surveillé" est fort similaire à l'opération de l'OCR en lot. La différence majeure est que, cette fois, l'utilisateur ne déclenche pas la reconnaissance. Au contraire, Readiris exécute systématiquement la reconnaissance sur un fichier de l'image qui est placé dans un dossier spécifique. Vous pouvez laisser tourner le logiciel OCR au quotidien... Numérisez des nouveaux documents et ils seront reconnus aussitôt.



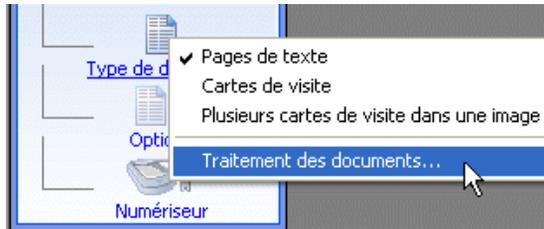
Cliquez sur "OK" pour activer la "supervision" du **dossier surveillé** et Readiris traite les images progressivement dès qu'elles sont créées-placées dans le dossier surveillé. Cliquez sur "Arrêter" pour interrompre la supervision.

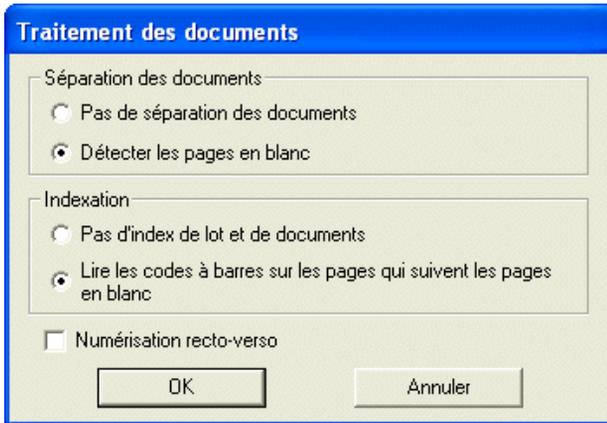


Comme c'était le cas avec l'OCR en lot, Readiris traite les images de tous les formats de fichier supportés. Vous ne pouvez pas limiter l'OCR à un format graphique spécifique: Readiris traite les images de *tous* les formats de fichier supportés. (Tous les fichiers avec un autre format fichier sont simplement ignorés.)

ORGANISER DES LOTS DE DOCUMENTS

Cela ne veut pas dire que vous êtes obligé de numériser les documents sans plus, de façon aveugle. Bien au contraire, Readiris Corporate offre des routines avancées qui traitent les lots de documents de façon intelligente. Accédez à la commande "Traitements des Documents" sur le bouton "Type de Document" (ou sous le menu "Configuration") pour découvrir ces fonctionnalités.



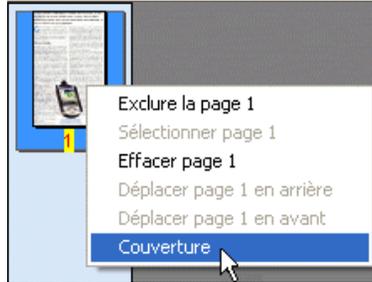


Vous pouvez insérer des couvertures blanches entre les pages pour **séparer** les **documents**. Une page "blanche" est une page qui ne contient pratiquement pas de pixels noirs - exception faite des bords noirs. La barre d'outils des pages met en évidence les couvertures, le panneau des documents les mentionne.



Page	Couverture	Code à barres	Source d'image	Numérisation
<input checked="" type="checkbox"/> 1	oui		C:\Program File...	1.48
<input checked="" type="checkbox"/> 2	-	TEST2	C:\Program File...	1.30

Mais vous pouvez aussi segmenter manuellement un flux d'images numérisées: sélectionnez une page - blanche ou pas - dans la barre d'outils des pages, cliquez avec le bouton droit de la souris et activez l'option "Couverture" du menu "Contexte". (Le contenu d'une couverture est de toute façon perdu: les couvertures permettent la séparation des documents, mais elles *ne* sont *pas* lues!)

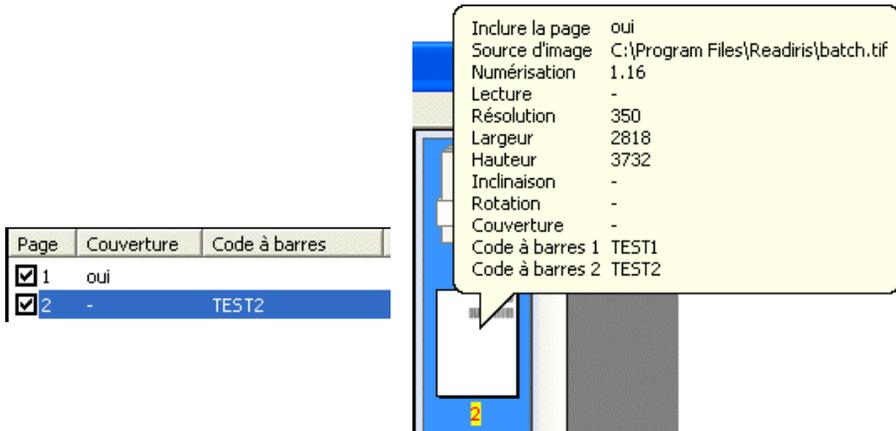


Isoler les documents individuels dans un lot d'images n'est que la première étape. Vous pouvez ensuite **indexer** ces documents correctement grâce à la lecture de codes à barres. Tous les codes à barres qui figurent sur une page qui suit une couverture blanche sont lus automatiquement. (Exécutez la reconnaissance sur l'image d'exemple BATCH.TIF si vous souhaitez essayer.)



L'option "Numérisation Recto-Verso" indique que vous numérisez la face avant et arrière des documents. Les faces arrières seront ignorées par la lecture automatique des codes à barres: la page suivante devient alors la *face avant* la suivante, non pas l'image suivante!

L'interface de Readiris vous assiste de plusieurs façons dans le traitement de lots de documents: le panneau des documents et les bulles explicatives sur les icônes des pages contiennent la valeur des codes à barres reconnus!



Les résultats de la lecture des codes à barres sont exclusivement enregistrés dans l'index XML, et non pas dans les documents reconnus!

Quand vous convertissez des lots manuellement, l'option "Créer un Fichier par Document" de la fenêtre de sauvegarde enregistre des documents séparés, pas des lots complets. (Cette option est uniquement disponible quand vous créez des fichiers externes; elle n'est pas offerte lorsque vous envoyez les résultats directement vers une application cible.)



RECONNAISSANCE DES TABLES

Jusqu'ici, nous avons reconnu des textes, des télécopies et sauvegardé des graphiques. Traitons maintenant une table. Prenez une table et numérisez-la ou ouvrez l'image en exemple TABLES.JPG de votre dossier Readiris.

En fait, l'image TABLES.JPG contient deux tables et ce n'est pas un hasard! L'analyse de page détecte les deux tables comme zone de table, et Readiris les reconstruira en créant une table cellule par cellule dans votre feuille de calcul ou en insérant un objet table dans votre traitement de texte.

Découvrons ensemble les possibilités, en commençant par la table encadrée. Une table "encadrée" a des bords autour de chaque cellule.



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.

Page	Source d'image	Numérisation	Lecture	Résolution	Largeur	Hauteur	Inclinaison	Rotation
<input checked="" type="checkbox"/> 1	C:\Program ...	4.49	-	300	2092	2517	-	-

Effectuez la reconnaissance avec l'option de mise en page "Retenir le Formatage des Mots et des Paragraphes" ou l'option "Recréer le Document d'Origine" et la table sera recréée. Ouvrez votre traitement afin de vérifier la sortie: les cellules et les bords sont recréés par Readiris un par un! (Vous pourriez aussi bien avoir inclus les paragraphes textes dans le fichier texte.)

The screenshot shows a Microsoft Word window titled 'table - Microsoft Word'. The table is centered on the page and contains the following data:

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					

Maintenant l'exemple où la table n'a pas de bordure autour des cellules. Remarquez que l'analyse de page détecte néanmoins la table!

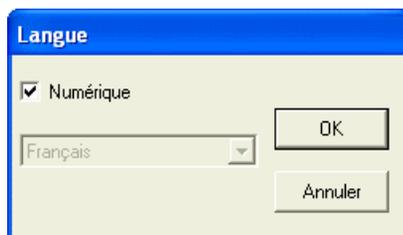


The screenshot shows the Readiris software window with the following elements:

- Menu Bar:** Fichier, Edition, Configuration, Vue, Action, Apprendre, S'enregistrer, Aide
- Left Panel:**
 - Assistant OCR**
 - 1 Numériser**
 - Numériser
 - Source
 - Options
 - Numériser
 - 2 Reconnaître**
 - Reconnaître Enregistrer
 - Anglais
 - Apprentissage
 - Format
- Main Content Area:**
 - Text: "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.
 - Text: When your tables exclusively contain **numeric characters**, enable the numeric reading mode with the "Language" button on the main toolbar for increased accuracy.
 - | | | | |
|---------|-----|------|-------|
| 123.985 | 69 | 3137 | 2.395 |
| 987.745 | 129 | 24 | 5.583 |
| 19 | 149 | 915 | 91.54 |
| 287.410 | 49 | 526 | 9.812 |
| 429.000 | 0 | 17 | 17.42 |
| 499.123 | 149 | 28 | 112.9 |
 - Text: 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.
 - Text: Copyright Image Recognition Integrated Systems
Web site: <http://www.irislink.com>
- Bottom Table:**

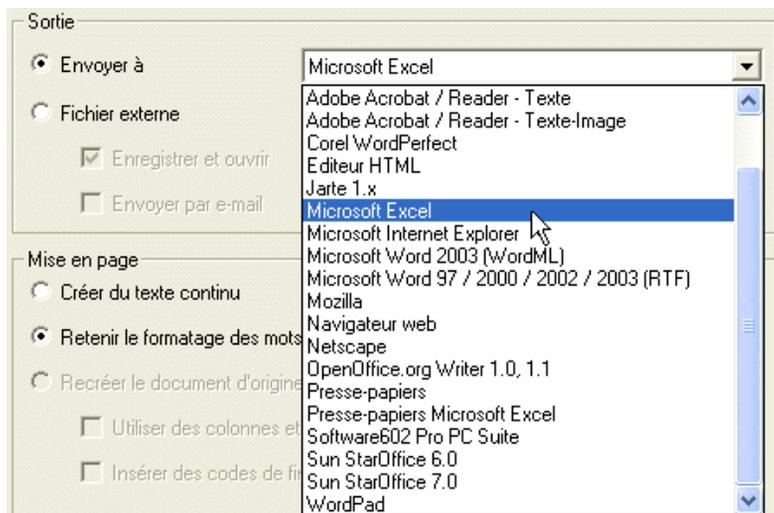
Page	Source d'image	Numérisation	Lecture	Résolution	Largeur	Hauteur	Inclinaison	Rotation
<input checked="" type="checkbox"/> 1	C:\Program Fi...	4.31	-	300	2092	2517	-	-

Pour obtenir une exactitude optimale d'OCR, vous devez limiter la reconnaissance aux **symboles numériques** avec le bouton "Langue". (Le mode numérique n'est pas strictement numérique, il inclut les symboles "0" à "9", "+", "*", "/", "%", "," (virgule), "." (point), "(", ")", "-", "=", "\$", "£", "¥" et le symbole "€".)



Comme vous ne pouvez faire ceci que lorsque la table ne contient pas de symboles alphabétiques - autrement les parties de texte ne seront pas reconnues correctement - nous pouvons maintenant activer le mode numérique mais ne pourrions pas le faire pour la première table.

Cette fois-ci, nous enverrons le résultat directement dans le tableur Microsoft Excel; nous sélectionnons donc Excel comme application cible à l'aide du bouton "Format".





Le tableur est lancé automatiquement et la sortie se présente de cette façon: la structure typique de table avec lignes et colonnes est recréée, et vous pouvez commencer immédiatement le traitement des données.

	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					

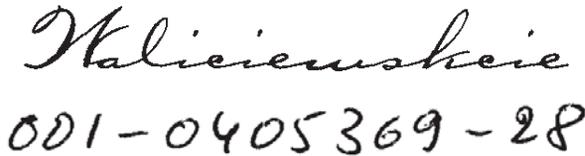
Il se peut que l'analyse de page ne détecte pas certaines tables non encadrées comme des fenêtres de tables; ceci peut arriver quand les colonnes sont trop espacées. Readiris essaye en effet d'éviter la confusion avec des blocs de texte organisés en colonnes. Afin de créer manuellement une fenêtre de table, cliquez sur l'outil "Fenêtre Table" dans la barre des outils image et créez la fenêtre. La bulle explicative du bouton indiquera le nombre de fenêtres table.



RECONNAÎTRE DES TEXTES MANUSCRITS

Nous avons déjà reconnus des documents scannés, des tables, des télécopies, des instantanés pris avec un appareil photo numérique, nous avons sauvés des graphiques et convertis des documents PDF. Sachez sur Readiris offre une capacité de lecture supplémentaire: la reconnaissance de textes manuscrits.

En réalité, nous devrions parler de texte manuscrit “imprimé”, pas que de texte manuscrit! En effet, le terme “écriture manuscrite” fait référence aux textes dits “cursifs” dans lesquels les symboles ou caractères se touchent. Dans l’écriture cursive, il est impossible de déterminer la limite précise entre la fin d’un caractère et le début du suivant. Si l’on considère l’écriture manuscrite *imprimée* cependant, les “lettres imprimées” ne se touchent pas et le logiciel de reconnaissance est capable d’isoler les caractères individuels.



Kalicienuskeie
001-0405369-28

La reconnaissance de symboles manuscrits imprimés requiert toutefois l’utilisation de logiciels hautement spécialisés - les logiciels “ICR” (“Intelligent Character Recognition”). La puissante technologie ICR d’I.R.I.S. est basée sur des échantillons d’écriture d’un million de personnes! Readiris supporte tous les styles d’écriture naturelle - y compris l’écriture américaine. Readiris n’impose pas de style d’écriture particulier.

La reconnaissance de textes manuscrits est limitée aux chiffres (0-9), aux lettres majuscules (A-Z) et aux symboles de ponctuation “,” (virgule), “.” (point) et “-” (trait d’union).

Si vos résultats de lecture ne s’avèrent pas suffisamment bons, vous pouvez adapter votre style d’écriture et utiliser le formulaire d’écriture optimisée d’I.R.I.S. Consultez l’aide en ligne de Readiris pour prendre connaissance des règles de bonne écriture. Quelques astuces simples peuvent vous faire éviter des substitutions de caractères. Le formulaire vierge d’I.R.I.S. sert de “gabarit”. Il vous permet de remplir les caractères imprimés dans des cases qui ont la bonne taille et qui sont correctement espacées. Vous trouverez ce formulaire vierge sur le CD-ROM de Readiris; vous pouvez donc l’imprimer et même le modifier.



Formulaire pour l'écriture manuscrite et règles d'écriture

Quand les résultats de lecture ne sont pas optimaux, vous pouvez utiliser le **formulaire d'écriture** d'I.R.I.S. et **adapter** votre **style d'écriture**.

- Le formulaire d'écriture I.R.I.S. blanc sert comme "gabarit" sur lequel les **lettres imprimées** sont remplis avec le bon espacement et la taille correcte!

Astuce: vous trouvez le formulaire d'écriture vide sur le **CD-ROM** de Readiris. Le fichier Adobe Acrobat PDF permet la **réimpression**, le fichier Word (*.doc) permet l'**édition** éventuelle. Imprimez le formulaire en **couleur** et remplissez-le. Ou imprimez-le en **noir et blanc** et placez-le comme "guide" en-dessous du papier sur lequel vous écrivez.

- Quelques simples **règles** d'écriture permettent de comprendre pourquoi des substitutions se produisent et comment on les évite.

L'écriture est dans le stylo!

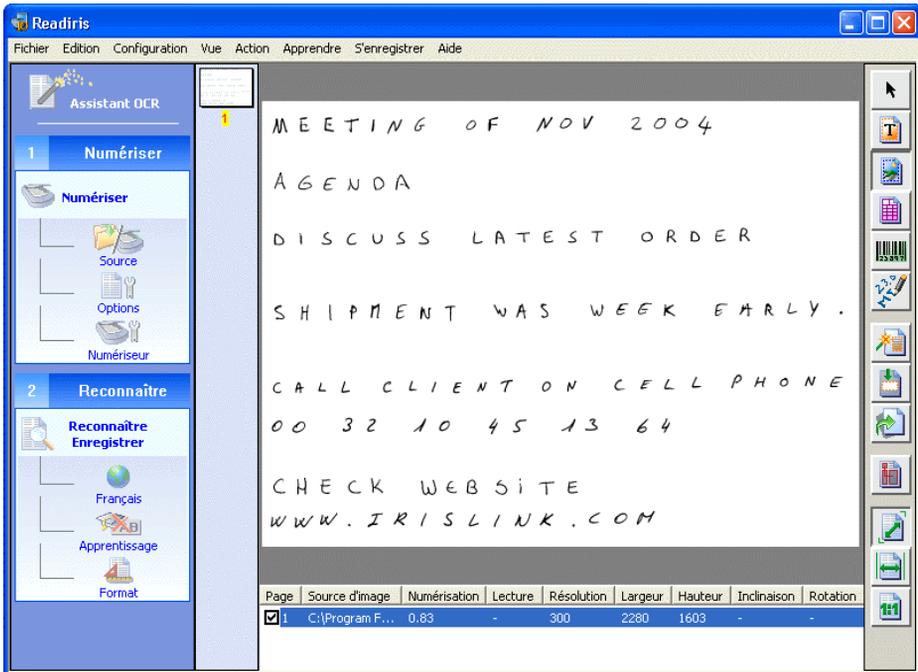
Il est important utiliser une **bille** suffisamment épaisse. Des billes trop fines ne fournissent pas

Vous savez à présent comment produire des notes manuscrites lisibles par la "machine". Voyons à présent comment faire pour faire "lire" ces notes par Readiris. Dans le logiciel, il vous suffira en fait de dessiner une fenêtre autour du texte et d'exécuter la reconnaissance. (Vous pouvez essayer avec l'image d'exemple HANDPRINTING.TIF.)

Dessiner une fenêtre de texte manuscrit: 0



Dans le cas de l'écriture manuscrite, les caractéristiques du document - langue, type de police et pas des caractères - ne sont pas prises en compte. Vous êtes limité à un jeu de caractères de base anglais - on pourrait parler du jeu de caractères "latin" - de surcroît composé essentiellement de caractères imprimés (majuscules). En ce qui concerne les routines d'apprentissage, celles-ci ne sont pas non plus d'application dans le cas de textes manuscrits: l'apprentissage n'a aucun sens dans un contexte où tout le monde a son écriture personnelle. (Rappelons que la technologie ICR est basée sur plus d'un million d'écritures différentes...!)



LIRE LES BARRES ET LES ESPACES

Et ce n'est pas tout... Readiris lit aussi les codes à barres! Les codes à barres qui figurent dans les images numérisées peuvent être lus et intégrés comme des données reconnues dans les documents de sortie.

Un code à barres est composé de barres et d'espaces entre ces barres. Des combinaisons prédéfinies de barres et d'espaces représentent des caractères

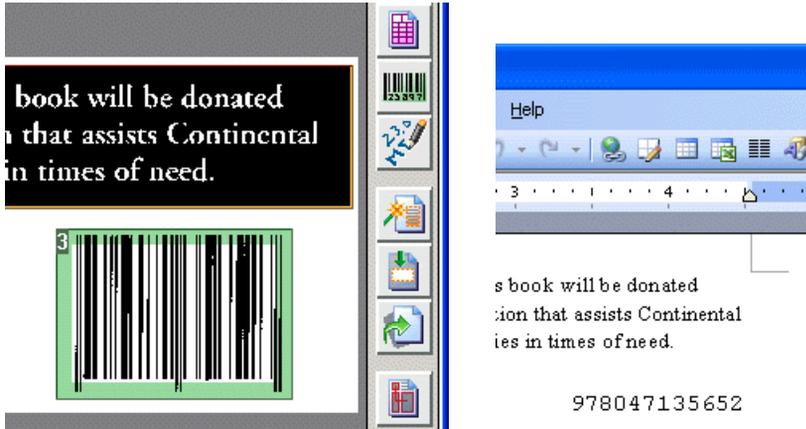
spécifiques. Il y a plusieurs standards ou "symbologies" de codes à barres; Readiris supporte toutes les symbologies standards.



Les codes à barres imprimés avec une imprimante laser ou une imprimante à jet d'encre offrent une qualité d'impression suffisante. Il faut par contre exclure les codes à barres imprimés avec une imprimante matricielle: ils n'offrent pas assez de contraste et leur résolution est normalement limitée à 60 ppp! Readiris reconnaît correctement les codes à barres bien contrastés. La plupart des codes à barres sont entourés d'une "zone tranquille" autour du code même. Un code à barres ne génère jamais un résultat partiel: si un caractère de début ou de fin est manquant ou si un caractère de contrôle est invalide, cela provoque toujours une erreur de lecture et le résultat est nul!



Pour lire les codes à barres, dessinez une fenêtre de code à barres autour de chaque code à barres - car l'analyse de pages ne les détecte pas - et lancez la reconnaissance. Les codes à barres sont lus et inclus dans le texte de sortie. Vous pouvez également procéder comme suit: cliquez avec le bouton droit de la souris sur une fenêtre de codes à barres et sélectionnez la commande "Copier comme Données" du menu "Contexte"; le code à barres est lu et envoyé au presse-papiers... (Les caractères de contrôle de certains standards de codes à barres sont vérifiés mais enlevés des résultats de lecture.) L'image d'exemple BARCODE.TIF illustre comment fonctionne la lecture des codes à barres.



RECONNAÎTRE DES CARTES DE VISITE

Nous avons déjà reconnus des documents scannés, des tables, des télécopies, des instantanés pris avec un appareil-photo numérique, nous avons sauvés des graphiques et converti des documents PDF. Readiris Corporate ajoute encore une autre capacité de lecture: la **reconnaissance de cartes de visite** (“BCR” - “Business Card Reading”)!

Avec Readiris, vous pouvez numériser vos cartes de visite, les reconnaître et les convertir en une **base de donnée d’adresses**. Dans ce contexte, l’OCR permet d’encoder vos cartes de visite sans devoir passer par la tâche fastidieuse de l’encodage. Rappelez-vous de votre dernière foire commerciale, lorsque vous êtes revenu avec une pile de cartes de visite que votre secrétaire a mis deux jours à encoder!

Les données de la carte sont extraites automatiquement de l’image et les données reconnues sont attribuées à des champs spécifiques de la base de données. Readiris utilise une base de données de connaissance, et dispose de

cette façon de l'intelligence nécessaire pour distinguer un prénom et un nom, une ville et son pays, un numéro de téléphone et un numéro de télécopie, etc. Chaque pays possède son propre style pour la composition des cartes de visite: les Américains composent les adresses d'une manière différente des Français, etc.

Ce logiciel supporte non moins de 28 **pays**: les cartes de visite de l'Amérique du Nord et du Sud et les cartes de visite des pays européens sont reconnues, y compris des pays de l'Europe de l'Est. (En option, vous pouvez lire les cartes de visite de la Chine, du Japon, de Corée et de Taiwan.)

Les données traitées sont disponibles pour l'**exportation**. Vous pouvez sauvegarder vos contacts dans un fichier texte structuré - par exemple dans un format délimité par une virgule ou au format vCard - afin de les importer dans une base de données d'adresses telle que Microsoft Access.

Vous pouvez également envoyer directement les résultats dans un logiciel de gestion de contacts Microsoft Outlook (Express) et vers votre logiciel **PDA** Palm Desktop. Cet outil de productivité complète parfaitement des applications telles que les gestionnaires de contact, les bases de données ou encore les traitements de texte dont la fonction de publipostage vous permet d'imprimer des lettres, des enveloppes et des étiquettes.

NUMÉRISER DES CARTES DE VISITE

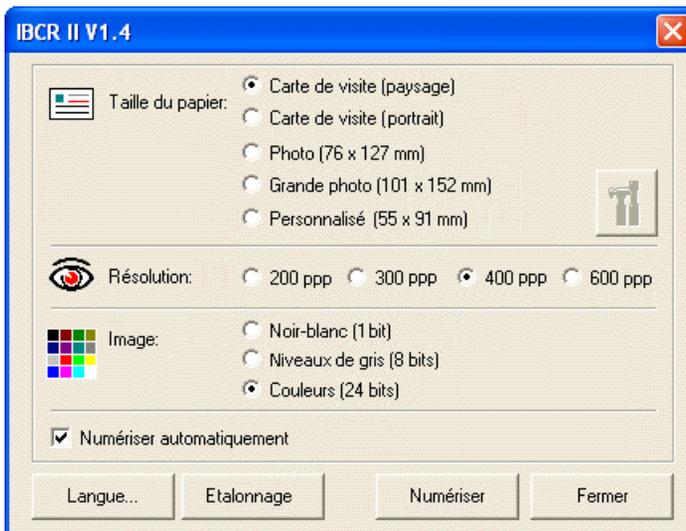
Comment fonctionne la lecture de cartes de visite? Tout d'abord, le scanner doit être installé de manière appropriée.

Si la lecture de cartes de visite est la principale application que vous faites du logiciel - vous devez par exemple numériser un très grand nombre de cartes de visite sur le stand d'une foire -, vous pouvez utiliser l'**IBCR-II**, le scanner spécialisé pour cartes de visite d'I.R.I.S. Les **numériseurs** spécialisés **pour cartes de visite** sont optimisés pour le scanning des cartes de visite et présentent de nombreux avantages par rapport aux autres types de scanners: ils permettent à l'utilisateur de ne pas devoir sélectionner le bon format de carte, de tels scanners



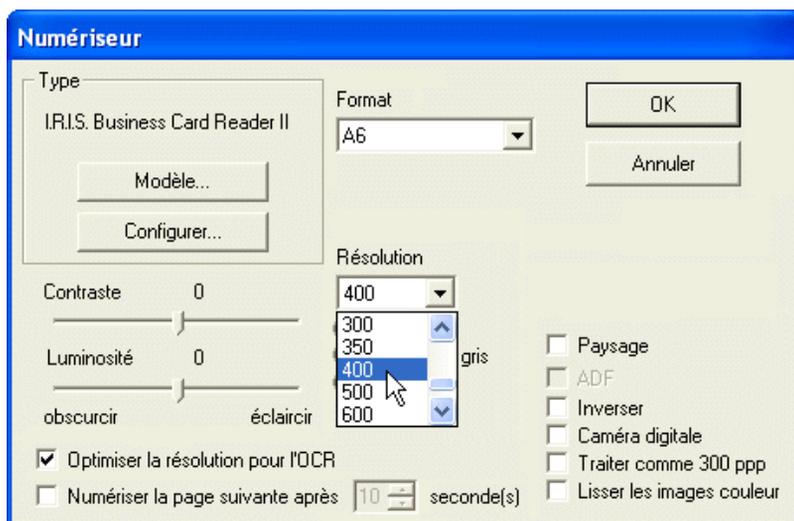
prennent peu de place sur un bureau et assurent une numérisation rapide et sans aucun problème, quels que soient le papier utilisé et la qualité de l'impression.

Servez-vous de l'interface Twain de l'IBCR-II pour activer le mode de **saisie automatique**! Cliquez sur la commande de numérisation pour activer l'option "Numériser Automatiquement" dans l'interface Twain de l'utilisateur .



Vous pouvez maintenant numériser les cartes de visite les unes après les autres: dès qu'une carte est insérée dans le scanner, la numérisation démarre...

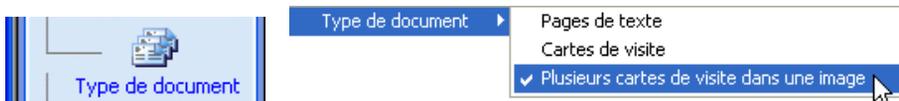
Pour reconnaître les cartes de visite avec succès, nous vous recommandons de sélectionner une **résolution** de 400 ppp.



Si vous utilisez un **scanner à plat**, vous pouvez numériser plusieurs cartes de visite à la fois sur le plateau du numériseur et les faire ensuite segmenter par le logiciel. Pour permettre à Readiris d'extraire les différentes cartes de visite, vous devez numériser les cartes avec un arrière-plan noir. Pour ce faire, nous vous suggérons une méthode très simple: laissez le couvercle du scanner ouvert quand vous numériser vos cartes!

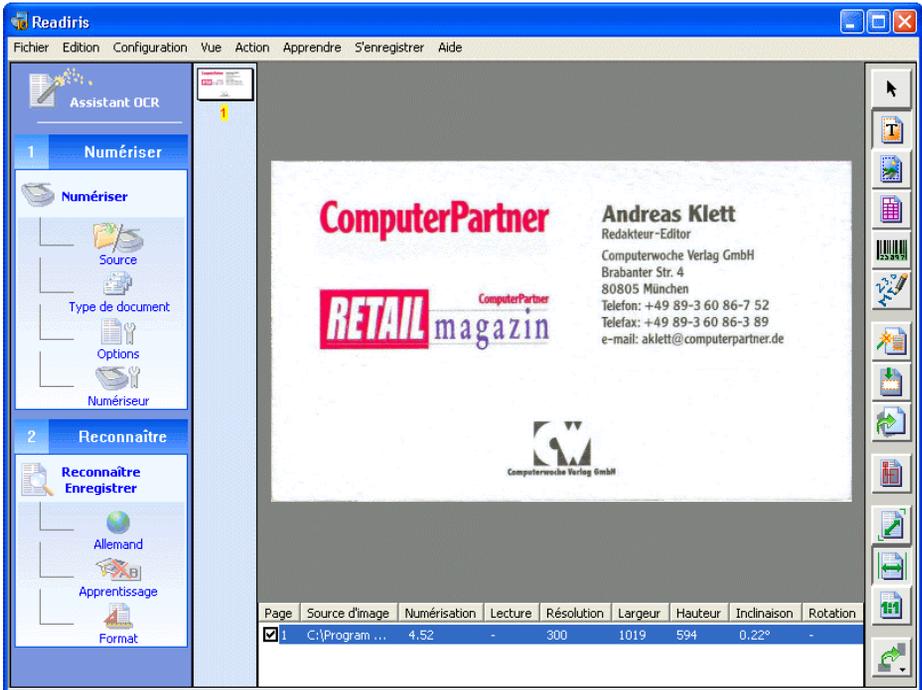


Sélectionnez l'option "Plusieurs Cartes de Visite dans une Image" avec le bouton "Type de Document" de la barre d'outils principale (ou avec la commande "Type de Document" sous le menu "Configuration") et numérisez vos cartes de visite.



Avec les scanners à plat, il suffit de cliquer une fois sur le bouton "Numériser" pour commencer la capture de l'image! (Vous pouvez également sélectionner la commande "Numériser" du menu "Action" pour numériser l'image.)

Vous ne verrez jamais apparaître sur votre écran la page entière, telle que vous l'avez numérisée: seules les cartes de visites segmentées seront affichées!



Si vous avez oublié d'activer la "fragmentation" de l'image avec le mode plusieurs cartes, cliquez sur la commande "Extraire les Cartes de Visite" sous le menu "Action" afin de segmenter la grande image en réelles images des cartes tout en vous débarrassant des bords noirs superflus.

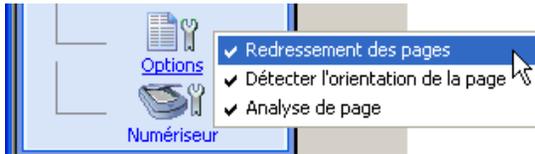
Extraire les cartes de visite



Mais quelque soit votre type scanner, certaines options sont activées de manière invisible pour optimiser le processus de lecture: le redressement des



cartes et la détection et l'orientation de la page. Cliquez sur le bouton "Options" de la barre d'outils principale pour les découvrir.



Il est pratiquement impossible que les cartes que vous avez positionnées sur le scanner soient (toutes) parfaitement droites. Laissez donc le logiciel agir à votre place... (Au cas où vous auriez désactivé cette option par accident, le bouton "Redresser la Page" est disponible sur la barre d'outils d'image. Il permet de redresser les images à posteriori, mais la tâche pourrait vite s'avérer fastidieuse si vous devez redresser chaque image manuellement!)

Il en va de même pour la détection de l'orientation. Laissez le logiciel agir à votre place, sinon vous devrez faire pivoter les cartes de visite qui ont été placées sur le scanner à plat à l'envers ou sous un angle de 90°. (Si vous désactivez cette option par accident, utilisez les boutons de rotations de la barre d'outils image...)



Redresser la page

ACTIVEZ LE MODE DE LECTURE DES CARTES DE VISITE

Lorsque les images individuelles des cartes de visite sont correctement positionnées dans la fenêtre d'image, vous devez encore activer le **mode de lecture des cartes de visite** avec le bouton "Type de Document" (ou avec la commande équivalente sous le menu "Configuration").



Si vous numérisez les cartes de visite une par une, sélectionnez l'option "Cartes de Visite"; si vous numérisez simultanément plusieurs cartes de visite sur le plateau de votre scanner, sélectionnez l'option "Plusieurs Cartes de Visite dans une Image". Cette option ne concerne pas uniquement la "fragmentation" de l'image: la sélection d'une de ces deux options est nécessaire pour la *reconnaissance* des cartes de visite! Sélectionnez "Pages de Texte" pour désactiver à nouveau le mode de lecture des cartes de visite.

Le bouton "Type de Document" confirme que le mode de lecture de la carte de visite est activé.



Vous devez activer ce mode spécial car des routines OCR spécifiques seront utilisées. Elles permettent au logiciel d'insérer les données reconnues dans des champs spécifiques de la base de données - nom de société, adresse e-mail et site Internet, etc. Lorsque Readiris reconnaît la chaîne de caractère "Jean Ricard", il "sait" que Jean est un prénom et Ricard un nom de famille. De même, le système reconnaît les titres, tel que "vice-président", "ingénieur", les villes telles que Toulouse et Lyon, les états (américains) tels que l'Oregon ou le Maryland, etc.

Ce logiciel supporte non moins de 28 pays: les cartes de visite de l'Amérique du Nord et du Sud et les cartes de visite des pays européens sont reconnues, y compris des pays de l'Europe de l'Est. (La lecture des cartes de visite de la Chine, du Japon, de Corée et de Taiwan nécessite l'option logicielle "Complément BCR asiatique". Les cartes de visite des pays asiatiques en anglais sont supportées par le logiciel "standard" Readiris Corporate!)

Dès que le mode de lecture de cartes de visite est activé, le contenu de la liste déroulante "Langue" change! Veillez à toujours sélectionner le pays correct.



En sélectionnant le pays des cartes de visite dans la barre de boutons, vous indiquez non seulement la langue du texte à reconnaître, mais aussi la présentation générale, le style de la carte de visite - s'agit-il d'une carte américaine, britannique, etc.? Chaque pays a un "style" différent, sa propre manière de présenter les informations sur les cartes de visite: les Américains composent une adresse différemment des Français, etc., les codes postaux et téléphones allemands ont une syntaxe différente des codes ou téléphones anglais, etc. (Certains styles de

carte correspondent à plusieurs langues. La Belgique et le Canada ont deux langues officielles, la Suisse en a trois! Readiris détecte la langue sélectionnée automatiquement...)



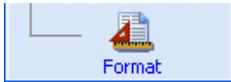
L'analyse des champs comprend une troisième tâche: Readiris effectue non seulement l'analyse mais également le *formatage* du texte reconnu. Le système filtre les données superflues, même si elles jouent un rôle actif! Si votre carte de visite mentionne l'information "Téléphone: (508) 898 42 89", le résultat de sortie sera "5088984289" dans le champ téléphone. Le mot "Téléphone" est effacé, même si Readiris l'utilise correctement pour détecter la localisation du numéro de téléphone. Les parenthèses, les traits d'union et les espaces inclus dans le numéro de téléphone sont également supprimés.



RECONNAISSANCE DES CARTES DE VISITE

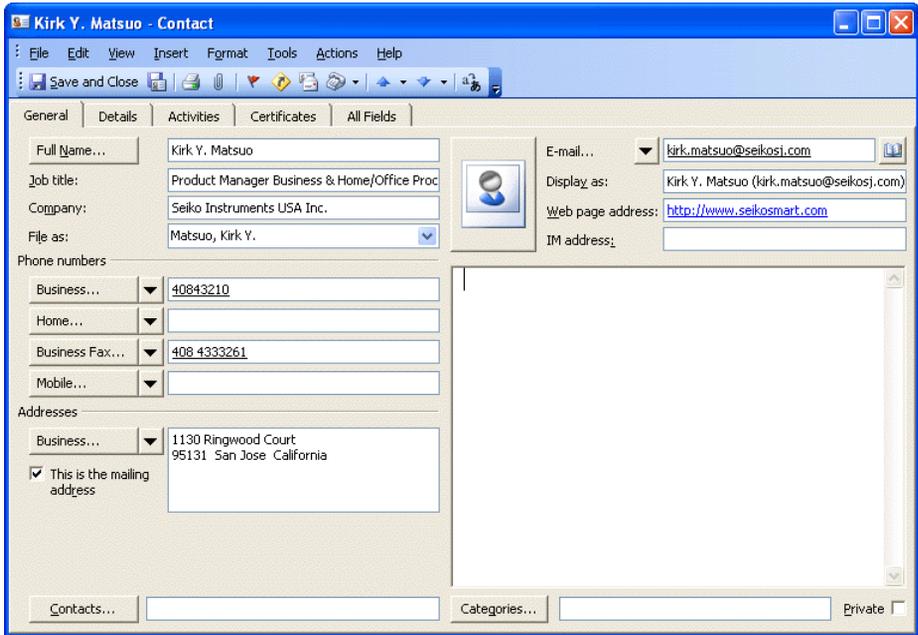
Traitez vos numérisations comme de coutume en cliquant sur le bouton "Reconnaître-Enregistrer".

Cependant, nous vous recommandons de déterminer le format de sortie avant tout. Pour cela, cliquez sur le bouton "Format".



Ces Personal Information Managers (“PIMs”) les plus répandus sont directement supportés: Microsoft Outlook (Express) et votre logiciel PDA Palm Desktop. Pour exporter les cartes vers d’autres applications, utilisez les formats “universels” vCard et les données séparées par des virgules. (Les fichiers vCard sont instantanément importés dans n’importe quelle application qui supporte le standard vCard: double-cliquez sur un fichier vCard et les contacts seront immédiatement ajoutés à votre Carnet d’Adresses Windows!)

Le texte qui en résulte - qui ne comprend ici qu’une seule adresse - ressemble par exemple à ceci lorsque vous envoyez les résultats directement vers Microsoft Outlook (Express).



OBTENIR DE L'AIDE EN LIGNE

Ceci termine l'aperçu de Readiris. Quelques informations de dernière minute peuvent ne pas se trouver dans ce manuel. Nous vous recommandons donc de consulter le système d'aide en ligne pour obtenir des renseignements complémentaires sur Readiris.

Pour ce faire, rendez-vous au menu "Aide". La commande "Rubriques d'Aide" et son raccourci clavier F1 vous permettent de naviguer au travers des nombreuses rubriques d'aide.



Aide de Readiris

Hide Back Forward Home Print Options

Contents Index Search

- Introduction à l'OCR
- Reconnaître des documents
- Reconnaître du texte manuscrit
- Reconnaître des codes à barres
- Comment faire...?
- Informations de référence
- Versions et options du logiciel
- Enregistrement du produit
- Support technique
- I.R.I.S.

Bienvenue dans l'aide de Readiris™...!

- Utilisez l'aide en ligne pour en savoir plus sur Readiris.
- Trouvez rapidement des réponses à vos questions.
- Connectez-vous au site web d'I.R.I.S. pour connaître les toutes dernières astuces et les mises à jour du produit.

©2005 Droits d'auteur [I.R.I.S.](#) Tous droits réservés

Les autres commandes du menu "Aide" vous expliquent comment obtenir le support technique, comment contacter I.R.I.S., et vous donnent un accès direct à la page d'accueil d'I.R.I.S.