

Background

package

READ IRIS OCR

yanubár

real

text

Save Time, No More Retyping!

typed

IRIS

Druckqualität

trainable

učici modul

OCR

paper

matrix

обраа



caractères accentués

omnifont

easy

recognition

fax

neue Symbole und Schriftar

IR

IMAGE RECOGNITION
INTEGRATED
SYSTEMS

JUST READ IT!

Readiris™

Manuel d'utilisation

**Logiciel intelligent de
reconnaissance optique de caractères**

**pour ordinateurs personnels IBM, PS/2 et
compatibles**

Image Recognition Integrated Systems

Readiris est une marque déposée de Image Recognition Integrated Systems

Le contenu de ce manuel est sujet à modification sans avertissement préalable et n'engage la société IRIS à aucune obligation. Le logiciel décrit dans ce document est fourni avec un contrat de licence ou un accord de confidentialité. Le logiciel ne peut être utilisé ou copié qu'en respectant les termes du contrat.

Proximity est une marque déposée de Proximity Technology Inc. IBM est une marque déposée de International Business Machines Corp. DOS et Windows sont des marques déposées de Microsoft Corp. Tous les types de scanners et formats de fichiers apparaissant dans ce manuel sont des marques déposées de leurs fabricants respectifs.

BASE DE DONNEES LINGUISTIQUE
PROXIMITY/MERRIAM-WEBSTER
Copyright 1984 Tous droits réservés
Proximity Technology Inc.
Copyright 1984 Tous droits réservés
Merriam-webster Inc.

BASE DE DONNEES LINGUISTIQUE
PROXIMITY/MERRIAM-WEBSTER
Copyright 1984 Tous droits réservés
Proximity Technology Inc.
Copyright 1984 Tous droits réservés
William Collins Sons & Co. Ltd.

L'orthographe de ce produit est basé sur
Proximity Linguistic Technology
© Copyright Image Recognition Integrated Systems 1994

Tous droits réservés
Numéro de document: READ-9405-FRN

Table des matières

Pour commencer

Ce dont vous avez besoin	7
Installation de Readiris	7
Installation de la fonction Connect	8

Pour une utilisation immédiate

Démarrer Readiris	9
Définir votre configuration par défaut	10
Faire l'OCR sur une image pré-scannée	10
Ajustements	10
Saisie de l'image	11
Agrandissement de l'image	12
Définition d'une fenêtre texte et graphique	12
Ajustage d'une fenêtre texte	13
Reconnaissance de caractères	13
Visualisation des résultats	14
Faire de l'OCR sur une nouvelle image	15
Réglages d'un numériseur	15
Saisie à l'aide d'un scanner à plat	16
Saisie à l'aide d'un scanner à main	16
Exécution de la reconnaissance	17

Pour aller plus loin avec Readiris

Ajustage des réglages par défaut	19
Le scanner à plat	20
Le scanner à main	21
Formats de sortie texte	22
Formats de sortie graphiques	23
Langue	24
Sauvegarde de la configuration par défaut	24
Sauvegarde et récupération de configurations	24
Sauvegarde d'une configuration	25
Récupérer une configuration	25
Saisie à l'aide des "Contrôles"	26

II

Document	27
Action	27
Fichiers	29
Saisie à l'aide d'un scanner à plat	30
Saisie à l'aide d'un scanner à main	31
Comment visualiser l'image	32
Contrôle de la qualité de la saisie	32
Les outils de fenêtrage	33
Création d'une fenêtre rectangulaire	33
Création d'une fenêtre polygonale	33
Sélection des fenêtres	34
Modification d'une fenêtre	34
Déplacement d'une fenêtre	35
Effacement d'une fenêtre	35
La reconnaissance de caractères	35
L'apprentissage de fontes	37
OK	37
Continue	38
Confirme	38
Annule	38
Efface	38
Fin	38
L'efficacité d'un dictionnaire de fontes	39
Aide	40
Reconnaissance des tables	40
Reconnaissance automatique	41
La fonctionnalité "Connect"	42
WordPerfect, Word et Excel	42
Autres applications Windows	43

Commandes de Readiris

Fichier	45
Visualiser fichier TIFF	45
Résumé de la configuration	45
Sélectionner source	45
Acquisition	46
Enregistrer un fenêtrage	46
Charger un fenêtrage	46
Enregistrer la configuration	46
Charger une configuration	46
Configuration par défaut	46
Quitter	46
Edition	47
Annuler	47
Couper	47
Copier	47
Coller	47
Effacer	47

Marquer tout	47
Préférences	48
Scanner	48
Format texte	48
Format graphique	48
Langue	48
Barre des outils	48
Vue	49
Page entière	49
Réduction 50%	49
Taille réelle	49
Agrandissement 200%	49
Toutes les fenêtres	49
Type sélectionné	49
Tri des fenêtres	49
Visualiser tri	50
Annuler tri	50
Rotation	50
Fonctions de fenêtrage	50
Coordonnées	50
Action	51
Contrôles	51
Saisie	51
OCR	51
Apprentissage interactif	51
Analyse de page	51
Vue après saisie	52
La boîte de contrôles	52
Espacement	52
Taille	52
Fonte	52
Page(s) à lire	52
Analyse de page	52
Vue après saisie	53
Apprentissage interactif	53
Vérification sévère	53
Source	53
Sortie	53
Fenêtrage	54
Dico	54
Saisie	54
Fermer	54
Aide	54
Contenu	54
A propos de Readiris	55

Pour en savoir plus sur l'OCR

Le processus d'OCR	57
--------------------------	----

IV

Champ d'action de Readiris	58
Nombre de caractères et contexte linguistique	59
Isolement des caractères	59
Taille de caractères	60
Les documents à exclure	60



Bienvenue

Readiris est un puissant système de Reconnaissance Optique de Caractères, technique souvent mentionnée par l'abréviation OCR (Optical Character Recognition). Cette technique fait de votre ordinateur personnel et de votre scanner un outil de lecture de documents particulièrement efficace.

Il vous fournit le moyen le meilleur et le plus rapide pour transformer des informations imprimées ou dactylographiées en fichiers texte éditables qui peuvent être exploités par vos outils de traitement de textes, de PAO, etc.

Readiris utilise les méthodes de manipulation conventionnelles des applications tournant dans l'environnement Windows: menus, boîtes de dialogue, icônes. La facilité d'utilisation est telle que vous pouvez faire démarrer Readiris à partir de votre application Windows favorite!



A propos de Readiris

Votre logiciel Readiris est très simple à installer et à utiliser. L'installation ne demande que deux minutes, la mise en route seulement cinq minutes. Cette opération se limite le plus souvent à cliquer sur la barre de boutons.

Readiris est un produit bon marché qui vous offre les techniques les plus avancées de l'intelligence artificielle utilisées dans le domaine de l'OCR.

Readiris est un système à "auto-apprentissage", c'est-à-dire qu'il peut apprendre de nouveaux signes par lui-même. Ceci est rendu possible grâce à un système d'analyse du contexte linguistique: Readiris utilise des bases de données linguistiques, les lexiques, afin de résoudre les cas difficiles de reconnaissance.

Enfin Readiris vous permet, à l'aide de votre scanner, de faire l'acquisition d'images depuis vos documents imprimés et de les importer dans votre application graphique.



A propos de ce manuel

Ce manuel décrit l'utilisation de Readiris. Le premier chapitre, "Pour commencer", donne les renseignements concernant les prérequis logiciels et matériels nécessaires et la procédure d'installation.

Le deuxième chapitre, "Pour une utilisation immédiate", fait un survol de la façon dont Readiris travaille et introduit les concepts de base.

Le troisième chapitre, "Pour aller plus loin avec Readiris", quant à lui, se propose de détailler plus en profondeur les différentes caractéristiques de Readiris.

Tous les menus et commandes sont détaillés dans le quatrième chapitre, "Les commandes de Readiris".

Le dernier chapitre enfin, "Pour en savoir plus sur l'OCR", donne une brève description des concepts de l'OCR et des paramètres entrant en jeu. Il explique l'impact de l'ajustement de ces paramètres sur les performances des systèmes OCR en général et sur celles de Readiris en particulier.

Vous pouvez trouver une information complémentaire - non fournie dans ce manuel - dans le fichier README.TXT. N'hésitez pas à consulter ce fichier si vous avez des problèmes lors de l'installation ou de la manipulation de Readiris. Pour plus d'information sur les scanners, consultez le fichier READSCAN.TXT.

1 Pour commencer

Ce chapitre explique comment installer Readiris avant la première utilisation.

Nous vous conseillons, avant d'utiliser Readiris, d'être familiarisé avec votre PC, le MS-DOS et Windows.

Ce dont vous avez besoin

Voici une liste de ce dont vous avez besoin pour installer Readiris.

- un système PC 386 100% compatible IBM-PC ou PS/2. Readiris supporte le fonctionnement multitâches, même sur un système 386!
- le système d'exploitation MS-DOS version 3.3 ou suivantes. Nous recommandons la version MS-DOS 6.x.
- 4 MB RAM. Vous pouvez avoir besoin de plus d'espace mémoire pour utiliser la fonction Connect.
- Windows version 3.1. N'utilisez pas Windows 3.1 en mode standard.

Installation de Readiris

Commencez par faire une copie de sécurité de vos disquettes Readiris et conservez-les dans un endroit sûr.

Exécutez les étapes suivantes pour installer Readiris:

- Assurez-vous que vous vous trouvez dans le gestionnaire de programme Windows. Choisissez "Exécuter" dans le menu "Fichiers", introduisez **a:setup** dans la ligne de commande et cliquez sur "OK".
- Le programme d'installation démarre. Indiquez dans quelle langue vous souhaitez utiliser Readiris et suivez les instructions affichées sur votre écran.

Quand le logiciel vous le demande, introduisez la seconde disquette.

Le groupe de programme et les icônes d'applications sont créés par le programme d'installation.

Installation de la fonction Connect

Il faut que Readiris soit correctement installé avant d'installer la fonction Connect.

Lancez Readiris, sélectionnez votre scanner et sauvegardez les réglages par défaut avant d'installer la fonction Connect. Reportez-vous au chapitre 3, "Pour aller plus loin avec Readiris", pour savoir comment procéder.

Pour WordPerfect, Word et Excel, vous pouvez installer une macro si la fonction Connect doit être utilisée. Le système d'aide en-ligne vous donnera les explications nécessaires - le menu d'aide approprié est automatiquement affiché quand vous lancez le programme d'installation de Readiris.

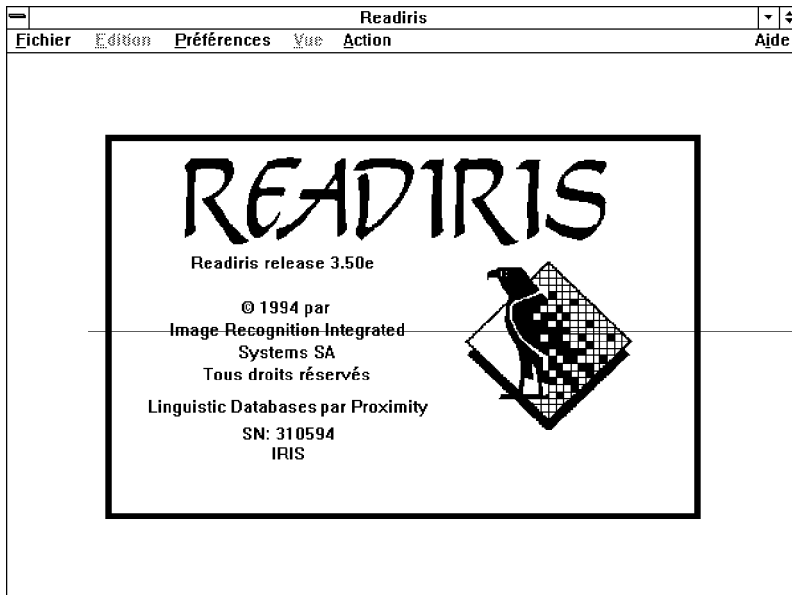
Pour toutes les autres applications Windows, un exécutable Connect a été installé par le programme d'installation Readiris. L'utilisateur n'a pas à intervenir.

2 Pour une utilisation immédiate

La meilleure façon de voir comment Readiris fonctionne, c'est de l'utiliser. Ce chapitre est destiné à ceux qui veulent commencer tout de suite. Afin d'être utilisable par ceux qui n'ont pas de scanner à disposition, une image pré-scannée, FRENCH.TIF, est fournie.

Démarrer Readiris

Assurez-vous que vous vous trouvez dans le gestionnaire de programme Windows. Cliquez deux fois sur l'icône Readiris.



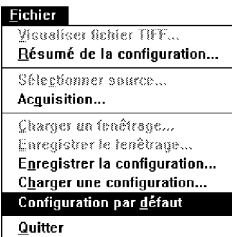
L'écran de démarrage s'affiche après quelques secondes.

La fenêtre d'application Readiris ne contient pas seulement les menus de commande, mais également une **barre de boutons**: cliquer sur ces boutons vous permettra d'exécuter de nombreuses manipulations. Cliquez sur la commande "Barre des outils" dans le menu "Préférences" pour activer ou désactiver la barre de boutons.



Définir votre configuration par défaut

- 1 Sélectionnez le menu "Préférences". Les commandes suivantes apparaissent: "Scanner", "Format Texte", "Format Graphique" et "Langue".
- 2 Examinez chaque commande et choisissez vos paramètres par défaut.
- 3 Sélectionnez la commande "Configuration par Défaut" dans le menu "fichier". Cliquez "OK" pour confirmer.



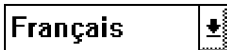
Votre système est maintenant prêt à l'emploi. Même si vous avez un scanner connecté, nous vous conseillons d'essayer d'abord le cours préparatoire.

Faire l'OCR sur une image pré-scannée

Ce bref cours préparatoire utilise une image pré-scannée d'un texte sur l'OCR. Nous allons reconnaître le texte et sauvegarder l'illustration dans un fichier graphique.

Ajustements

- 1 Vérifiez que l'option "Langues" dans la barre des boutons est bien positionnée sur français.
- 2 Dans le menu "Préférences", sélectionnez le format ASCII de l'option "Format Texte" et le format "Paintbrush" de l'option "Format Graphique".
- 3 Dans le menu "Action", cliquez sur "Contrôles".



- 4 Vérifiez si la taille de caractères et le type de fontes sont mis sur normal et le pas de caractères sur proportionnel.
- 5 Assurez-vous que le dictionnaire de fontes est READIRIS.DUS et que le dictionnaire se trouve dans le mode "Nouveau".
- 6 Désactivez l'option "Analyse de page" dans la barre de boutons, si nécessaire.
- 7 Fermez la boîte de "Contrôles" en cliquant sur "Fermer".

Action	
Contrôles...	
Saisir une Image	^S
OCR	^O
✓ Apprentissage Interactif	
Analyse de Page	
✓ Vue après Saisie	

Saisie de l'image

- 1 Activez l'option "Vue après saisie" dans le menu "Action".

Cette option est activée quand il y a une marque de vérification. Si ce n'est pas le cas, cliquez sur "Vue après saisie".

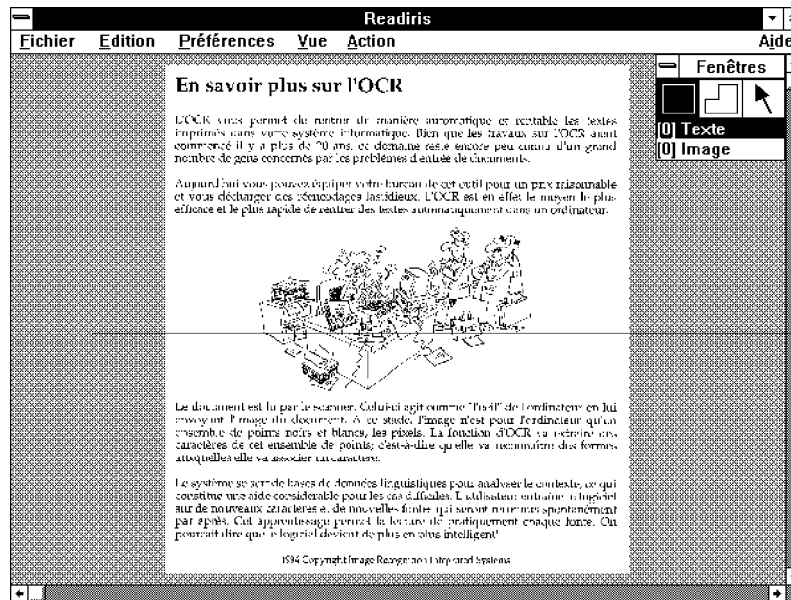
- 2 Cliquez sur le disque dans le bouton "Source" de la barre de boutons pour indiquer la source d'entrée exacte.



- 3 Cliquez sur le bouton "Saisie" dans la barre des boutons. Vous devez à présent choisir le fichier image. Sélectionnez l'image pré-scannée FRENCH.TIF installée dans le répertoire de Readiris.



- 4 L'image est alors lue depuis le disque et affichée progressivement sur votre écran. La fenêtre boîte à outils est également affichée.



Agrandissement de l'image

Vous pouvez agrandir n'importe quelle partie de l'image.

- 1 Positionnez la souris sur la partie de l'image que vous désirez agrandir.
- 2 Cliquez le bouton droit de la souris. En cliquant à nouveau sur ce bouton, vous obtiendrez l'image dans sa totalité.

Vous pouvez également utiliser les commandes du menu "Vue" ainsi que les touches d'accélération associées qui peuvent même être utilisées pendant les déplacements de la souris. Ceci peut être très utile pour sélectionner des zones de texte dispersées sur toute la page.

Définition d'une fenêtre texte et graphique

La seconde phase consiste à indiquer quels blocs de texte vous intéressent. Vous allez tracer une fenêtre texte autour du texte que

vous voulez reconnaître et une fenêtre graphique autour de l'image.

- 1 Sélectionnez l'outil rectangle en cliquant sur l'outil rectangle dans la boîte à outils. Le pointeur de la souris devient une croix.
- 2 Sélectionnez le type de fenêtre: texte.
- 3 Positionnez la souris sur le coin supérieur gauche du texte et cliquez.
- 4 Positionnez la souris sur le coin inférieur droit du texte et cliquez à nouveau.



La fenêtre est alors définie.

Définir une fenêtre image se fait de la même façon, avec cette seule différence que vous sélectionnez un type de fenêtre graphique dans la fenêtre boîte à outils.

Ajustage d'une fenêtre texte

- 1 Dans la boîte de fonctions, sélectionnez l'icône en cliquant sur l'outil "flèche". Cliquez ensuite sur un côté de la fenêtre.
- 2 Déplacez le curseur sur l'un des marqueurs.
- 3 Cliquez et maintenez le bouton gauche de la souris enfoncé.
- 4 Déplacez la souris. La fenêtre est ajustée en conséquence.



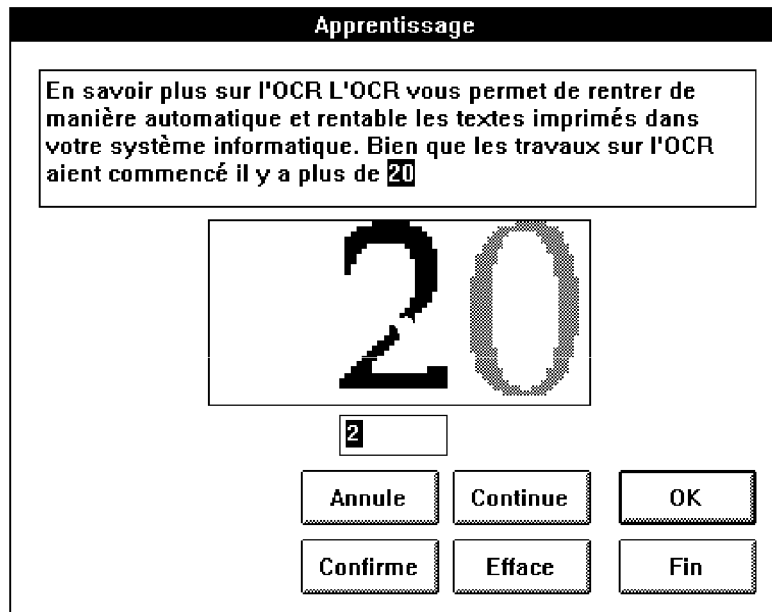
Reconnaissance de caractères

Vos fenêtres étant correctement définies, vous pouvez exécuter la reconnaissance de caractères.

- 1 Sélectionnez l'option "OCR" du menu "Action".

A la fin de la phase de reconnaissance, Readiris entre en phase d'apprentissage interactif. Il affiche le texte reconnu et s'arrête sur les caractères douteux: il affiche leur image à l'écran et propose une solution.





- 2 Cliquez sur "OK" ou pressez la touche retour chariot si la solution est correcte. Sinon tapez au clavier la solution correcte puis cliquez "OK" ou pressez la touche retour chariot. Le bouton "Continuer" permet de corriger une solution sans l'apprendre. Ignorez les autres commandes pour l'instant.

(Vous apprendrez la signification exacte de ces boutons dans le chapitre suivant).

- 3 Quittez Readiris quand vous avez fini.

Puisque le bouton "OCR" sert aussi à sauvegarder des graphiques, la zone graphique a été sauvegardée en fichier de façon automatique.

Visualisation des résultats

Maintenant votre répertoire contient le fichier texte: READIRIS.TXT et le fichier graphique READIRIS.PCX.

- 1 Exécutez par exemple le programme Notepad et ouvrez READIRIS.TXT.

Le texte reconnu est affiché à l'écran. Ce qui était à l'origine un document papier est devenu un fichier texte éditable!

- 2 Quittez Notepad.
- 3 Lancez le programme Paintbrush et ouvrez le fichier READIRIS.PCX.

Le module graphique de Windows affiche l'image. Vous pouvez lancer maintenant l'édition graphique habituelle sur cette image.

Faire de l'OCR sur une nouvelle image

Nous supposons ici que vous avez correctement installé votre scanner.

Réglages d'un numériseur

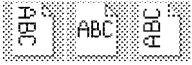
- 1 Sélectionnez l'option "Scanner" dans la barre des boutons. Une liste des scanners supportés s'affiche à l'écran.
- 2 Sélectionnez le scanner installé et cliquez OK. Les formats de pages supportés pour les scanners à plat apparaissent à l'écran, ainsi que les différentes résolutions.
- 3 Pour des raisons de reconnaissance, sélectionnez une résolution de 300 ppp ou 400 ppp si la qualité de l'image est faible.
- 4 Il se peut que vous ayez à ajuster la brillance et le contraste, à l'aide du curseur, si disponible.



Pour les scanners à main, la brillance et le contraste sont réglés à l'aide de commutateurs sur le scanner lui-même.

- 5 Indiquez la source d'entrée correcte en cliquant le scanner choisi dans la barre des boutons.

Saisie à l'aide d'un scanner à plat



- 1 Placez votre document dans votre scanner à plat.
- 2 Sélectionnez l'orientation de lecture correcte sur la barre des boutons.
- 3 Cliquez le bouton "Saisie".

La page est saisie et progressivement affichée à l'écran.

Si l'orientation de lecture choisie est paysage, l'image subira immédiatement une rotation après la saisie.

Saisie à l'aide d'un scanner à main



- 1 Sélectionnez la direction de saisie correcte dans la barre des boutons.
- 2 Cliquez le bouton "Saisie". Le curseur se transforme en un petit scanner à main.
- 3 Placez votre scanner sur la section de document à reconnaître. Votre document doit être placé sur une surface plane.
- 4 Pressez le bouton de saisie de votre scanner tout en le déplaçant dans la direction sélectionnée.

Le déplacement doit être aussi lent et régulier que possible; sans quoi l'image lue sera inexploitable pour l'OCR. L'image saisie est affichée au fur et à mesure du déplacement du scanner.

- 5 Lorsque vous avez saisi votre image, pressez la barre d'espacement pour rendre le contrôle à Readiris.

Si la direction de saisie choisie allait de la gauche vers la droite, ou de la droite vers la gauche, en pressant la barre d'espacement, vous provoquerez la rotation de l'image.

Exécution de la reconnaissance

Vous pouvez maintenant utiliser Readiris de la même façon que pour l'image pré-scannée: agrandissement, fenêtrage, reconnaissance...

3 Pour aller plus loin avec Readiris

Les pages qui suivent décrivent dans le détail les caractéristiques de Readiris. Vous allez apprendre comment contrôler la qualité de l'image lors de son acquisition. Sur l'image correctement scannée vous pourrez définir des fenêtres sur les parties du texte à "lire" par Readiris ou à faire analyser automatiquement par le système. Ensuite Readiris va effectuer l'OCR sur l'image et générer un fichier contenant le texte reconnu.

Nous étudierons également certaines caractéristiques avancées telles que la reconnaissance automatique, la fonctionnalité Connect et la saisie de tables.

Ajustage des réglages par défaut

La première chose à faire est de régler la configuration par défaut de Readiris, c'est-à-dire:

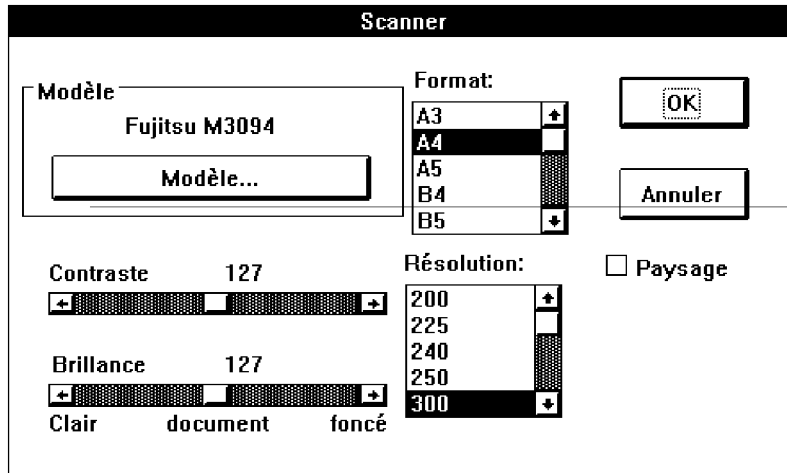
- le scanner que Readiris utilise et ses réglages
- le format des fichiers texte et graphique
- la langue des textes à lire
- les paramètres définis avec le panneau de contrôle

Ces réglages ne nécessitent pas de changements fréquents et peuvent être sauvegardés dans un fichier de configuration par défaut.

Le scanner à plat



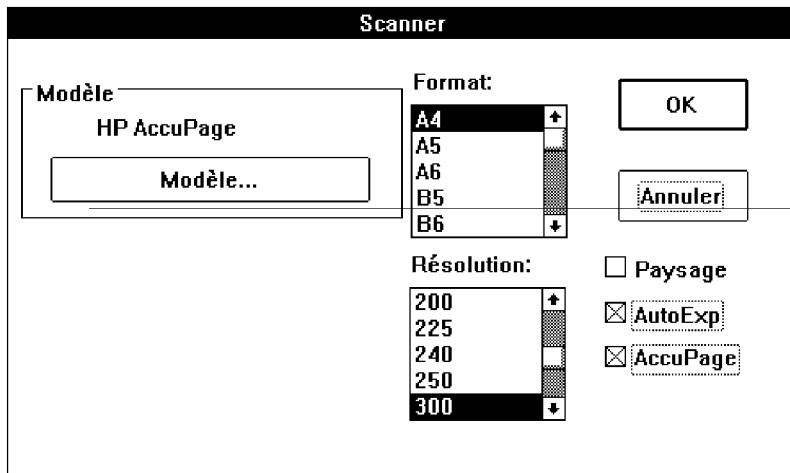
- 1 Cliquez le bouton "Scanner" de la barre des boutons.



- 2 Sélectionnez le scanner connecté à votre PC et cliquez "OK". La résolution par défaut, le format de page, la brillance, et peut-être aussi le contraste, sont affichés.
- 3 Cliquez sur le format du document correspondant aux pages à lire. A4 est le format par défaut.
- 4 Sélectionnez votre résolution de scanning. La résolution par défaut est 300 ppp.
- 5 Ajustez la brillance et, si possible, le contraste. La position du milieu est la valeur par défaut.

L'option contraste n'est peut-être pas disponible sur votre scanner. Le réglage de la brillance et du contraste permet d'adapter la lecture aux documents foncés et clairs. Les scanners HP mettent à votre disposition l'option AutoExp qui ajuste automatiquement la brillance de la saisie à votre document.

Readiris supporte également la technologie **HP AccuPage**. Cette technologie ajuste non seulement la brillance du scanner automatiquement, mais se sert également de techniques de traitement d'images avancées pour optimiser l'image scannée en fonction de la reconnaissance ultérieure.



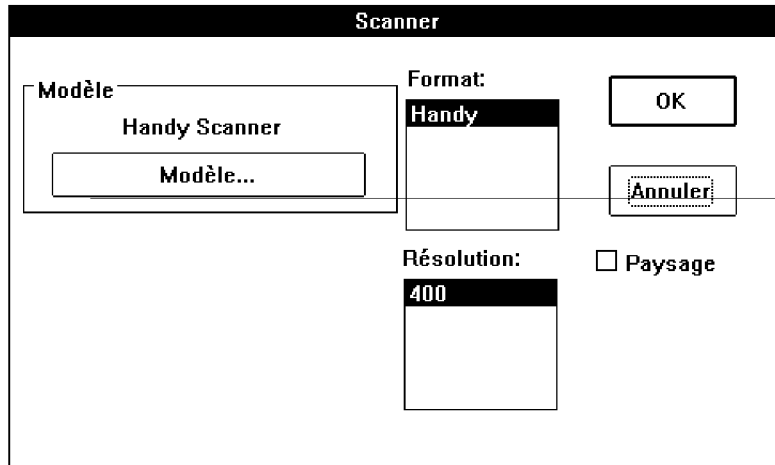
Cette technologie n'est disponible que sur les scanners HP. Sélectionnez le "scanner" AccuPage au lieu d'un scanner HP et cliquez sur l'option AccuPage.

La documentation technique jointe à votre scanner vous donnera plus de précisions sur son installation et son utilisation.

Le fichier READSCAN.TXT dans votre répertoire Readiris peut également vous donner des informations complémentaires sur l'installation d'un scanner.

Le scanner à main

- 1 Sélectionnez le bouton "Scanner" de la barre des boutons. La liste des scanners supportés s'affiche à l'écran.
- 2 Sélectionnez le scanner à main connecté à votre PC et cliquez sur "OK". La résolution utilisée est affichée sur l'écran.



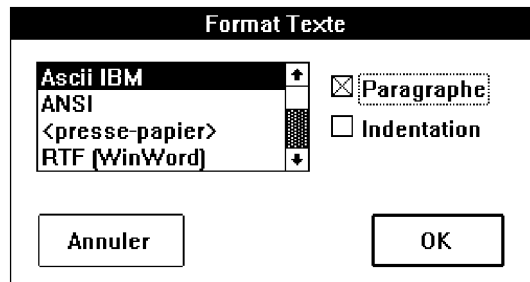
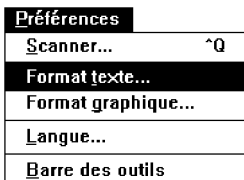
La résolution et la brillance sont réglées avec des commutateurs sur votre scanner.

La documentation technique jointe à votre scanner à main vous donnera plus de précisions sur son installation et son utilisation.

Le fichier READSCAN.TXT dans votre répertoire Readiris peut également vous donner des informations complémentaires sur l'installation d'un scanner.

Formats de sortie texte

- 1 Sélectionnez l'option "Format texte" du menu "Préférences".
- 2 Choisissez le format de sortie que vous désirez utiliser.



Readiris génère des fichiers de sortie vers un grand nombre de **traitements de texte** et de **formats texte**. Utilisez les formats ASCII ou ANSI pour d'autres traitements de texte.

Readiris génère également des tables de sortie directement vers les **feuilles de calcul** dans un format table générique.

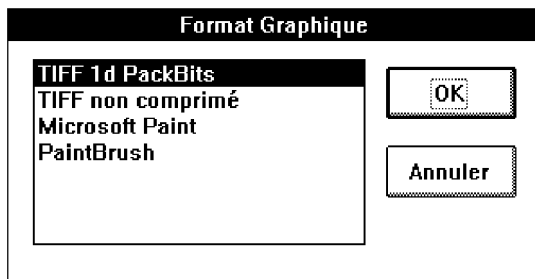
Le "format" **presse-papiers** vous permet de sauvegarder directement les résultats dans le presse-papiers et de les copier directement dans votre application Windows.

Deux options supplémentaires permettent de paramétrer le format de sortie:

- **Paragraphe**: activation de la détection automatique de paragraphes. Ceci permet de générer un texte "au kilomètre" c'est-à-dire sans code de retour chariot en fin de ligne. Seules les fins de paragraphes sont ponctuées du code retour chariot.
- **Indentation**: activation du codage des marges. Cette option est intéressante si vous désirez garder l'apparence originale du texte: marge, etc. Cependant elle est déconseillée dans le cas d'une importation dans un système de PAO.

Formats de sortie graphique

- 1 Sélectionnez l'option "Format graphique" du menu "Préférences".

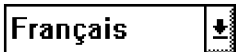


- 2 Choisissez le format graphique utilisé par votre application graphique.

Readiris sauvegardera votre graphique dans ce format graphique.

Langue

Il est important que l'utilisateur indique correctement la langue du document à reconnaître si les bases de données linguistiques doivent jouer un rôle positif dans le processus de reconnaissance.



- 1 Cliquez la liste des langues dans la barre de boutons ou choisissez la commande "Langue" du menu "Préférences".
- 2 Sélectionnez la langue que vous utiliserez le plus souvent.

Readiris est équipé de nombreux modules linguistiques qui influencent positivement le degré de reconnaissance.

Readiris peut aussi être mis en **mode numérique** pour reconnaître des tables de chiffres avec un maximum de précision et de rapidité.

Sauvegarde de la configuration par défaut

Maintenant vous pouvez sauvegarder tous ces paramètres sur disque afin de ne pas devoir les redéfinir chaque fois que vous utilisez Readiris. Vous devrez charger ces paramètres chaque fois que vous démarrerez Readiris. Vous pouvez afficher ces paramètres avec la commande "Info" du menu "Fichier".

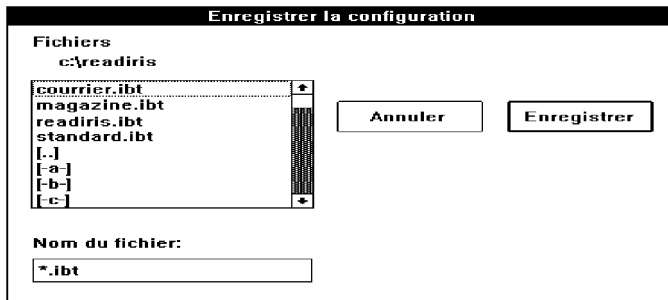
Sauvegarde et récupération de configurations

De plus, vous pouvez aussi sauvegarder des configurations adaptées à des utilisations particulières de Readiris. Ces configurations pourront être sauvegardées et récupérées à l'aide des options "Enregistrer la configuration" et "Charger une configuration" du menu "Fichier".

Sauvegarde d'une configuration

- 1 Sélectionnez l'option "Enregistrer la configuration" du menu "Fichier".

La liste des fichiers de configuration présents dans le répertoire est affichés.



- 2 Entrez au clavier un nouveau nom de fichier pour sauvegarder la configuration d'un nouveau fichier ou remplacez un fichier existant.
- 3 Cliquez sur le bouton "Enregistrer".

Votre configuration est maintenant sauvegardée dans le fichier spécifié. Ces fichiers ont l'extension par défaut *.ibt

Récupérer une configuration

- 1 Sélectionnez l'option "Charger une configuration" du menu "Fichier".

Une boîte de dialogues apparaît alors affichant une liste des fichiers de configuration présents dans le répertoire.

- 2 Sélectionnez le fichier correspondant à la configuration désirée ou sélectionnez-le dans un autre répertoire.
- 3 Cliquez sur "Chargement".

Saisie à l'aide des "Contrôles"

La boîte contrôles vous donne accès à tous les paramètres dont le réglage est nécessaire lors de la saisie du document.

Voici ces paramètres:

- la langue et qualité du document
- le type de fonte
- le pas et la taille du caractère

Vous pouvez également spécifier s'il y a un dictionnaire de fontes pour les types de caractères du document à lire.

La boîte de contrôle vous donne le contrôle de tout paramètre nécessaire pour ajuster votre logiciel aux différents types de document que l'on peut reconnaître. Elle comprend les parties suivantes:

- Document
- Fichier
- Action

Contrôles	
Document	
Espacement	
<input type="radio"/> Fixe <input checked="" type="radio"/> Proportionnel	
Taille	
<input type="radio"/> Petite <input checked="" type="radio"/> Normale	
Font	
<input checked="" type="radio"/> Normal <input type="radio"/> Dot Matrix	
Fichiers	
Source	Scanner
Sortie	Texte: (N) READIRIS.rtf Graphiques: READIRIS.TIF
Fenêtrage	Par défaut
Dico	(E) READIRIS.dus
Saisie	
1	Page(s) à lire
<input checked="" type="checkbox"/> Analyse de page	
<input checked="" type="checkbox"/> Vue après saisie	
<input checked="" type="checkbox"/> Apprentissage interactif	
<input type="checkbox"/> Strong Checking	
<input type="button" value="Saisir"/> <input type="button" value="Fermer"/>	

Document

Ici vous spécifiez les caractéristiques du document à lire:

- l'espacement: proportionnel ou fixe.
- la taille de caractères: normale ou petite
- le type de fonte: normal ou impression à aiguilles

Un **espacement fixe** signifie que tous les caractères ont la même largeur et sont espacés de manière identique, comme sur la plupart des documents produits par des machines à écrire.

Au contraire un **espacement proportionnel** tient compte de la largeur réelle du caractère qui dépend de sa forme: un "m" est plus large qu'un "i". Il en est ainsi de la plupart des livres, magazines et journaux.

La taille est considérée comme **petite** lorsque la fonte utilisée est inférieure au point 10 ou lorsque l'interligne est très fin. Dans ce cas, Readiris peut éprouver quelques difficultés à séparer correctement deux lignes consécutives si l'indicateur n'est pas positionné sur "petite".

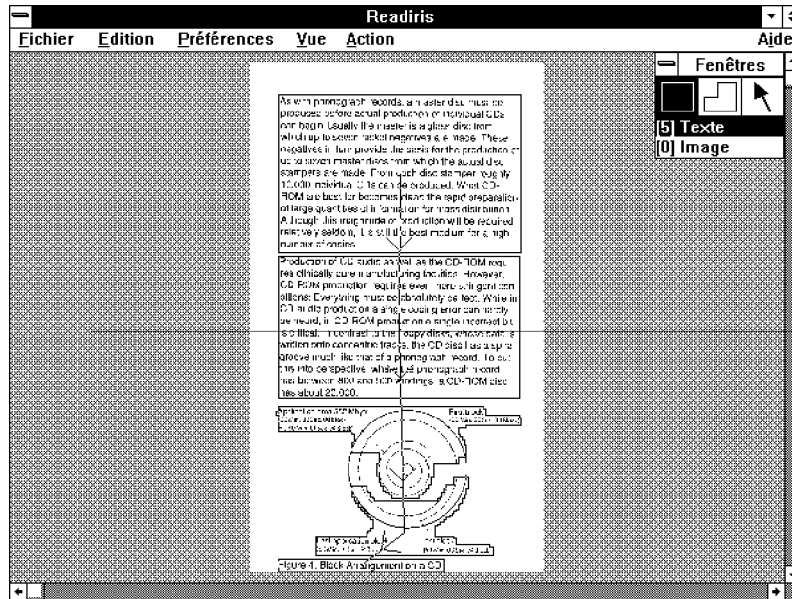
L'utilisateur peut également indiquer le type de fonte: vous ne pourrez pas reconnaître des caractères **dot matrix** sans préalablement sélectionner le paramètre correct. Les caractères dot matrix sont très dégradés et souvent simplement coupés ou composés de points individuels. Voir l'exemple ci-dessous.

ape--descended life

Action

Ici vous définissez le mode de fonctionnement de Readiris.

Quand vous activez l'**analyse de page**, les blocs de texte et de graphiques sur la page sont détectés et triés automatiquement. L'utilisateur n'est plus obligé de fenêtrer les pages numérisées manuellement, ce qui est fort intéressant pour les textes en colonnes et les documents avec une présentation complexe.



Dans un premier temps, l'analyse de page considère chaque bloc détecté comme un bloc de texte. N'intervenez pas quand il y a des graphiques sur la feuille: dans une seconde phase, le logiciel découvre spontanément qu'il n'y a pas de caractères à reconnaître. (Ces zones ne seront cependant pas sauvegardées comme graphiques).



Par défaut, les zones sont triées du haut vers le bas, de gauche à droite, mais vous pouvez modifier cet ordre en sélectionnant le bouton "Tri des Fenêtres", puis cliquez sur les marqueurs des fenêtres à reconnaître. Les fenêtres que vous ne sélectionnez pas à ce moment se seront pas reconnues!

L'analyse de page fonctionne avec rapidité et précision et est tolérante envers la pente du texte.

A l'aide de l'option "Vue après saisie" vous spécifiez si l'image doit être ou non affichée à l'écran avant l'OCR (**prévisualisation**). Pour les scanners à main, nous recommandons de laisser l'option "prévisualisation" toujours active.

Le troisième paramètre est l'**apprentissage interactif** qui permet la phase de validation pendant laquelle Readiris peut apprendre de

nouvelles fontes. Il vaut mieux désactiver cette option pour faire de la reconnaissance automatique sur de multiples pages et quand le degré de reconnaissance est excellent.

Vous pouvez également influencer le degré d'interaction. L'option "Contrôle sévère" permet un degré de vérification élevé: Readiris fera appel à l'utilisateur au moindre petit doute sans prendre la moindre décision autonome. Quand la reconnaissance est aisée, vous pouvez limiter vous-même la vérification aux formes qui constituent un réel problème.

Fichiers

Dans cette partie, vous trouverez tous les fichiers nécessaires au fonctionnement de Readiris. Il y en a quatre:

Source: définition du moyen d'entrée. Vous pouvez scanner une page ou ouvrir une image pré-scannée.

Sortie: définition du nom du fichier de sortie et de son mode d'écriture: ajout ou remplacement. Readiris peut créer deux fichiers de sortie: un fichier texte et un fichier graphique. .

Fenêtrage: définition du nom du fichier de fenêtrage - fichier par défaut, fichier vide ou fichier avec un fenêtrage spécifique. Ce sont des configurations de fenêtrage qui contiennent une structure de fenêtre définie pendant la prévisualisation du document. Outre la définition manuelle de fenêtre et l'analyse de page par le système, les fenêtrages constituent la troisième méthode pour indiquer les zones d'intérêt.

Dico: définition du nom du fichier dictionnaire et son état.

Le dictionnaire contient les informations sur les caractères que vous lui enseignez durant la phase d'apprentissage interactive.

Le dictionnaire peut avoir trois états:

- Nouveau dico: création d'un nouveau dictionnaire. Votre Readiris débute sans connaissances préalables.
- Enrichissement du dico: votre Readiris utilise les connaissances acquises lors d'utilisations précédentes et continue à apprendre les nouveaux symboles qui seront mémorisés dans le dictionnaire.

- Lecture du dico: Readiris utilise les connaissances déjà acquises mais n'apprend plus rien.

Il est important de bien choisir le mode d'utilisation car cela va influencer les performances de la reconnaissance.

Si vous n'avez qu'une page à lire, sélectionnez le mode nouveau dico. Si, au contraire, vous devez lire plusieurs pages du même type - nous entendons par même type des pages imprimées avec la même fonte et ayant la même qualité d'impression - sélectionnez le mode nouveau pour la première page, le mode enrichissement pour les deuxième et troisième pages et le mode lecture pour les pages suivantes.

Utiliser le même dictionnaire tout le temps n'a pas de sens: la taille de ces fichiers est limitée!

Saisie à l'aide d'un scanner à plat



- 1 Sélectionnez la direction de saisie correcte dans la barre des boutons (vous pouvez également activer l'option "Paysage" en cliquant sur le bouton "Scanner" dans la barre des boutons).



- 2 Cliquez sur le bouton "Scanner".

Le document est scanné immédiatement. Si vous avez activé la prévisualisation, l'image saisie se forme progressivement sur l'écran.

Si l'orientation de scanning choisie était paysage, la rotation de l'image s'effectue automatiquement. Vous pouvez également procéder à la rotation d'images après le scanning, grâce à la commande "Rotation" dans le menu "Vue".

Faites attention à l'**inclinaison des lignes**. Bien que l'analyse de page et la reconnaissance soient tolérantes envers la pente du texte, vous pouvez éprouver des difficultés à fenêtrer et reconnaître correctement une page scannée quand l'inclinaison de la ligne est trop importante.

Appuyez sur le bouton "Scanner" pour modifier les paramètres du scanner - résolution, format de page, etc.

Saisie à l'aide d'un scanner à main

Quand vous avez activé tous les paramètres nécessaires, vous êtes prêt à scanner votre document.

- 1 Sélectionnez la direction de saisie correcte dans la barre des boutons.
- 2 Cliquez sur le bouton "Scanner" dans la barre des boutons. Le curseur se transforme en un petit scanner à main.
- 3 Vous pouvez modifier la brillance et la résolution avec les commutateurs sur votre scanner à main. Utilisez une résolution de 300 dpi pour des documents de très bonne qualité, sélectionnez une résolution de 400 dpi pour des images de mauvaise qualité.
- 4 Placez votre scanner sur la section de document que vous souhaitez reconnaître.
- 5 Pressez une fois le bouton de démarrage de votre scanner et déplacez-le dans la direction sélectionnée.
- 6 L'image s'affiche progressivement à l'écran à mesure que votre scanner se déplace.
- 7 Le mouvement devra être aussi lent et régulier que possible, sinon, l'image saisie ne pourra pas être utilisée en OCR.
- 8 Les instructions du fabricant jointes à votre scanner vous donneront plus d'information sur la manière d'utiliser votre scanner.
- 9 Quand vous êtes prêt à scanner l'image, appuyez sur la barre d'espacement pour rendre le contrôle à Readiris.



Si la direction de scanning sélectionnée était de gauche vers droite, ou vice versa, vous provoquerez directement la rotation de l'image en appuyant sur la barre d'espacement. Vous pouvez également procéder à la rotation d'images après le scanning, grâce à la commande "Rotation" dans le menu "Vue".

Faites attention à l'**inclinaison des lignes**. Bien que l'analyse de page et la reconnaissance soient tolérantes envers la pente du texte, vous pouvez éprouver des difficultés à fenêtrer et reconnaître correctement une page scannée quand l'inclinaison de la ligne est trop importante.

Comment visualiser l'image

L'affichage par défaut est pratique pour avoir une vue d'ensemble du document (blocs de texte et d'images). Ce mode d'affichage est appelé "Page entière".

Cependant, pour contrôler la qualité de la saisie, vous devez avoir une visualisation plus précise. Readiris vous donne le choix entre trois modes de "vues":

Vue	
✓ Page entière	^W
Réduction à 50%	^5
Taille réelle	^1
Agrandissement 200%	^2
✓ Toutes les fenêtres	
Type sélectionné	
Tri des fenêtres	
✓ Visualiser tri	^T
Annuler tri	
Rotation...	
^R	
Fonctions de fenêtrage...	
Coordonnées...	
^C	

- Taille Réelle: donne une vue du document tel qu'il est digitalisé, c'est-à-dire un point écran pour un point image.
- Réduction 50%: donne une vue réduite à 50% de l'image.
- Agrandissement 200%: donne une vue agrandie à 200% de l'image.

Vous pouvez aussi passer de la vue "Page entière" à la vue "Taille réelle" et vice versa en cliquant sur le bouton droit de la souris.

Contrôle de la qualité de la saisie

Les performances de la reconnaissance optique de caractères dépendent énormément de la qualité du document et du scanner. C'est pourquoi une vue détaillée de l'image est nécessaire pour contrôler cette qualité.

Readiris donnera les meilleurs résultats sur des documents de bonne qualité, c'est-à-dire où les caractères sont bien formés et séparés.

Sur des documents trop clairs, les signes sont parfois interrompus; un "O" peut devenir "()". Les documents trop foncés contiennent des caractères au dessin très gras et des lettres ouvertes peuvent se fermer lors de la saisie: un "c" peut devenir "o" par exemple.

Vous pouvez améliorer cette qualité en réglant la brillance et le contraste du scanner.

Quand vous aurez rescanné l'image, vous verrez la différence sur votre écran.

Les outils de fenêtrage

Une fois le document affiché sur l'écran, l'étape suivante est d'indiquer les zones d'intérêt.

Vous pouvez procéder de différentes manières: l'analyse de page détecte et trie automatiquement les différents blocs. Vous pouvez modifier le fenêtrage résultant en utilisant le bouton "Trier". Vous pouvez également définir vos fenêtres manuellement. Enfin, vous pouvez sauvegarder des structures de fenêtrage spécifiques, appelées "fenêtrage", dans des fichiers que vous utiliserez plus tard.



Nous recommandons l'utilisation des fenêtres quand votre document est disposé en colonnes. Pour reconnaître des tableaux correctement, définissez des fenêtres autour de chaque colonne pour placer les résultats dans les cellules correspondantes.

Voici une description des outils de fenêtrage de Readiris.

Création d'une fenêtre rectangulaire

- 1 Sélectionnez la fonction rectangle en cliquant sur l'icône rectangle de la boîte "Fenêtres".
- 2 Sélectionnez le type de la fenêtre: texte ou graphique.
- 3 Positionnez la souris à un coin de la fenêtre à créer et cliquez.
- 4 Positionnez la souris au coin diagonalement opposé de la fenêtre et cliquez à nouveau.



La fenêtre est dessinée.

Des côtés de moins d'un millimètre ne sont pas autorisés.

Création d'une fenêtre polygonale

- 1 Sélectionnez la fonction polygone en cliquant sur l'icône polygonale de la boîte "Fenêtres."
- 2 Positionnez la souris à un coin du polygone à définir et cliquez.



- 3 Déplacez la souris. Un côté du polygone est dessiné et se déplace selon le mouvement de la souris.
- 4 Cliquez pour figer le premier côté.
- 5 Déplacez la souris et cliquez pour fixer chaque segment du polygone.
- 6 Cliquez le bouton droit pour fermer le polygone.

Vous ne pouvez passer à travers un côté que vous avez déjà dessiné.

Vous pouvez également dessiner des polygones en mélangeant des rectangles. Quand deux fenêtres du même type se touchent, elles se mêlent pour ne plus former qu'une seule fenêtre.



Sélection des fenêtres

- 1 Sélectionnez la fonction flèche de la boîte de fonctions.
- 2 Cliquez sur le bord d'une fenêtre.

Des marqueurs de fenêtres apparaissent à chaque coin et au milieu de chaque côté.

- 3 Pour sélectionner une autre fenêtre tout en gardant la première, pressez la touche majuscule pendant que vous sélectionnez la deuxième fenêtre.
- 4 Pour sélectionner une fenêtre et toutes celles qui y sont incluses, pressez la touche Ctrl quand vous sélectionnez la fenêtre principale.
- 5 Pour détruire la sélection, cliquez le bouton de la souris à l'extérieur des fenêtres.

Note: Vous pouvez afficher le type et les **coordonnées** d'une fenêtre en pressant la touche Alt quand vous sélectionnez la fenêtre.

Modification d'une fenêtre

- 1 Sélectionnez la fenêtre à modifier.
- 2 Positionnez la souris sur l'un des marqueurs de fenêtre.
- 3 Pressez et gardez enfoncé le bouton gauche de la souris.

- 4 Déplacez la souris.

Les bords de la fenêtre bougent en suivant les mouvements de la souris.

- 5 Relâchez le bouton lorsque la fenêtre a la taille voulue.

Déplacement d'une fenêtre

- 1 Sélectionnez la fenêtre à déplacer.
- 2 Positionnez la souris entre deux marqueurs de fenêtre.
- 3 Pressez et gardez enfoncé le bouton gauche de la souris.
- 4 Déplacez la souris.

La fenêtre bouge en suivant les mouvements de la souris.

- 5 Relâchez le bouton lorsque la fenêtre a la position voulue.

Effacement d'une fenêtre

- 1 Sélectionnez la fenêtre.
- 2 Sélectionnez l'option "Couper" ou "Effacer" du menu "Edition" selon que vous voulez ultérieurement coller ou détruire définitivement la fenêtre.

Edition	
Annuler	Alt+BkSp
Couper	Sh Del
Copier	Ins
Coller	Sh Ins
Effacer	Del
Marquer tout	*A

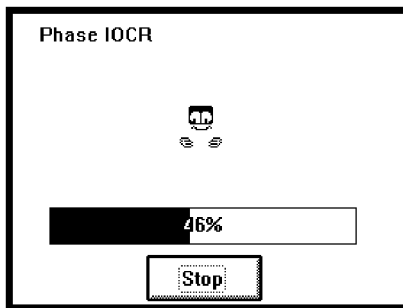
La reconnaissance de caractères

Cliquez sur le bouton "OCR" dans la barre des boutons pour lancer la reconnaissance de caractères sur les zones sélectionnées, même s'il ne s'agit que de graphiques.



La progression de la reconnaissance est indiquée; à tout moment, vous pouvez interrompre le processus en cliquant sur le bouton "Stop".

Comme Readiris supporte la **fonction multi-tâches**, vous pouvez aussi lancer la reconnaissance en arrière-plan. Minimisez la fenêtre Readiris et accomplissez une autre tâche sur votre PC. La fenêtre d'apprentissage interactif émettra un petit son pour vous avertir que le logiciel est prêt.

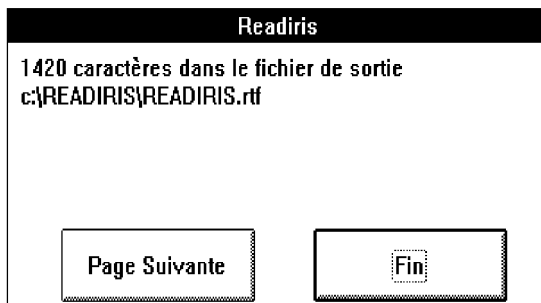


Durant cette phase interactive, le texte reconnu est progressivement affiché. Quand Readiris n'est pas certain de son choix, il affiche l'image graphique du mot et met en évidence le caractère douteux accompagné de la solution proposée puis attend votre correction éventuelle. Référez-vous à la section suivante pour plus de détails sur la manipulation des commandes de cette boîte de dialogue.

Lorsque cette phase de validation/correction est terminée, le texte reconnu est écrit dans le fichier de sortie ou dans le presse-papiers.

Readiris vous demande alors si vous désirez lire d'autres pages.

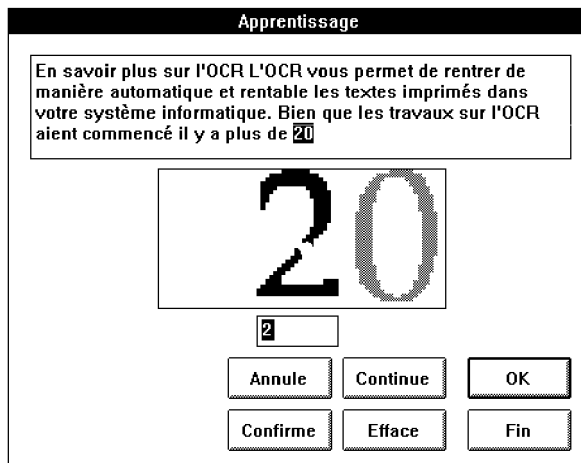
Si votre document est multi-page, le fait de cliquer sur "Page Suivante" place automatiquement les fichiers de sortie et les dictionnaires de fontes en mode d'ajout. Un scanner à plat scanne la page suivante; avec un scanner à main vous devez recommencer la saisie. Si la fonte est identique sur l'ensemble des pages de votre document, vous constaterez aisément que la phase de validation/correction vous demande beaucoup moins d'interactions pour la seconde page et les suivantes.



L'apprentissage de fontes

Comme expliqué précédemment, la dernière phase de la reconnaissance est la décision. (Le degré d'interaction dépend de l'option "Vérification sévère" dans la boîte des Contrôles).

Durant cette phase, lorsqu'un caractère n'est pas correctement reconnu, l'image du mot auquel il appartient est affichée. Le caractère non reconnu (ou la chaîne de caractères s'ils n'ont pu être séparés) est mis en évidence et la solution de l'OCR est présentée. Les caractères non reconnus sont signalés par le symbole "~".



Si nécessaire, entrez la correction et sélectionnez l'un des boutons suivants.

OK

Vous acceptez ou corrigez la solution proposée.

Readiris mémorise la solution dans le dictionnaire de fontes: toute reconnaissance future de cette forme ne nécessitera plus votre intervention.

Alternative: pressez la touche retour chariot.

Continue

Vous acceptez ou corrigez la solution proposée. A la différence du bouton "OK", Readiris garde l'information dans le dictionnaire de fontes, tout en notant qu'elle ne doit pas servir pour le reste de la reconnaissance. Toute reconnaissance future de cette forme nécessitera votre intervention.

Utilisez cette commande pour des caractères endommagés qui présentent un risque de confusion si le bouton "OK" est sélectionné.

Confirme

Les '1' et "l" ont une forme identique dans de nombreuses fontes. Dans ce cas, seul le contexte permet de distinguer l'un de l'autre. Quand cela n'est pas suffisant, ces caractères sont automatiquement présentés pour validation ou correction.

Si ces deux caractères ont des formes différentes, utilisez le bouton "Confirme". Sinon, utilisez le bouton "Continue".

Annule

A l'aide de cette commande vous pouvez revenir en arrière, ceci est utile pour corriger une erreur de frappe. Cette commande permet de revenir jusqu'à neuf fois en arrière.

Efface

Cette commande permet d'effacer le caractère du fichier de sortie. Cette commande est utile pour éliminer les points parasites interceptés par Readiris.

Fin

Cette commande met fin au processus d'apprentissage interactif. La phase de décision se termine en mode automatique. Vous acceptez chaque solution proposée par le logiciel.

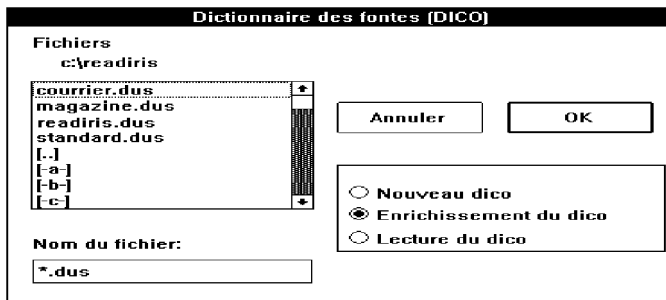
L'efficacité d'un dictionnaire de fontes

Si vous utilisez Readiris sur de nombreux documents du même type, il est intéressant de ne pas avoir à recommencer chaque fois l'apprentissage. C'est pourquoi Readiris peut sauvegarder les caractères appris dans un dictionnaire de fontes et les récupérer à chaque fois que vous utilisez un même type de document.

Pour créer ou utiliser un dictionnaire utilisateur:

- 1 Sélectionnez l'option "Contrôles" du menu "Action".
- 2 Cliquez sur "Dico".

Une boîte de dialogue affiche une liste des dictionnaires utilisateurs présents dans le répertoire. Ils ont l'extension par défaut *.dus.



- 3 Tapez un nom de fichier pour créer un nouveau dictionnaire utilisateur ou sélectionnez un dictionnaire existant en cliquant dans la liste.
- 4 Cliquez sur l'un des modes disponibles: "nouveau dico" pour créer un nouveau dictionnaire utilisateur, "enrichissement du dico" pour continuer à apprendre en utilisant un dictionnaire utilisateur existant déjà, et "lecture du dico" pour utiliser un dictionnaire utilisateur existant sans l'enrichir.
- 5 Cliquez sur "OK".

Le nom du dictionnaire choisi est affiché à côté du bouton "Dico" dans la boîte de "Contrôles". Le mode choisi est indiqué entre parenthèses.

Aide

Ce paragraphe termine notre étude générale du fonctionnement de Readiris. Readiris contient une **aide en ligne** qui donne une documentation complète sur tous les aspects du logiciel. Vous y trouverez aussi les dernières informations non comprises dans le manuel.



Sélectionnez la commande "Contenu" dans le menu "Aide" pour accéder à cette information. Court-circuit: appuyez sur F1. "A propos de Readiris" affiche l'écran initial avec vos données d'enregistrement.

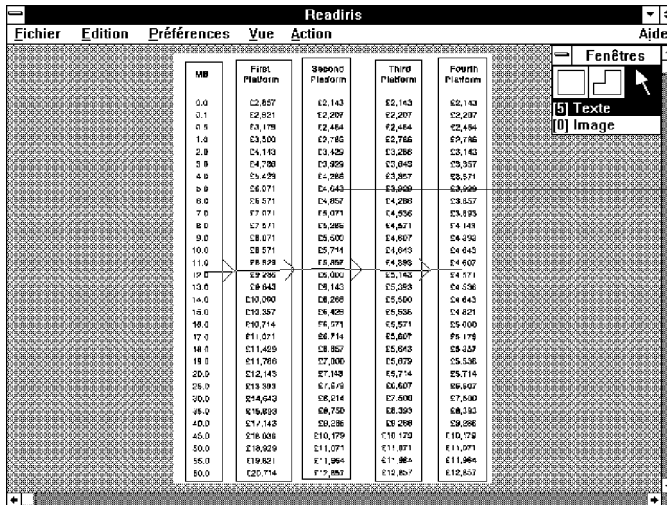
Readiris contient des **caractéristiques avancées** qui seront détaillées plus. Ainsi, nous vous expliquerons la reconnaissance des tables, la reconnaissance automatique, l'utilisation de la fonction "Connect".

Reconnaissance des tables

Vous pouvez importer des tables directement dans vos tableurs.

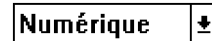
Il est obligatoire de dessiner des fenêtres autour de chaque colonne pour placer les résultats dans les cellules correspondantes. Voir exemple ci-dessous.

Sélectionnez un des deux formats de texte suivants: Excell ASCII ou Excell presse-papiers. Bien qu'ils forment l'un et l'autre un filtre de table générique, le format Excel ASCII crée un fichier table qui peut être directement importé dans votre tableur; Excel presse-papiers envoie la table reconnue dans le presse-papiers de Windows. La commande "Coller" de votre tableur insèrera la table dans votre tableur.



La détection des paragraphes et de l'indentation est désactivée quand un format table est sélectionné.

Readiris peut limiter la reconnaissance aux signes numériques quand vous faites l'OCR sur des tables de chiffres sans texte. Sélectionnez le **mode numérique** dans la liste enroulante "Langue" de la barre de boutons.



Reconnaissance automatique

Readiris exécute aussi la reconnaissance automatique sur un grand nombre de feuilles. Le degré de fonctionnement automatique dépend du type de scanner.

Les **scanners à plat** permettent la reconnaissance entièrement automatique. Désactivez les options "Vue après saisie" et "Apprentissage interactif" et activez l'option "Analyse de page" du menu "Action". (Vous pouvez aussi utiliser un fenêtrage pour indiquer les zones qui vous intéressent.) Spécifiez le nombre de pages à reconnaître dans les "Contrôles".

Mettez un paquet de feuilles dans l'alimentation du scanner et lancez la saisie en appuyant sur le bouton "Saisie". Le paquet de documents est scanné; les zones d'intérêts sont détectées et reconnues sans aucune intervention de votre part. Si l'alimentation

se vide avant la lecture du nombre de pages spécifié, un message d'erreur apparaît.

Les **scanners à main**, eux, demandent toujours une certaine intervention de l'utilisateur. La reconnaissance n'est jamais entièrement automatique. Cependant, vous pouvez rendre automatiques le fenêtrage et la reconnaissance.

La reconnaissance automatique implique la désactivation de l'apprentissage interactif. Il peut s'avérer utile de former d'abord Readiris sur quelques pages; sinon, vous courez le risque que la reconnaissance ne soit pas optimale.

La fonctionnalité "Connect"

La fonctionnalité "Connect" offre un lien direct entre le scanner et votre application Windows. Vous pouvez maintenant démarrer Readiris à partir de vos applications Windows!

Une nouvelle commande apparaît dans les menus de votre application Windows après l'installation de Readiris Connect. Cliquez sur cette commande et Readiris démarre. Vous scannez et reconnaissez le document, après quoi le résultat est importé directement dans votre application Windows, à la position actuelle du curseur.

WordPerfect, Word et Excel

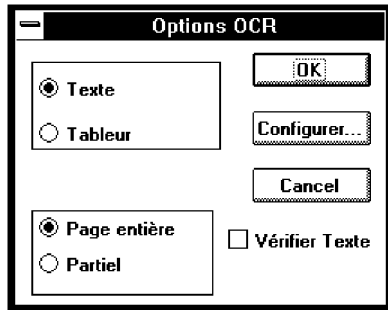
Assurez-vous que Readiris ne fonctionne pas en arrière-plan.

Cliquez sur la nouvelle commande "Saisie de texte" dans votre menu d'application pour faire démarrer Readiris à partir de votre application Windows. Pour WordPerfect 5.x, cette commande se trouve dans le menu "Macro"; pour WordPerfect 6.0 dans la commande "Macro" du menu "Outils" et pour Word et Excel dans le menu "Fichiers".

L'interface Connect peut vous poser quelques questions avant que ne démarre Readiris. Agissez comme d'habitude et quittez l'interface Connect quand vous avez fini. Les résultats de la reconnaissance sont insérés à la position actuelle du curseur.

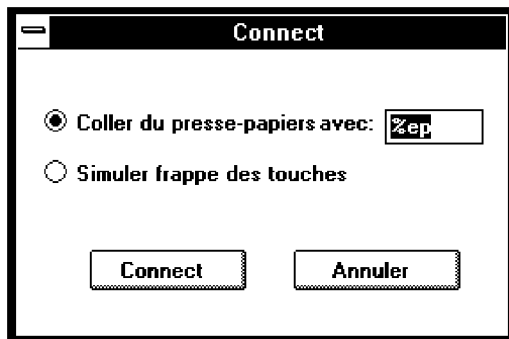
Autres applications Windows

Cliquez deux fois sur la fenêtre Connect pour lancer le programme Connect en arrière-plan avant de lancer votre application Windows. Cliquez sur la nouvelle commande "OCR" dans le menu "Fichier" de votre application. Une boîte de dialogue apparaît.



Indiquez si vous allez lire un texte ("Texte") ou une table de chiffres ("Tableur"). Sélectionnez "Page Entière" pour activer l'analyse de page ou "Partiel" pour fenêtrer le document manuellement.

Activez l'option "Vérifier texte" pour inclure l'apprentissage



interactif. La validation est par défaut désactivée.

Le bouton "Configurer" donne accès aux paramètres avancés. Sélectionnez "Simuler frappe des touches" pour insérer le texte reconnu comme si vous l'aviez introduit au clavier ou sélectionnez "Coller du presse-papiers avec" pour coller le texte invisiblement à partir du presse-papiers.

Contrôlez si la touche d'accélération correcte est spécifiée pour la commande "Coller" de votre application Windows.

Certaines touches sont représentées par un caractère spécial.

TOUCHE	REPRESENTATION
Majuscule	+
Ctrl	^
Alt	%
Insérer	{INSERT}

Quelques exemples: si la touche accélératrice de la commande Coller est CTRL+V, introduisez **^V**, si la touche accélératrice est MAJUSCULE+ INSERER, introduisez **+{INSERT}**, etc.

4 Commandes de Readiris

Ce chapitre contient une référence complète de tous les menus de Readiris. Toutes les options de ces menus sont détaillées. Vous devriez déjà connaître beaucoup d'entre elles si vous avez étudié les chapitres précédents.

Fichier

Ce menu permet de sauvegarder et récupérer des configurations et les fichiers de fenêtres. Il permet aussi d'ouvrir une image pré-scannée.

Visualiser fichier TIFF

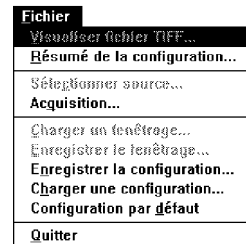
Lit un fichier image (format TIFF) et affiche son image à l'écran. Cette option n'est utilisable que pendant la phase de visualisation.

Résumé de la configuration

Affiche un récapitulatif de la configuration de Readiris.

Sélectionner source

Readiris est compatible avec le nouveau standard de scanners Twain. Cette commande vous permet de sélectionner un scanner Twain et vous donne accès aux paramètres de scanning.



Acquisition

Démarre le scanning. Cette commande équivaut à la commande "Saisie" dans le menu "Action" et au bouton "Saisie" dans la barre des boutons et les "Contrôles".

Enregistrer un fenêtrage

Sauvegarde la structure des fenêtres définies pendant la visualisation. Le fenêtrage est enregistré dans un fichier avec extension *.wdw. Cette option n'est utilisable que pendant la visualisation.

Charger un fenêtrage

Récupère un fenêtrage qui a été sauvegardé auparavant dans un fichier à extension *.wdw. Cette option n'est utilisable que pendant la visualisation.

Enregistrer la configuration

Sauvegarde la configuration de Readiris dans un fichier dont l'extension par défaut est *.ibt.

Charger une configuration

Récupère une configuration de Readiris sauvegardée auparavant dans un fichier.

Configuration par défaut

Sauvegarde la configuration de Readiris comme configuration par défaut. C'est cette configuration qui sera d'usage à chaque démarrage de Readiris.

Quitter

Quitte le programme Readiris.

Edition

Les commandes de ce menu ne sont actives que lors de la visualisation. Ces options agissent sur les fenêtres quand le document est affiché à l'écran.

Pour sélectionner une fenêtre, cliquez sur l'un des bords de la fenêtre. Si plusieurs fenêtres doivent être sélectionnées en même temps, pressez la touche Majuscule lorsque vous cliquez sur les fenêtres.

Edition	
Annuler	Alt+BkSp
Couper	Sh Del
Copier	*Ins
Coller	Sh Ins
Effacer	Del
Marquer tout ^A	

Annuler

Annule la commande précédente.

Couper

Efface la fenêtre sélectionnée et la place dans une mémoire tampon.

Copier

Copie la fenêtre placée dans la mémoire tampon.

Coller

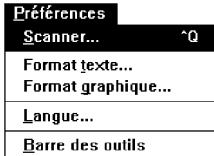
Insère le contenu placé dans la mémoire tampon. Les fenêtres insérées apparaissent à leur position d'origine. Pour les déplacer, cliquez sur l'un des côtés et, tout en gardant le bouton de la souris enfoncé, déplacez-les avec la souris.

Effacer

Efface toutes les fenêtres sélectionnées.

Marquer tout

Sélectionne toutes les fenêtres.



Préférences

Spécifie les réglages de base de Readiris: le format texte et graphique utilisé, la langue du document à lire. Active la barre des boutons.

Scanner

Permet de sélectionner le scanner et ses réglages.

Format texte

Permet de sélectionner le format du fichier pour les caractères reconnus. Permet d'activer la détection de paragraphe et d'indentation. Quand vous reconnaissez des tables, vous devez sélectionner le format Excel.

Format graphique

Permet de sélectionner le format graphique. Vérifiez quels formats de fichiers peut lire votre application graphique.

Langue

Permet de sélectionner la langue dans laquelle est écrit le document à traiter. Readiris a plusieurs bases de données linguistiques. Vous pouvez également limiter la reconnaissance de caractères aux symboles numériques, par ex. pour reconnaître des tables numériques qui ne contiennent pas de texte, avec un maximum de précision et de vitesse.

Barre des outils

Active la barre des boutons. La barre des boutons est affichée sur le côté gauche de l'écran. Nous vous conseillons d'utiliser cette barre des boutons. Les boutons donnent un accès direct aux fonctions de menus les plus utilisées.

Vue

Les commandes de ce menu ne sont utilisables que lors de la visualisation.

Vue	
√ Page entière	^W
Réduction à 50%	^5
Taille réelle	^1
Agrandissement 200%	^2
√ Toutes les fenêtres	
Type sélectionné	
Tri des fenêtres	^T
√ Visualiser tri	
Annuler tri	
Rotation...	
	^R
Fonctions de fenêtrage...	
Coordonnées...	^C

Page entière

Permet de visualiser la totalité de la page sur l'écran.

Réduction 50%

Visualisation de la page avec un facteur de réduction de 50%.

Taille réelle

Visualisation de la page en taille réelle, c'est-à-dire un point écran pour un point image. Vous pouvez passer du mode "Page entière" au mode "Taille réelle" et vice versa en cliquant sur le bouton droit de la souris.

Agrandissement 200%

Donne une vue agrandie à 200% de l'image.

Toutes les fenêtres

Affiche toutes les fenêtres définies sur le document.

Type sélectionné

Affiche seulement les fenêtres sélectionnées.

Tri des fenêtres

Indique à Readiris dans quel ordre les fenêtres texte doivent être reconnues et placées dans le fichier de sortie. Cette commande est utilisée, entre autres, pour la lecture de documents

multi-colonnes. Cliquez sur le bord de chaque fenêtre en suivant l'ordre dans lequel vous désirez que les blocs de texte soient lus.

Visualiser tri

Affiche la séquence logique des fenêtres définie lors du tri des fenêtres. Cette séquence est représentée par des flèches joignant les fenêtres.

Annuler tri

Annule le tri défini précédemment.

Alternative: utilisez le bouton "Tri" dans la barre des boutons.

Rotation

Permet d'effectuer la rotation de l'image du document. Lorsque vous activez cette commande, une nouvelle fenêtre vous demande de choisir le sens de rotation (sens horlogique ou anti-horlogique). Nous vous recommandons de scanner dans le sens correct de la saisie.

Fonctions de fenêtrage

Affiche la boîte de fonctions de fenêtrage permettant de choisir le type de fenêtres (texte ou graphique) et les fonctions (rectangle, polygone ou flèche).

Coordonnées

Affiche une fenêtre contenant les coordonnées de la souris en pixels document.

Action

Contient les commandes liées au processus OCR: saisie, reconnaissance et tous les paramètres concernant ces opérations.

Action	
Contrôles...	~P
Saisir une Image	~S
OCR	~O
<input checked="" type="checkbox"/> Apprentissage Interactif <input checked="" type="checkbox"/> Analyse de Page <input checked="" type="checkbox"/> Vue après Saisie	

Contrôles

Affiche la boîte "Contrôles".

Saise

Procède à la saisie d'un document ou ouvre une image pré-scannée, selon la source d'entrée sélectionnée.

OCR

Démarre la reconnaissance optique de caractères et sauvegarde les graphiques. Cette option est équivalente au bouton "OCR" dans la barre des boutons.

Apprentissage interactif

Active l'apprentissage interactif. Détermine si vous vérifiez les caractères douteux ou si vous acceptez les solutions proposées par le système. (L'option "Vérification sévère" dans les "Contrôles" détermine le degré de vérification voulu par l'utilisateur).

Le mode d'apprentissage est aussi disponible sur la barre des boutons. Désactivez ce mode pour exécuter de l'OCR automatique.

Analyse de page

Active l'analyse de page. Détermine si la saisie est suivie ou non par la décomposition automatique de page.

Vous pouvez aussi activer ou non l'analyse de page dans la barre des boutons. Activez cette option pour obtenir une reconnaissance automatique.

Vue après saisie

Active le mode de visualisation qui permet d'afficher l'image du document sur l'écran et de définir les fenêtres texte et graphique à traiter. Désactiver le mode de prévisualisation n'a aucun sens pour les scanners à main.

La boîte de contrôles

Toutes les spécifications de la boîte de contrôles sont décrites en détail dans le chapitre 3, "Pour aller plus loin avec Readiris".

Espacement

Sélectionne l'espacement de la fonte: fixe ou proportionnel.

Taille

Sélectionne la taille des caractères: petite ou normale. Une fonte est dite petite si sa taille est inférieure au point 10 ou si Readiris ne réussit pas à séparer deux lignes adjacentes.

Fonte

Indique la qualité d'impression de votre texte: vous lirez des fontes "normales" ou des textes imprimés en dot matrix.

Page(s) à lire

Indique le nombre de pages à saisir. Cette option n'est pas disponible pour les scanners à main.

Analyse de page

Active l'analyse de page. Détermine si la saisie est suivie ou non par la décomposition automatique de page.

Vous pouvez aussi activer l'analyse de page dans la barre des boutons. Activez cette option pour obtenir une reconnaissance automatique.

Vue après saisie

Active la visualisation de l'image du document après sa saisie. Cette commande est équivalente à l'option "Vue après saisie" du menu "Action".

Apprentissage interactif

Active l'apprentissage interactif. Détermine si vous vérifiez les caractères douteux ou si vous acceptez les solutions proposées par le système.

Le mode d'apprentissage est aussi accessible depuis la barre des boutons. Désactivez cette option pour obtenir une reconnaissance automatique. L'option suivante détermine le degré d'interaction.

Vérification sévère

Quand vous activez cette option, le degré de "rejet" est élevé: Readiris fera intervenir l'utilisateur pour tout caractère douteux au lieu d'agir de façon autonome. Quand le texte à reconnaître est de bonne qualité, vous pouvez vous en tenir à la seule vérification de ces formes réellement douteuses.

Source

Sélectionne le mode d'entrée soit depuis le scanner, soit depuis un fichier image stocké sur disque. Equivaut au bouton "Source" dans la barre des boutons.

Sortie

Définit le nom du fichier de sortie. Si ce fichier existe déjà, vous devez spécifier son mode d'écriture (ajout ou remplacement). L'extension du fichier sera automatiquement ajoutée en fonction du format que vous avez choisi.

Fenêtrage

Sélectionne le fichier contenant la structure de fenêtres adaptée au document à lire. L'extension par défaut est *.wdw.

Dico

Sélectionne le dictionnaire de fontes utilisateur et son mode d'utilisation (nouveau, enrichissement et lecture). L'extension par défaut est *.dus.

Saisie

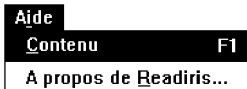
Procède à la saisie du document. Equivaut au bouton "Saisie" dans la barre des boutons.

Fermer

Ferme la boîte de contrôles.

Aide

Ce menu contient une aide en ligne et la fenêtre "A propos de".



Contenu

Donne accès à un système d'aide en ligne qui fournit une documentation complète sur tous les aspects du logiciel. Cette aide peut aussi contenir des informations de dernière minute.

Court-circuit: appuyez sur F1.

L'information peut non seulement être consultée mais aussi imprimée. Votre manuel de Windows vous dira comment faire.

A propos de Readiris

Affiche l'écran initial avec les données d'enregistrement.

5 Pour en savoir plus sur l'OCR

L'OCR vous permet de rentrer de manière automatique et rentable les textes imprimés dans votre système informatique. Bien que les travaux sur l'OCR aient commencé il y a plus de 20 ans, ce domaine reste encore peu connu d'un grand nombre de gens, qui pourraient, et devraient, être concernés par les problèmes d'entrée de documents.

Aujourd'hui vous pouvez équiper votre bureau de cet outil pour un prix raisonnable et vous décharger des réencodages fastidieux.

Les produits OCR peuvent lire des documents très variés. Cependant, ils ne peuvent pas lire n'importe quoi. Les pages suivantes donnent un bref aperçu des procédés de l'OCR ainsi que les spécifications de Readiris.

Le processus d'OCR

Le document à lire est placé sur le scanner. Celui-ci agit comme l'oeil de l'ordinateur en lui envoyant l'image du document. A ce stade l'image n'est pour l'ordinateur qu'un ensemble de points noirs et blancs. La fonction d'OCR va extraire de cet ensemble de points des caractères; c'est-à-dire qu'elle va reconnaître des formes auxquelles elle va associer une information symbolique: un nom.

Il existe deux types de systèmes d'OCR: les systèmes "automatiques" et les systèmes "à apprentissage". Les premiers peuvent reconnaître des caractères parmi une grande panoplie de fontes, mais pas tout. Les derniers peuvent apprendre chaque nouveau symbole rencontré. Cette méthode demande une

intervention de l'utilisateur mais permet de traiter virtuellement toutes les fontes.

Les deux systèmes utilisent des procédés de reconnaissance équivalents dont les principales étapes sont:

- la segmentation des lignes
- la segmentation de mots et caractères
- la reconnaissance des caractères
- la sauvegarde des résultats dans un fichier

La **segmentation des lignes** consiste à isoler chaque ligne du texte. Cette étape analyse la pente du texte ainsi que la distance d'interligne. Les textes fortement penchés ou ayant un interligne très faible sont les plus difficiles à traiter.

La **segmentation des mots et caractères** consiste à isoler chaque mot de la ligne et chaque caractère du mot. Si chaque caractère occupe la même place sur la ligne, le travail est très facile (texte imprimé à l'aide d'une machine à écrire).

Le problème est plus intéressant lorsque les caractères occupent une place proportionnelle à leur taille (espacement proportionnel). Les cas les plus difficiles sont les fontes qui génèrent des caractères qui se touchent aléatoirement. Les caractères en dot matrix sont composés de petits points et requièrent des techniques de segmentation hautement spécifiques.

La **reconnaissance de caractères** extrait les caractéristiques de chaque forme isolée et lui associe un nom.

La **sauvegarde des résultats** génère un fichier contenant le texte reconnu avec un codage permettant de l'importer dans un système de traitement de texte, feuille de calcul ou de PAO.

Champ d'action de Readiris

Readiris est un produit d'OCR de haute performance facile à utiliser. Ses fonctionnalités sont issues du savoir-faire d'I.R.I.S. en matière d'OCR. Ses spécifications sont les suivantes:

- Il utilise les techniques d'**apprentissage automatique**, ce qui en fait un système ouvert. Cela signifie que Readiris apprend par lui-même la plus grande partie des caractères et peut le faire de manière automatique : le logiciel trouve de façon autonome la bonne solution pour la plus grande partie du document.
- Il utilise le contexte linguistique pour trouver les solutions. Cependant il peut travailler sans contexte pour traiter les tables numériques par exemple.
- Il guide l'utilisateur à travers un **processus en 5 étapes** : saisie, visualisation, fenêtrage, reconnaissance et validation de l'apprentissage automatique.

Cependant, comme beaucoup de systèmes d'OCR, Readiris ne peut pas tout lire, mais ce qu'il peut reconnaître, est bien reconnu. Ce système fiable et bon marché est supposé lire des documents originaux ou des copies d'excellente qualité.

La méthode d'apprentissage automatique est très efficace et donne les meilleurs résultats dans les conditions suivantes.

Nombre de caractères et contexte linguistique

Readiris a besoin d'un nombre raisonnable de caractères et du contexte linguistique pour chaque fonte afin que le processus d'auto-apprentissage soit rentabilisé. Si le contexte automatique est pauvre ou inexistant, Readiris demandera plus d'interventions de l'utilisateur.

Isolement des caractères

Afin de réduire la taille mémoire nécessaire et le temps de traitement, Readiris utilise des algorithmes de segmentation limités.

Aussi Readiris ne traite que les caractères raisonnablement bien séparés. Cependant, dans un second temps, Readiris utilise de puissants algorithmes afin de décomposer les caractères crénelés et les ligatures.

Ceci considéré, Readiris a besoin d'un minimum de caractères isolés afin de procéder à une segmentation correcte des caractères liés. La conséquence immédiate de ceci est que la reconnaissance de toute fonte est possible pourvu que Readiris arrive à isoler suffisamment de caractères élémentaires.

Taille de caractères

Le champ des tailles de caractères va du point 8 au point 20, ce qui couvre la majorité des tailles utilisées pour la plupart des corps de textes. Des fontes plus petites sont lisibles si vous disposez d'un scanner proposant une résolution vraie de 400 ppp.

Nous suggérons, d'une manière générale de ne pas reconnaître les parties titres imprimées en grande taille en même temps que le corps même du texte qui est imprimé dans une taille standard. En général les titres ne représentent que quelques mots que l'utilisateur encode lui-même.

Documents à exclure

Readiris ne peut pas lire les documents trop denses, c'est-à-dire où presque tous les caractères se touchent. Ceci explique l'exclusion des documents de mauvaise qualité du type photocopie très foncée.

Index

A

Action	51
Agrandissement	12, 49
Analyse du contexte	38
Annuler	38, 47
Annuler tri	50
Apprentissage	37, 57
Annuler	38
Confirmer	38
Continuer	38
Effacer	38
Fin	38
OK	37
Auto-apprentissage	59

B

Backup	7
Boîte de contrôles	52
Boîte de fonctions	
Icône	13
Brillance	20

C

Caractère	
Douteux	13
Reconnaissance	35
Colonnes	49
Commandes	
Agrandissement	49
Annuler	47
Annuler tri	50
Charger un fenêtrage	46
Charger une configuration	46
Contrôle	26
Contrôles	51 - 52
Coordonnées	50

Copier	47
Couper	47
Edition	35
Effacer	47
Enregistrer la configuration	46
Enregistrer un fenêtrage	46
Fenêtres sélectionnées	49
Format graphique	48
Format texte	48
Langue	24, 48
Marquer tout	47
OCR	13, 51
Page entière	49
Page suivante	36
Réduction	49
Résumé de la configuration	45
Saisir une image	51
Scanner	48
Taille Réelle	49
Toutes les fenêtres	49
Tri des fenêtres	49
Visualiser	45
Visualiser tri	50
Vue après saisie	52
Configuration	
Défaut	19
Récupération	24 - 25, 46
Réglages	15
Résumé	45
Sauvegarde	24 - 25, 46
Configuration par défaut	46
Ajustage	19
Sauvegarde	24
Confirmer	38
Contexte linguistique	59
Continuer	38
Contrôles	26, 51 - 52
Action	27
Dico	54
Document	27
Espacement	52
Fenêtrage	54

Fichier d'entrée	29	.DUS	54
Fichier de fenêtrage	29	.WDW	46
Fichier de sortie	29		
Fichier dictionnaire	29		
Fichiers	29		
Page(s) à lire	52		
Saisie	54		
Sortie	53		
Source	53		
Taille	52		
Vue après saisie	53		
Coordonnées	50		
Copier	47		
Couper	47		

D

Détection			
Indentation	23, 48		
Paragraphe	23, 48		
Dico	39		
Dictionnaire	54		
Création	39		
Enrichissement	39		
Etat	29, 39		
Lecture	30, 39		
Nouveau	29		
Sélection	39		
Utilisateur	39		
Document	27		
A exclure	60		
Clair	32		
Foncé	32		
multi-colonne	49		
Multi-page	36		
Paysage	50		
Rotation	30, 50		
Saisie	11, 16, 51		
Type	30		
DOS			
Version	7		

E

Edition	47		
Effacer	38, 47		
Espacement	27, 52		
Fixe	27, 58		
Proportionnel	27, 58		
Extension			

F

Fenêtrage	29, 54		
Récupération	46		
Sauvegarde	46		
Fenêtre			
Effacement	35		
Fenêtres	33, 47, 50		
Ajustage	13		
Coordonnées	34		
Définition	12, 33		
Déplacement	35		
Fenêtre Texte	12		
Marqueurs	34 - 35		
Modification	34		
Polygone	33		
Rectangle	33		
Sélection	34, 49		
Type	33 - 34		
Fichier	45		
Dico	54		
Entrée	53		
Fenêtrage	54		
Sortie	53, 58		
Fonctions de fenêtrage	50		
Fonte			
Apprentissage	37		
Espacement	52		
Taille	27, 52, 60		
Format	19		
Graphique	23, 48		
Presse-papier	23		
Texte	22, 48, 58		
TIFF	9		

I

Image			
Agrandissement	12		
Qualité	32		
Rotation	30, 50		
Sur disque	11		
Visualisation	32		
Indentation	48		
Installation	7		

L

Langue 19, 24, 48

M

Marquer tout 47

Mémoire

 Mémoire vive 7

Menu

 Action 51

 Edition 47

 Fichier 45

 Préférences 48

 Vue 49

Microsoft

 Windows 7

Multi-colonne 49

Commandes 45 - 47, 49 - 55

Commencer 7

Configuration 45 - 46

Disquettes 9

Fonctions de fenêtrage 33

Installation 7

Mode de fonctionnement 27

Reconnaissance 13

Utilisation immédiate 9

Reconnaissance 10, 57 - 58

 Automatique 57

 Solution 14

 Sur nouvelle page 15

 Sur plusieurs pages 30, 36

Réduction 49

Résolution 20

 400 PPP 60

ROC 35

Rotation 30, 50

O

OCR 15, 51, 57 - 58

P

Page 52

 Entière 49

 Multiple 36

 Rotation 50

 Saisie 26, 30, 51

 Suivante 36

Paragraphe 48

 Détection 23

 Indentation 48

Paramètres 26

Paysage 50

Points 57

Pour en savoir plus sur l'OCR 57 - 60

Préférences 48

Presse-papier 23

Processus

 Interruption 35

R

Readiris

 Apprentissage interactif 13

 Auto-apprentissage 13

 Champ d'action 58

S

Saisie 11, 51, 54

Scanner 48

 A main 16, 21, 31

 A plat 20, 30

 Brillance 20, 32

 Résolution 20

 Saisie 16

Segmentation

 De caractères 58

 De lignes 58

 De mots 58

Sortie 29, 53

Source 29, 53

T

Taille 52

 Petite 27

 Réelle 49

TIFF 9, 45

Toutes les fenêtres 49

Tri 49 - 50

U

Utilisation immédiate 9

V

Visualiser	45, 52 - 53
Tri	50
Vue	32, 49
Page entière	32
Taille réelle	32

W

Windows	
Version	7