Readiris™ Pro 12

Guide utilisateur

Table des matières

Copyrights 1

Chapitre 1 Présentation de Readiris 3

Gain de temps et élimination du travail de réencodage 3

La famille Readiris 6

Chapitre 2 Installation de Readiris 11

Configuration requise 11

Installation du logiciel 11

Désinstallation du logiciel 13

Enregistrement du logiciel 13

Assistance produits 14

Chapitre 3 Mise en route 15

Exécution de Readiris 15

Utilisation de l'Assistant OCR 15

Interface utilisateur 17

Modification de la langue de l'interface utilisateur 20

Chapitre 4 Readiris SmartTasks 21

Utilisation des SmartTasks Readiris 21

Chapitre 5 Numérisation des documents 27

Sélection des options 27

Ouverture de fichiers image 28

Numérisation de documents papier 30

Chapitre 6 Optimisation des documents numérisés 37

Chapitre 7 Enregistrement des documents en tant que fichiers image 41

Chapitre 8 Fenêtrage des documents 43

Fenêtrage automatique des documents 43

Fenêtrage manuel des documents 45

Utilisation des modèles de fenêtrage 50

Chapitre 9 Reconnaissance des documents 53

Introduction 53

Sélection de la langue du document 54

Définition des caractéristiques du document 57

Utilisation de l'apprentissage interactif 59

Utilisation des dictionnaires de polices 61

Chapitre 10 Formatage et enregistrement de documents 65

Formatage de documents 65

Formatage de documents texte 67

Formatage de documents comportant des tableaux 71

Création de documents PDF 75

Sélection des options PDF 76

Documents PDF compressés avec iHQC 78

Adaptation des documents PDF 79

Création de documents XPS 80

Sélection des options XPS 81

Documents XPS compressés avec iHQC 82

Sélection des options des graphiques 83

Chapitre 11 Enregistrement et chargement des paramètres 87

Chapitre 12 Reconnaissance des documents de plusieurs pages 89

Ouverture et reconnaissance de plusieurs fichiers image 89

Numérisation et reconnaissance des documents de plusieurs pages 91

Modification des documents de plusieurs pages 92

Chapitre 13 Reconnaissance de texte manuscrit..... 95

Règles d'écriture..... 96

Chapitre 14 Reconnaissance des codes à barres 99

Index 103

Copyrights

ReadirisPro12-dgi-090211-02

Copyrights © 1987-2009 I.R.I.S. Tous droits réservés.

I.R.I.S. détient les droits d'auteur couvrant le logiciel Readiris, le système d'aide en ligne et la présente publication.

Les informations figurant dans ce document sont la propriété d'I.R.I.S. Elles sont sujettes à modification sans préavis et ne représentent pas un engagement de la part d'I.R.I.S. Le logiciel décrit dans le présent document est fourni sous accord de licence stipulant les conditions d'utilisation de ce produit. Le logiciel peut être utilisé ou copié uniquement conformément aux conditions de cet accord. Aucune partie de la présente publication ne peut être reproduite, transmise, stockée dans un système de récupération ou traduite dans une autre langue sans l'autorisation écrite préalable d'I.R.I.S.

Ce guide utilisateur utilise des noms fictifs à des fins de démonstration. Les références à des personnes, sociétés ou organisations réelles sont strictement fortuites.

Marques commerciales

Le logo Readiris et Readiris sont des marques commerciales d'Image Recognition Integrated Systems S.A.
Technologies d'OCR, d'ICR et de codes à barre par I.R.I.S.
Technologies AutoFormat et Linguistic par I.R.I.S.
Technologies de BCR et d'analyse de champs par I.R.I.S.
Technologie de compression iHQC par I.R.I.S.
Segmenteur XML développé par Apache. Ce produit comprend le logiciel développé par Apache Software Foundation.

Tous les autres produits mentionnés dans le présent guide utilisateur sont des marques commerciales ou déposées de leurs propriétaires respectifs.

CHAPITRE 1
PRESENTATION DE READIRIS

GAIN DE TEMPS ET ÉLIMINATION DU TRAVAIL DE RÉENCODAGE

Merci d'avoir acheté Readiris. Ce logiciel vous apportera sans nul doute une aide précieuse dans le réencodage de vos textes, tableaux, graphismes, codes à barres et textes manuscrits.

Quelle que soit l'efficacité des ordinateurs, le travail préalable de saisie des informations reste une étape incontournable. Si vous avez déjà retapé un rapport de 15 pages ou un grand tableau de chiffres, vous savez combien cette tâche peut être longue et fastidieuse. Utilisez ce logiciel d'OCR à la pointe du progrès pour convertir automatiquement des documents papier ou des fichiers image numérisés en documents modifiables et entièrement indexés qu'il est possible d'archiver et de partager. Readiris propose deux modes de reconnaissance : le premier garantit une vitesse maximale et le second une précision d'OCR optimale.

Numérisez un document imprimé ou dactylographié, indiquez les zones à reconnaître avec Readiris (ou laissez au système le soin de le faire pour vous), exécutez la reconnaissance des caractères et exportez le document dans votre programme de traitement de texte. Les documents comptant de nombreuses pages sont traités d'une seule traite. Quelques clics de souris vous permettent d'éviter de longues heures de travail tandis que Readiris convertit vos documents papier en fichiers électroniques modifiables : il offre des performances 40 fois supérieures à la frappe manuelle.

Grâce à l'Assistant qui vous guide dans les différentes étapes de configuration requises, Readiris est rapidement opérationnel. Vous avez également la possibilité d'utiliser les SmartTasks pour accélérer encore le processus. Vous pouvez envoyer les résultats de la reconnaissance directement à votre application de traitement de texte ou de tableur, les archiver au format PDF ou XPS, etc. Pour reconnaître les télécopies et convertir les documents PDF, faites glisser leur fichier image de l'Explorateur Windows vers la fenêtre d'application de Readiris. Vous pouvez aussi envoyer rapidement une image à Readiris à l'aide du menu contextuel.

Readiris reconnaît les données tabulaires et les recrée sous la forme de feuilles de calcul dans votre tableur ou en tant qu'objets tableau dans votre application de traitement de texte. Les données numériques sont directement prêtes à être traités.

Readiris repose sur les technologies de reconnaissance les plus évoluées. La reconnaissance du texte indépendante de la police est complétée par des techniques d'auto-apprentissage. Le système est capable d'apprendre de nouveaux mots et caractères grâce à une analyse contextuelle et linguistique. En d'autres termes, la précision de l'OCR s'améliore avec le temps et l'utilisation.

Readiris prend en charge jusqu'à 128 langues : toutes les langues d'Amérique et d'Europe, y compris les langues d'Europe centrale, les langues baltes, le grec, le turc et le cyrillique. Grâce à des modules complémentaires, Readiris peut lire des documents en arabe, persan et hébreu et dans quatre langues asiatiques, à savoir le japonais, le coréen, le chinois simplifié et le chinois traditionnel. Readiris peut même traiter des alphabets mélangés : le logiciel détecte les mots « occidentaux » qui ressortent des documents en grec, en cyrillique, en langue asiatique, en hébreu et en arabe puisque beaucoup de noms propres intranscriptibles, des noms de marques, etc., sont écrits à l'aide des symboles occidentaux.

Pour reconnaître des documents, Readiris applique une analyse linguistique *pendant* de l'étape de reconnaissance et non après. En conséquence, il reconnaît tous les types de documents avec une

précision optimale, y compris les documents de piètre qualité, les télécopies et les impressions matricielles. Il n'a aucune difficulté à traiter les documents mal copiés ou numérisés contenant des caractères trop clairs ou trop foncés. Il est en mesure de résoudre des caractères liés et de reconstituer les caractères incomplets ou fragmentés, tels que les symboles matriciels.

En outre, Readiris offre une fonction de vérification utilisateur (facultative). Lorsqu'elle est activée, la fonction de vérification utilisateur (« Apprentissage interactif ») marque les caractères considérés comme douteux par le système de reconnaissance et permet également d'améliorer la précision du système. Toutes les solutions que vous validez sont enregistrées dans la mémoire du système, ce qui améliore la vitesse et la fiabilité du système et accroît son intelligence au fur et à mesure de son utilisation. Ce puissant outil d'apprentissage interactif permet également à Readiris de reconnaître des caractères spéciaux, notamment les symboles mathématiques et les signes typographiques et de traiter les polices déformées.

Pour accroître encore votre productivité, Readiris, non content de reconnaître les textes, peut également les formater pour vous. Plusieurs niveaux de formatage sont disponibles. Lorsque vous utilisez le « formatage automatique », Readiris recrée une copie fidèle du document numérisé : le formatage des mots, des paragraphes et des pages du document d'origine est préservé. Le système utilise des pas de caractères, des tailles et des styles similaires à ceux utilisés dans le document source pendant la reconnaissance. L'emplacement des colonnes, des blocs de texte et des images respecte les documents d'origine. Readiris peut même inclure la photo en arrière-plan de la page numérisée dans le document reconnu. En outre, puisqu'il prend en charge la numérisation couleur et à niveaux de gris, vous pouvez reproduire n'importe quelle image, dessin au trait, photos en noir et blanc et illustrations en couleur. Lorsqu'un document contient des tableaux, Readiris les réorganise dans des véritables cellules et recrée les bordures de cellule des tableaux d'origine.

En d'autres termes, Readiris vous offre la possibilité d'archiver une copie fidèle de vos documents et de les convertir en fichiers texte compacts et modifiables au lieu de simples images numérisées.

Les codes à barres présents sur une page numérisée peuvent être lus, au même titre que du texte manuscrit, pour autant que vous écriviez en « lettres capitales » bien espacées.

Readiris est compatible avec la norme Twain et prend en charge un large éventail de scanners à plat et à feuilles, de périphériques « tout-en-un », de périphériques multifonctions et d'appareils photo numériques. La numérisation avec pause vous permet de numériser efficacement des documents de plusieurs pages lorsque votre numériseur n'est pas équipé d'un bac d'alimentation.

LA FAMILLE READIRIS

Le tableau ci-dessous présente les caractéristiques des différentes versions :

Readiris Home 12
Fonctionnalités réduites
25 langues reconnues
Prise en charge des images PDF, DCX, DJV, DJVU, JPG, JPEG, J2C, J2K, JP2, PNG, TIF, TIFF, BMP, PCX
Génération de fichiers PDF Image-Texte, DOCX, ODT, WordML, SpreadsheetML, RTF, HTM, XML, TXT, TIFF, etc.

Readiris Pro 12	**Readiris Corporate 12**
Fonctionnalités de base	Fonctionnalités de base
128 langues reconnues	128 langues reconnues
Prise en charge des images PDF, DCX, DJV, DJVU, JPG, JPEG, J2C, J2K, JP2, PNG, TIF, TIFF, BMP, PCX	Prise en charge des images PDF, DCX, DJV, DJVU, JPG, JPEG, J2C, J2K, JP2, PNG, TIF, TIFF, BMP, PCX
Génération de quatre types de fichiers PDF, PDF-iHQC (niveau I), quatre types de fichiers XPS, XPS-iHQC (niveau I), DOCX, ODT, XLS, WordML, SpreadsheetML, RTF, HTM, XML, TXT, TIFF, etc.	Génération de quatre types de fichiers PDF, PDF-iHQC (niveau I-III), PDF/A, quatre types de fichiers XPS, XPS-iHQC (niveau I), DOCX, ODT, XLS, WordML, SpreadsheetML, RTF, HTM, XML, TXT, TIFF, etc.
	Reconnaissance d'importants volumes
	Traitement automatisé
	Indexation de documents
	Reconnaissance des cartes de visite

Readiris Pro 12 Asian	**Readiris Corporate 12 Asian**
Fonctionnalités de base	Fonctionnalités de base
128 langues reconnues	128 langues reconnues
Prise en charge des images PDF, DCX, DJV, DJVU, JPG, JPEG, J2C, J2K, JP2, PNG, TIF, TIFF, BMP, PCX	Prise en charge des images PDF, DCX, DJV, DJVU, JPG, JPEG, J2C, J2K, JP2, PNG, TIF, TIFF, BMP, PCX
Génération de quatre types de fichiers PDF, PDF-iHQC (niveau I), quatre types de fichiers XPS, XPS-iHQC (niveau I), DOCX, ODT, XLS, WordML, SpreadsheetML, RTF, HTM, XML, TXT,	Génération de quatre types de fichiers PDF, PDF-iHQC (niveau I-III), PDF/A, quatre types de fichiers XPS, XPS-iHQC (niveau I), DOCX, ODT, XLS, WordML, SpreadsheetML, RTF, HTM, XML, TXT,

TIFF, etc. Reconnaissance du chinois simplifié et traditionnel Reconnaissance du japonais Reconnaissance du coréen	TIFF, etc. Reconnaissance du chinois simplifié et traditionnel Reconnaissance du japonais Reconnaissance du coréen Reconnaissance d'importants volumes Traitement automatisé Indexation de documents Reconnaissance des cartes de visite

Readiris Pro 12 Middle-East*	**Readiris Corporate 12 Middle-East***
Fonctionnalités de base 128 langues reconnues Prise en charge des images PDF, DCX, DJV, DJVU, JPG, JPEG, J2C, J2K, JP2, PNG, TIF, TIFF, BMP, PCX Génération de quatre types de fichiers PDF, PDF-iHQC (niveau I), quatre types de fichiers XPS, XPS-iHQC (niveau I), DOCX, ODT, XLS, WordML, SpreadsheetML, RTF, HTM, XML, TXT, TIFF, etc. Reconnaissance de l'arabe et du persan Reconnaissance de l'hébreu	Fonctionnalités de base 128 langues reconnues Prise en charge des images PDF, DCX, DJV, DJVU, JPG, JPEG, J2C, J2K, JP2, PNG, TIF, TIFF, BMP, PCX Génération de quatre types de fichiers PDF, PDF-iHQC (niveau I-III), PDF/A, quatre types de fichiers XPS, XPS-iHQC (niveau I), DOCX, ODT, XLS, WordML, SpreadsheetML, RTF, HTM, XML, TXT, TIFF, etc. Reconnaissance de l'arabe et du persan Reconnaissance de l'hébreu Reconnaissance d'importants volumes Traitement automatisé

	Indexation de documents Reconnaissance des cartes de visite

(*) Aucune version Mac n'est disponible.

CHAPITRE 2
INSTALLATION DE READIRIS

CONFIGURATION REQUISE

La **configuration minimale** suivante est nécessaire pour exécuter Readiris :

- PC Intel 486 ou équivalent. Un PC Pentium est recommandé.
- Mémoire RAM de 256 Mo.
- 120 Mo d'espace libre sur le disque.

  (105 Mo d'espace disque sont suffisants si vous n'installez pas les fichiers exemple.)

- Système d'exploitation Windows Vista, Windows XP ou Windows.

Notez que certains pilotes de scanner ne fonctionnent pas toujours sous les dernières versions de Windows. Consultez la documentation de votre scanner pour connaître les plates-formes prises en charge.

INSTALLATION DU LOGICIEL

Procédure d'installation de Readiris :

- Connectez-vous à Windows en tant qu'administrateur et assurez-vous de posséder les droits d'administration requis.

- Connectez votre scanner à votre ordinateur et installez le logiciel qui l'accompagne.
  Testez votre scanner. En cas de problème, contactez le fabricant de votre scanner.

- Insérez le CD-ROM de Readiris dans le lecteur de CD-ROM et suivez les instructions d'installation qui s'affichent à l'écran.

- Cliquez sur **Readiris** pour démarrer l'installation (d'autres produits logiciels sont proposés : Copernic Desktop Search Home Edition et Cardiris 4 LE).

- Sélectionnez la langue d'installation et cliquez sur **OK**.

- Acceptez les conditions du contrat de licence.

- Vous avez le choix entre une installation **complète** et **personnalisée**. Sélectionnez les options voulues et cliquez sur **Suivant** chaque fois que vous avez défini les options requises dans un écran.
  Tous les lexiques et les exemples d'images sont installés par défaut ainsi qu'un guide utilisateur électronique et l'aide en ligne.

- Cliquez sur **Finir** pour terminer l'installation.
  Le sous-menu **Applications I.R.I.S. - Readiris** dans le menu Windows **Tous les programmes** est créé automatiquement par le programme d'installation.
  Le programme d'installation crée également un raccourci vers l'application Readiris sur le Bureau Windows.

- Répétez le processus d'installation pour installer un autre logiciel à partir du CD-ROM.

DÉSINSTALLATION DU LOGICIEL

Il n'existe qu'un seul moyen pour désinstaller correctement Readiris : utiliser l'Assistant d'installation/désinstallation Windows. Il est vivement recommandé de *ne pas* désinstaller Readiris ou l'un de ses modules en supprimant manuellement les fichiers du programme.

Pour désinstaller Readiris :

- Fermez l'application.
- Dans le menu Démarrer, cliquez sur Panneau de configuration.
- Sélectionnez l'icône **Ajouter ou supprimer des programmes** puis cliquez sur l'option **Modifier ou supprimer des programmes**.
- Sélectionnez **Readiris** dans la liste et cliquez sur **Supprimer**.
- Suivez les instructions qui s'affichent à l'écran.

ENREGISTREMENT DU LOGICIEL

N'oubliez pas d'enregistrer votre licence Readiris. Cet enregistrement vous procure divers avantages :

- vous serez informé du développement de nouveaux produits et des produits I.R.I.S. connexes ;
- vous aurez droit à une assistance produits gratuite ;
- vous pourrez bénéficier des offres spéciales relatives aux produits I.R.I.S.

Pour s'enregistrer :

Lancez l'**Assistant d'enregistrement** dans le menu **S'enregistrer**. Suivez les instructions qui s'affichent à l'écran.

ASSISTANCE PRODUITS

Dès que vous avez enregistré votre produit, vous bénéficiez de l'assistance I.R.I.S. pour toutes les fonctionnalités de base du logiciel. Pour contacter I.R.I.S.:

Europe :
support@irislink.com
Tél. :+32 10 45 13 64

Etats-Unis :
support@irisusa.com
Tél. :+1 800 447 4744

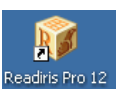
Asie-Pacifique :
support@irislink.com
Tél. : +852 22646133

CHAPITRE 3
MISE EN ROUTE

EXÉCUTION DE READIRIS

Pour exécuter Readiris :

- Lancez **Readiris** à partir du menu **Démarrer** de Windows ou double-cliquez sur le raccourci de votre Bureau.



- Cliquez n'importe où dans l'écran de démarrage pour lancer Readiris.

  L'Assistant OCR s'ouvre automatiquement.

UTILISATION DE L'ASSISTANT OCR

L'Assistant OCR permet de définir rapidement tous les paramètres nécessaires au fonctionnement de Readiris.

Lorsque vous démarrez Readiris, cliquez à un endroit quelconque de l'écran de démarrage pour lancer l'Assistant OCR.

Etape 1
Sélectionnez la **source de l'image**.

Vous pouvez capturer des images à l'aide de votre **numériseur** ou ouvrir des **fichiers image**.
Sélectionnez les options de **rotation** et de **redressement** de votre choix.

> Pour plus d'informations, consultez la rubrique **Sélection des options**.
>
> Pour vous familiariser avec Readiris, utilisez les exemples d'images fournis avec le logiciel. Ils se trouvent sur le CD-ROM de Readiris et dans le sous-dossier **Samples** du dossier d'installation de Readiris.

Cliquez sur **Suivant** pour passer à l'étape suivante.

Etape 2
Si vous avez sélectionné un numériseur, cliquez sur le bouton **Modifier** pour définir les paramètres du numériseur.

> Pour plus d'informations sur la configuration du numériseur, consultez la rubrique **Numérisation des documents papier**.

Cliquez sur **OK** pour enregistrer les paramètres.
Cliquez sur **Suivant** pour passer à l'étape suivante.

Etape 3
Cliquez sur le bouton **Modifier** pour changer la langue du document. La langue du document a la valeur Anglais américain par défaut.
Sélectionnez la langue ou la combinaison de langues requise ainsi que les langues secondaires dans la liste puis cliquez sur **OK**.
Utilisez le curseur pour définir les paramètres **Vitesse-Précision** requis.

> Pour plus d'informations, consultez la rubrique **Sélection de la langue du document**.

Cliquez sur **OK** pour enregistrer les paramètres.
Cliquez sur **Suivant** pour passer à l'étape suivante.

Etape 4
Cliquez sur le bouton **Modifier** pour changer le format de sortie ou

l'application de destination. Microsoft Word est l'application de destination par défaut.
Sélectionnez l'application ou le format de sortie requis dans la liste déroulante **Envoyer à** ou **Fichier externe**.
Accédez aux différents onglets et sélectionnez les options de votre choix. Les options indisponibles pour l'application/le format de sortie sélectionné sont affichées en grisé.

> Pour plus d'informations, consultez la section **Formatage et enregistrement des documents**.

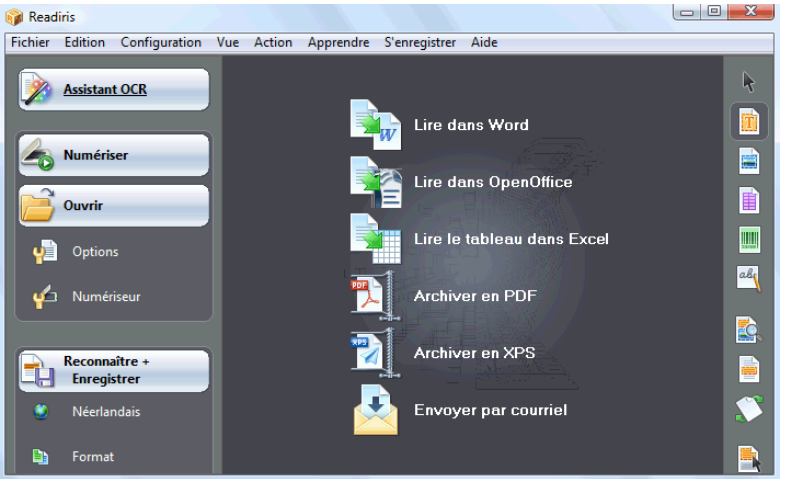
Cliquez sur **OK** pour enregistrer les paramètres.
Cliquez sur **Suivant** pour passer à l'étape suivante.

Etape 5
Cliquez sur le bouton **Démarrer** pour ouvrir/numériser et reconnaître le document.

INTERFACE UTILISATEUR

Pour découvrir l'interface de Readiris, cliquez à un endroit quelconque de l'écran de démarrage **Readiris** et cliquez sur **Annuler** lorsque l'Assistant OCR est exécuté.

L'interface vide de Readiris s'affiche.



Cette interface se compose des éléments suivants :

- **SmartTasks** (au centre)

  Les SmartTasks sont des commandes prédéfinies qui vous permettent d'utiliser les fonctions Readiris les plus utilisées d'un simple clic de bouton.

  Utilisez ces SmartTasks pour numériser, reconnaître et envoyer vos documents à l'application cible ou au format de sortie de votre choix.

  Les SmartTasks appliquent des paramètres par défaut mais peuvent être facilement configurées par un clic droit pour répondre à des besoins particuliers.

- **Barre d'outils principale** (barre d'outils à gauche)

  Utilisez les commandes de la barre d'outils principale pour numériser et reconnaître manuellement les documents.
  L'ordre à respecter pour ce faire est décrit dans l'**Assistant OCR**.

- **Barre d'outils Image** (barre d'outils à droite)

Utilisez les boutons de la barre d'outils Image pour modifier des documents dans l'interface de Readiris.
Pointez sur les différents boutons pour afficher leur info-bulle.

Lorsqu'un document a été ouvert ou numérisé dans Readiris, trois zones principales sont ajoutées à l'interface.

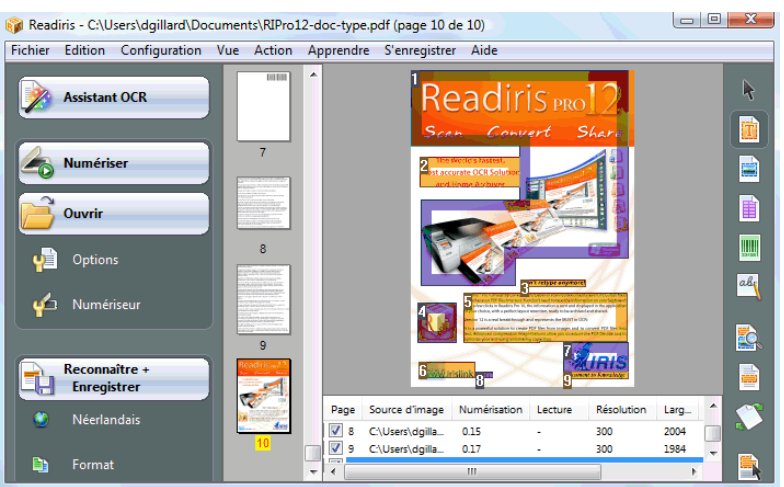
- **Barre d'outils des pages** (à droite de la barre d'outils principale)

  La barre d'outils des pages affiche les miniatures des pages, lesquelles affichent des informations de configuration lorsque vous pointez sur celles-ci.

- **Fenêtre Image** (au centre)

- **Panneau de document** (en bas)

  Le panneau du document affiche des informations statistiques sur les documents ouverts dans Readiris, notamment la date et l'heure de la numérisation et de l'OCR, la résolution, la largeur et la hauteur des documents, etc.



MODIFICATION DE LA LANGUE DE L'INTERFACE UTILISATEUR

L'interface utilisateur de Readiris est disponible dans plusieurs langues.

Pour modifier la langue de l'interface utilisateur :

- Dans le menu **Configuration**, cliquez sur **Langue interface utilisateur**.
- Dans la liste **Langue**, sélectionnez la langue requise puis cliquez sur **OK** pour confirmer.

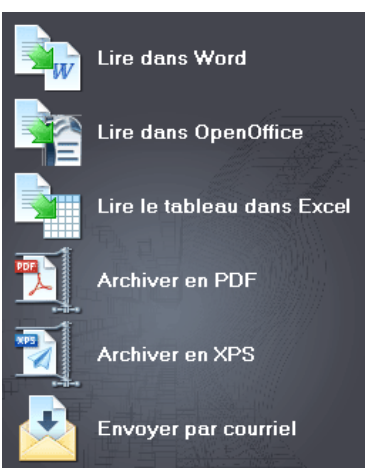
> Remarque : si vous avez sélectionné une langue incorrecte, cliquez sur **Ctrl+U**. Dans la boîte de dialogue **Langue** qui apparaît, vous pouvez choisir une autre langue de la liste.

CHAPITRE 4
READIRIS SMARTTASKS

UTILISATION DES SMARTTASKS READIRIS

Lorsque vous démarrez Readiris, cliquez à un endroit quelconque de l'écran de démarrage Readiris et cliquez sur **Annuler** lorsque l'Assistant OCR est exécuté. Les SmartTasks Readiris s'affichent au centre de l'écran.

Il s'agit de commandes prédéfinies qui vous permettent d'utiliser les fonctions Readiris les plus utilisées d'un simple clic de bouton.



Les différents boutons des SmartTasks permettent d'effectuer les opérations suivantes :

1. Numériser et reconnaître les documents et les envoyer directement vers **Word** pour traitement de texte.

   Microsoft Word est l'application de destination par défaut. Consultez la section **Formatage des documents texte** pour en savoir plus sur les autres applications disponibles.

2. Numériser et reconnaître les documents et les envoyer directement vers **OpenOffice** pour traitement de texte.

   OpenOffice.org Writer est l'application de destination par défaut. Consultez la section **Formatage des documents texte** pour en savoir plus sur les autres applications disponibles.

3. Numériser et reconnaître des tableaux et les envoyer directement vers **Excel** et d'autres tableurs.

   Microsoft Excel est l'application de destination par défaut. Consultez la section **Formatage de documents comportant des tableaux** pour en savoir plus.

4. Numériser et reconnaître des documents et les **archiver** en tant que fichiers **PDF**.

   Adobe Acrobat PDF Image-Texte est le format de sortie par défaut. Consultez la section **Création de documents PDF** pour en savoir plus sur les autres formats disponibles.

5. Numériser et reconnaître des documents et les **archiver** en tant que fichiers **XPS**.

   XPS Image-Texte est le format de sortie par défaut. Consultez la section **Création de documents XPS** pour en savoir plus sur les autres formats disponibles.

6. Numériser et reconnaître les documents et les envoyer directement par **e-mail**.

   Les documents sont envoyés par défaut au format PDF Image-Texte via votre application de messagerie par défaut. Consultez la section

Formatage de documents pour en savoir plus sur les autres formats disponibles.

Lorsque vous utilisez Readiris pour la première fois, vous devez configurer les SmartTasks.

Pour configurer les SmartTasks :

- Cliquez avec le bouton droit sur la SmartTask que vous souhaitez utiliser.
- Sélectionnez **Numériseur** ou **Fichiers image** comme source d'image.
    - Lorsque vous sélectionnez **Numériseur**, Readiris démarre votre numériseur dès que vous cliquez sur la SmartTask. Le ou les documents numérisés sont affichés dans l'interface, traités et enregistrés.

      Votre numériseur doit être correctement configuré pour que les SmartTasks fonctionnent.

      Pour ce faire, procédez comme suit :

        - Cliquez sur le bouton **Numériseur** dans la barre d'outils principale.
        - Cliquez sur **Modèle** de numériseur et sélectionnez votre numériseur dans la liste.

          Si votre numériseur ne figure pas dans la liste, sélectionnez Twain other models.
        - Le cas échéant, cliquez sur **Configurer** pour sélectionner la source Twain.
        - Cliquez sur **OK** pour enregistrer les paramètres.

Pour plus d'informations sur la configuration du numériseur et la numérisation des documents papier, consultez la section **Numérisation des documents papier**.

- Lorsque vous sélectionnez **Fichiers image** et que vous cliquez sur la SmartTask, Readiris ouvre la boîte de dialogue **Source**, dans laquelle vous pouvez sélectionner les fichiers image à traiter.

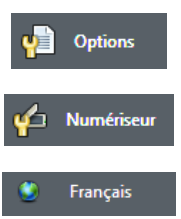
  Pour plus d'informations sur l'ouverture des fichiers image, consultez la rubrique **Ouverture de fichiers image**.

- Cliquez sur **Configurer** pour modifier le **format** de sortie et ses **options**.

  Notez que les options et formats de sortie disponibles dépendent du type de SmartTask sélectionnée.

  Consultez la section **Formatage de documents** pour en savoir plus sur les autres formats disponibles.

- Lorsque vous avez terminé de configurer les SmartTasks, utilisez les boutons de la barre d'outils principale pour définir les paramètres de **langue**, les **options** d'optimisation d'images et, le cas échéant, la **configuration du numériseur**.



  Pour en savoir plus sur les options susmentionnées, consultez les sections **Sélection des options**, **Numérisation de documents papier** et **Sélection de la langue du document**.

- Pour terminer, cliquez sur la SmartTask pour l'utiliser.

Readiris exécute automatiquement tout le processus de reconnaissance.

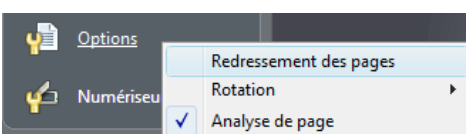
CHAPITRE 5
NUMERISATION DES DOCUMENTS

SÉLECTION DES OPTIONS

Avant de numériser des documents papier ou d'ouvrir des fichiers image, vous pouvez sélectionner plusieurs options d'optimisation d'images. Lorsqu'elles sont activées, ces options sont appliquées lors de l'ouverture et de la numérisation des documents.

Procédure

- Cliquez sur le bouton **Options** dans la barre d'outils principale pour sélectionner diverses options d'optimisation d'images.



  - Cliquez sur **Redressement des pages** pour corriger l'angle éventuel des pages numérisées.
  - Pointez sur **Rotation** et précisez si vous souhaitez que Readiris applique une rotation des pages **automatique** ou de **90 degrés dans le sens inverse des aiguilles d'une montre**, de **90 degrés dans le sens des aiguilles d'une montre** ou de **180 degrés**.

    Notez que ces deux options ralentissent quelque peu la numérisation. Ne les sélectionnez qu'en cas de nécessité.

- L'option **Analyse de page** est activée par défaut.

  De cette façon les images numérisées ou ouvertes sont automatiquement découpées en fenêtres.

  Vous pouvez également utiliser les outils de fenêtrage de la barre d'outils des images pour modifier les résultats de l'analyse de page ou pour appliquer un fenêtrage manuel aux documents. Pour plus d'informations, consultez la rubrique **Fenêtrage de documents**.

- Lorsque vous avez terminé de définir tous les paramètres (**configuration du numériseur**, **options**), cliquez sur le bouton **Numériser** ou **Ouvrir** pour numériser ou ouvrir les fichiers image.

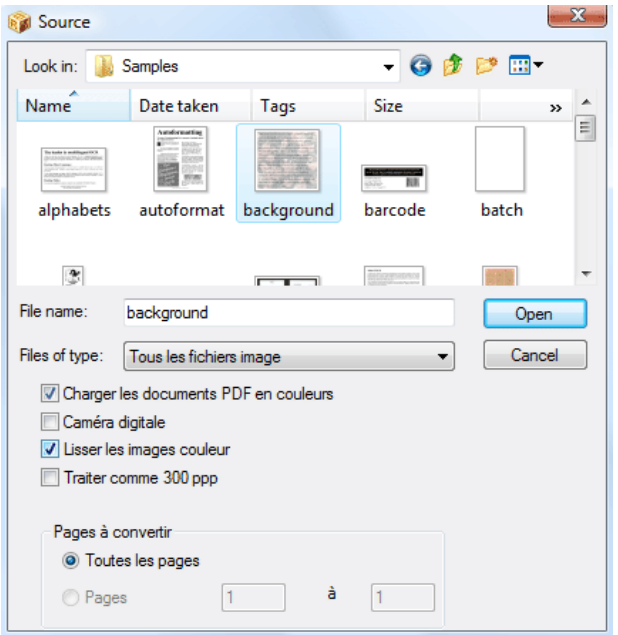
Notez que les options de traitement susmentionnées sont également disponibles dans le menu **Configuration**.

OUVERTURE DE FICHIERS IMAGE

Readiris vous permet de traiter des documents papier que vous pouvez numériser avec votre scanner ou de traiter des fichiers image existants de différents formats.

Pour ouvrir des fichiers image existants :

- Cliquez sur le bouton **Ouvrir** pour rechercher des fichiers image.



Conseil : vous pouvez également **faire glisser** les fichiers image vers la fenêtre des images de Readiris pour les ouvrir.

Conseil : cliquez avec le bouton droit sur un fichier image que vous souhaitez ouvrir, pointez sur **Ouvrir avec** et cliquez sur **Application IOCR**. Le logiciel Readiris s'ouvre et affiche l'image.

Conseil : lorsque vous chargez des fichiers image de plusieurs pages (images TIFF et télécopies DCX) et des documents PDF, vous pouvez définir l'étendue de pages (dans le cas où vous avez uniquement besoin d'un chapitre spécifique d'un document, par exemple). Pour ce faire, cliquez sur **Ouvrir** dans la barre d'outils principale. Dans la zone **Pages à convertir**, sélectionnez **Pages** et spécifiez les pages à charger. Voir aussi la rubrique **Ouverture et reconnaissance de plusieurs fichiers image**.

Conseil : pour accélérer le chargement, cliquez sur le bouton **Ouvrir** et désactivez l'option **Charger les documents PDF en couleurs** lors du traitement des documents PDF.

- Readiris prend en charge les formats d'image suivants :

Format graphique	Extension de fichier
Adobe Acrobat PDF	*.pdf
Fax DCX	*.dcx
Images DjVu	*.djv, *.djvu
Images JPEG	*.jpg, *.jpeg
Images JPEG 2000	*.j2c, *.j2k, *.jp2
Portable Network Graphics	*.png
Images TIFF (compressées et multipages) Non compressées et LZW, PackBits, Groupe 3, Groupe 4 et JPEG compressées	*.tif, *.tiff
Bitmaps Windows	*.bmp
Images ZSoft Paintbrush	*.pcx

- Sélectionnez le fichier image de votre choix et cliquez sur **Ouvrir**.

Remarque : les options de la boîte de dialogue **Source** s'appliquent également à la numérisation des documents et sont décrites dans la rubrique **Numérisation des documents papier**.

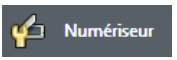
NUMÉRISATION DE DOCUMENTS PAPIER

Readiris vous permet de traiter des documents papier que vous pouvez numériser avec votre scanner mais aussi de traiter des fichiers image existants de différents formats.

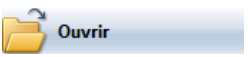
Lorsque vous traitez des documents papier, Readiris lance votre scanner dès que vous cliquez sur le bouton **Numériser** et affiche le document numérisé dans l'interface.

Pour numériser des documents :

- Cliquez sur le bouton **Numériseur** pour définir les paramètres du scanner.



Remarque : plusieurs options de la boîte de dialogue **Numériseur** sont également disponibles dans la boîte de dialogue **Ouvrir**.



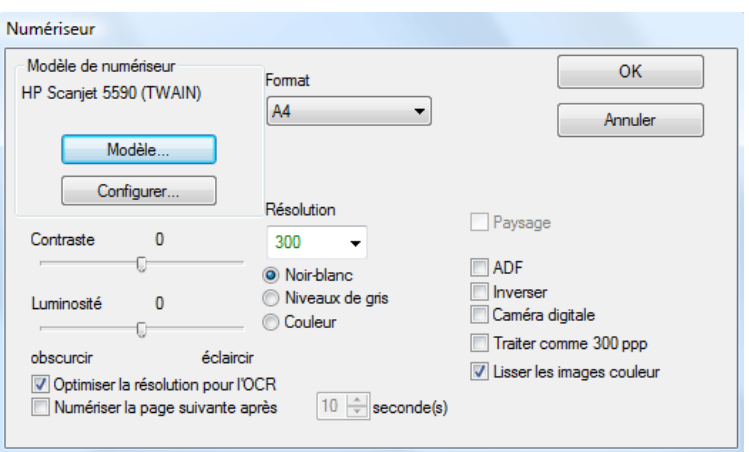
- Sélectionnez le modèle de numériseur correct.

  Si votre numériseur ne figure pas dans la liste, sélectionnez **Twain other models** et cliquez sur **OK**.

- Readiris prend en charge la plupart des scanners à plat et à feuilles, les périphériques « tout-en- un » (« MFP » - « périphérique multifonction »), les appareils photo numériques et numériseurs standard actuellement disponibles. Readiris est compatible avec Twain.

  Remarque : le bouton **Configurer** est disponible uniquement lorsqu'il y a lieu de configurer certains paramètres. Cliquez sur ce bouton pour sélectionner la source Twain.

- Sélectionnez les paramètres du numériseur.



Format et résolution

Readiris prend en charge un large éventail de résolutions et de formats papier.

Notez qu'il est recommandé d'utiliser une résolution de numérisation de 300 ppp.

Utilisez une résolution de 400 ppp pour la reconnaissance de texte asiatique ou de caractères très petits.

Mode couleur

Readiris peut numériser des documents et ouvrir des fichiers image en **couleur**, en **noir et blanc** et à **niveaux de gris**.

Contraste - Luminosité

Utilisez le curseur pour définir les valeurs de contraste et de luminosité adéquats, afin d'obtenir une numérisation optimale.

Optimisation de la résolution pour l'OCR

Sélectionnez **Optimiser la résolution pour l'OCR** afin de corriger la résolution des images numérisées avec trop de détails (plus de 600 ppp). Readiris réduira la résolution.

Remarque : cette option n'augmente jamais la résolution des images numérisées avec trop peu de détails.

Numérisation de documents de plusieurs pages

Lorsque vous numérisez des documents de plusieurs pages et que vous utilisez un scanner à feuilles, sélectionnez l'option **ADF** (alimentation automatique de documents). Placez les pages à numériser dans le bac d'alimentation et lancez la numérisation.

Vous pouvez également utiliser l'option **Numérisation avec pause** lorsque vous employez un scanner à plat **:** sélectionnez

l'option **Numériser la page suivante après** et spécifiez le délai en secondes après lequel Readiris doit numériser une autre page.

Pour plus d'informations, consultez la rubrique **Numérisation et reconnaissance de documents de plusieurs pages**.

Important : les options non disponibles pour le numériseur sélectionné sont affichées en grisé.

Numérisation des images au format paysage

Sélectionnez l'option **Paysage** lorsque vous numérisez des images au format paysage.

Luminosité automatique

Avec certains numériseurs, l'option **Luminosité automatique** est sélectionnée par défaut. Cette option ajuste automatiquement les paramètres de contraste et de luminosité.

Numérisation des images inversées

Sélectionnez l'option **Inverser** lorsque vous numérisez des images en noir et blanc inversées.

Utilisation d'un appareil photo numérique

Sélectionnez **Appareil photo numérique** lorsque vous utilisez un appareil photo numérique comme source de numérisation. Readiris fait appel à des routines de reconnaissance spéciales pour traiter les images d'appareil photo numérique.

Conseils en cas d'utilisation d'un appareil photo numérique comme source de numérisation :

- Etalonnez l'appareil en prenant une photo d'un document de couleur blanche.
- Sélectionnez toujours la résolution d'image la plus élevée.

- Activez le mode macro de l'appareil afin de faire un gros plan.
- Utilisez uniquement le zoom optique et non numérique.
- Placez l'appareil directement au-dessus du document. Evitez de photographier ce dernier sous un angle quelconque.
- Prenez des photos stables. Le cas échéant, utilisez un pied.
- Désactivez le flash si vous photographiez un papier brillant.
- Evitez d'ouvrir les photos compressées.
- Adaptez les paramètres de contraste et de luminosité de Readiris à l'environnement (lumière du jour, artificielle, néon).
- Sélectionnez couleur ou niveaux de gris comme mode couleur.

Traitement avec une résolution de 300 ppp

Sélectionnez **Traiter comme 300 ppp** lorsque vous traitez des images dont la résolution n'est pas correcte ou connue. Les images seront traitées comme si elles possédaient une résolution de 300 ppp.

La plupart du temps, la résolution des photos numériques est inconnue.

Lissage des images couleurs

Cette option est sélectionnée par défaut dans la mesure où certains scanners nécessitent un lissage d'images pour obtenir une reconnaissance correcte des images couleur et à niveaux de gris.

- Lorsque vous avez terminé de définir tous les paramètres (**configuration du numériseur**, **options**), cliquez sur **Numériser** pour numériser les documents.

  Remarque : faites attention au désalignement. Un désalignement supérieur à 0,5 augmente le risque d'erreurs d'OCR.

CHAPITRE 6
OPTIMISATION DES DOCUMENTS NUMÉRISÉS

Lorsque vous ouvrez ou numérisez des images couleur ou à niveaux de gris très claires ou très foncées, il est parfois nécessaire d'optimiser ces images avant de procéder à leur reconnaissance afin d'obtenir des résultats d'OCR satisfaisants.

Pour optimiser les images :

- Ouvrez ou numérisez un document couleur ou à niveaux de gris.

  Vérifiez que les **paramètres du numériseur** sont corrects.

  Important: les **paramètres du numériseur** et les **paramètres d'ajustement** semblent identiques mais notez que les deux ensembles de paramètres sont appliqués à des étapes différentes du processus de numérisation/reconnaissance.

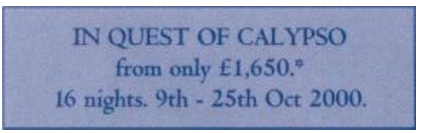
- Dans le menu **Action**, cliquez sur **Ajuster l'image**.

  Readiris utilise des routines de **binarisation** intelligentes pour convertir les images couleur ou à niveaux de gris en images en noir et blanc, utilisées pour réaliser l'OCR.

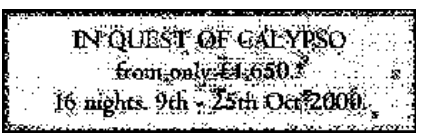
  - Sélectionnez **Lisser l'image couleur** pour obtenir une image plus homogène.

    Grâce à cette option, qui lisse les différences d'intensité, les images couleur et à niveaux de gris sont plus uniformes. En conséquence, le contraste est plus marqué entre l'avant-plan (texte) et l'arrière-plan (illustration).

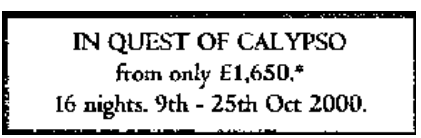
Remarque : le lissage représente parfois la seule manière de distinguer le texte d'un arrière-plan coloré.



(Image d'origine)



(Image en noir et blanc binarisée)



(Image lissée)

- Utilisez le curseur pour augmenter ou diminuer la valeur de **Luminosité**.

Les paramètres **Luminosité** déterminent la luminosité globale de l'image. Utilisez-les pour foncer ou éclaircir l'image lorsque le texte est illisible.

Exemple 1 : éclaircir une image sombre afin d'éliminer l'arrière-plan de la page.

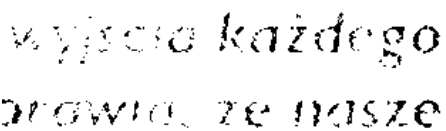
(Image couleur)

(Image numérisée. Les paramètres de binarisation par défaut génèrent une image noire)

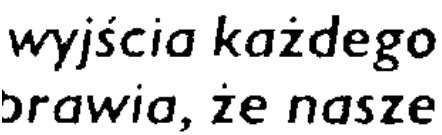
(L'image éclaircie donne des résultats de reconnaissance satisfaisants)

Exemple 2 : foncer une image lorsque le texte est tellement clair qu'il n'apparaît pas dans l'image binarisée.

(Image couleur)



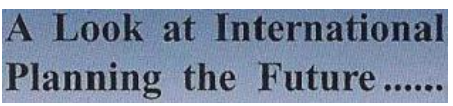
(Image numérisée. Les paramètres de binarisation par défaut génèrent des caractères fragmentés)



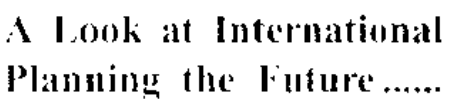
(L'image plus foncée donne des résultats de reconnaissance satisfaisants)

- Utilisez le curseur pour augmenter ou diminuer la valeur de **Contraste**.

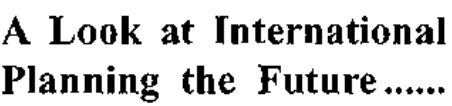
  Les paramètres **Contraste** déterminent le contraste entre les zones plus claires et plus foncées d'une image. Utilisez ces paramètres pour faire ressortir les formes des caractères dans un arrière-plan coloré.



(Image couleur)



(Les paramètres de contraste par défaut génèrent des caractères fragmentés)



(Des paramètres de contraste plus élevés donnent des résultats de reconnaissance satisfaisants)

- Utilisez le curseur pour augmenter ou diminuer les valeurs des options **Eliminer le bruit**.

  L'élimination du bruit permet d'enlever des petites tâches dans les images en noir et blanc.

- Cliquez sur **Appliquer** pour afficher un aperçu des résultats.
- Si les résultats sont satisfaisants, cliquez sur **OK**. Dans le cas contraire, modifiez à nouveau les paramètres.
- Cliquez sur **Reconnaître + Enregistrer** pour reconnaître le document.

CHAPITRE 7
ENREGISTREMENT DES DOCUMENTS EN TANT QUE FICHIERS IMAGE

Les documents papier que vous numérisez ne doivent pas nécessairement être reconnus directement. Ils peuvent être enregistrés en tant que fichiers image.

Pour ce faire, procédez comme suit :

- Numérisez le document.
- Dans le menu **Fichier**, cliquez sur les commandes **Sauvegarder la pleine page comme image** ou **Sauvegarder toutes les pages comme image**.
- Après quoi, ouvrez le fichier image sauvegardé et lancez la reconnaissance.

Enregistrement des graphiques uniquement

Vous pouvez décider d'enregistrer les fenêtres de graphiques sans le texte du document.

Pour ce faire, procédez comme suit :

- Numérisez ou ouvrez le document.
- Dans le menu Fichier, cliquez sur **Sauvegarder les graphiques**.

- Tous les graphiques du document seront sauvegardés dans un même fichier.

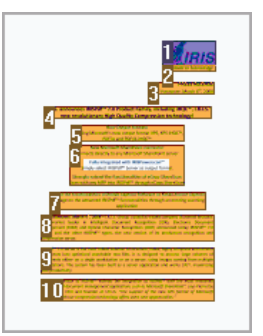
CHAPITRE 8
FENÊTRAGE DES DOCUMENTS

FENÊTRAGE AUTOMATIQUE DES DOCUMENTS

Lorsque vous numérisez ou ouvrez des documents, Readiris exécute automatiquement la fonction **Analyse de page** pour fractionner les documents en plusieurs fenêtres.

L'option **Analyse de page** est activée par défaut. Cliquez sur le bouton **Options** et désactivez **Analyse de page** si vous souhaitez éviter l'analyse de page automatique.

Les résultats de l'analyse de page peuvent être modifiés manuellement après l'analyse de page automatique. Pour plus d'informations, consultez la rubrique **Fenêtrage manuel des documents**.



Types de fenêtre

Readiris utilise cinq types de fenêtre : **blocs de texte**, **zones d'image**, **tableaux**, **zones de codes à barres** et **zones manuscrites**.

L'analyse de page détecte automatiquement les zones de texte, d'image et de tableau. Les zones manuscrites et de codes à barres doivent être délimitées manuellement.

> Pour plus d'informations, consultez la rubrique **Fenêtrage manuel des documents**.

Chaque type de fenêtre possède son propre code couleur : les fenêtres de texte sont en orange, les images en mauve et les tableaux en rose. Les zones de codes à barres sont vertes et les zones manuscrites sont bleues.

Les fenêtres sont triées de haut en bas et de gauche à droite. Les numéros indiquent l'ordre de tri des fenêtres. Il est toutefois possible de modifier l'ordre de tri et les types de fenêtre. Pour plus d'informations, consultez la rubrique **Fenêtrage manuel des documents**.

Ignorer les zones de texte en bordure de page

Lorsque votre numériseur génère des bordures noires autour de l'image, l'analyse de page a tendance à voir des zones là où il n'y a que du bruit.

Pour l'éviter, cliquez sur **Ne pas détecter de fenêtre sur les bords** dans le menu **Configuration** et renumérisez le document.

Ignorer la zone extérieure

En guise d'alternative au fenêtrage automatique des documents, il est possible d'utiliser la fonction **Ignorer la zone extérieure**. Cette fonction est utile lorsqu'une seule zone spécifique doit être reconnue dans les pages du document.

Sélectionnez **Ignorer la zone extérieure** en cliquant sur le bouton correspondant dans la barre d'outils Image.

Dessinez un cadre autour de la partie de la page que Readiris doit reconnaître.

La partie de la page sélectionnée est automatiquement analysée.

Vous êtes invité à préciser si vous souhaitez exclure la même zone extérieure de l'analyse de page pour chaque page du document.

FENÊTRAGE MANUEL DES DOCUMENTS

En plus du fenêtrage automatique des documents réalisé à l'aide de l'analyse de page, Readiris permet d'appliquer un fenêtrage manuel aux documents.

Le fenêtrage manuel est pratique lorsque vous devez modifier les résultats de l'analyse de page automatique. Il est également utile pour la création de modèles de fenêtrage.

Pour plus d'informations sur les modèles de fenêtrage, consultez la rubrique **Utilisation des modèles de fenêtrage**.

Notez que les zones de texte manuscrit et de codes à barres doivent toujours être délimitées manuellement.

Procédure

- Pour appliquer un fenêtrage manuel à un document, commencez par cliquer sur le bouton **Options** puis désactivez **Analyse de page**.
- Ouvrez ou numérisez le document en cliquant sur le bouton **Numériser** ou **Ouvrir**.

- Sélectionnez le **type de fenêtre** correct en cliquant sur le type correspondant dans la barre d'outils Image.

  Readiris utilise cinq types de fenêtre : **blocs de texte**, **zones d'image**, **tableaux**, **zones de codes à barres** et **zones manuscrites**.

- Dessinez un cadre autour des blocs de texte, images, tableaux, codes à barres et zones de texte manuscrit auxquels vous souhaitez appliquer un fenêtrage.

  Pour plus d'informations sur la reconnaissance des codes à barres et du texte manuscrit, consultez les rubriques **Reconnaissance des codes à barres** et **Reconnaissance de texte manuscrit**.

- Lorsque le fenêtrage est terminé, cliquez sur **Reconnaître + Enregistrer** pour lancer la reconnaissance.

Tri des fenêtres

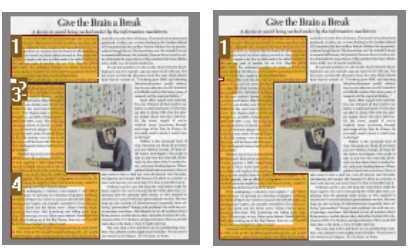
Pour modifier l'ordre de tri des fenêtres, cliquez sur le bouton **Trier** dans la barre d'outils Image et cliquez successivement sur les différentes fenêtres dans l'ordre voulu.

Après quoi, cliquez sur **Reconnaître + Enregistrer** pour lancer la reconnaissance.

Les fenêtres non sélectionnées sont exclues de la reconnaissance.

Dessin de polygones

Le fenêtrage manuel de documents n'est pas limité à des formes rectangulaires. Vous pouvez créer des fenêtres polygonales en fusionnant des fenêtres rectangulaires. Chaque fois que deux fenêtres du même type se touchent, elles deviennent automatiquement un polygone.



Analyse de page automatique

Si la page active est trop complexe pour un fenêtrage manuel, cliquez sur le bouton **Analyser la page** dans la barre d'outils Image afin de réaliser un fenêtrage automatique de la page.

Notez que les zones de texte manuscrit et de codes à barres doivent toujours être délimitées manuellement.

Modification du type de fenêtre

Pour modifier le type d'une fenêtre, cliquez avec le bouton droit sur celle-ci, pointez sur **Fenêtre**, puis sur **Type** et enfin cliquez sur le type de fenêtre voulu.

Vous pouvez également changer simultanément le type de plusieurs fenêtres :

- Cliquez sur le bouton du **pointeur** dans la barre d'outils Image.

- Tout en maintenant la touche **Maj** enfoncée, sélectionnez plusieurs fenêtres.
- Cliquez avec le bouton droit sur l'une des fenêtres sélectionnées, pointez sur **Fenêtre**, puis sur **Type** et enfin cliquez sur le type de fenêtre voulu.

Modification de la taille de la fenêtre

Cliquez sur la fenêtre à modifier.

Placez le pointeur de la souris sur un marqueur (sur les côtés et dans les angles de la fenêtre).

Cliquez sur le marqueur et faites glisser la souris afin de modifier la taille de la fenêtre.

Déplacement des fenêtres

Sélectionnez la fenêtre à déplacer.

Cliquez à l'intérieur de la fenêtre et faites glisser la souris afin de modifier la position de la fenêtre.

Reconnaissance d'une fenêtre spécifique

Cliquez avec le bouton droit sur la fenêtre à reconnaître et sélectionnez **Copier comme texte**.

Les résultats sont envoyés au Presse-papiers en tant que texte continu. Cette commande fonctionne également pour le texte manuscrit.

Il est également possible de copier les fenêtres d'images et de codes à barres dans le Presse-papiers.

Suppression de fenêtres

Sélectionnez la ou les fenêtres à supprimer ou cliquez sur la commande **Sélectionner tout** dans le menu **Edition** afin de sélectionner toutes les fenêtres.

Sélectionnez les commandes **Couper** ou **Effacer** dans le menu **Edition** pour couper ou supprimer les fenêtres.

ou

Cliquez avec le bouton droit sur les fenêtres sélectionnées, pointez sur **Fenêtre**, puis cliquez sur **Supprimer**.

Suppression de petites fenêtres

Certains documents, les télécopies par exemple, comportent des points « dispersés » sur les pages, ce qui conduit Readiris à créer des fenêtres inutiles qui ne contiennent pas de texte.

Pour effacer toutes les petites fenêtres, cliquez sur **Supprimer les petites fenêtres** dans le menu **Edition**.

Cette option efface toutes les fenêtres inférieures à 0,5" et retrie les zones restantes.

UTILISATION DES MODÈLES DE FENÊTRAGE

Lorsque vous effectuez une reconnaissance d'un grand nombre de nombreux documents présentant une mise en page similaire, il peut être utile d'utiliser les **modèles de fenêtrage** au lieu de l'analyse de page. De cette façon, la même structure de fenêtres est appliquée à tous les documents numérisés ou ouverts, ce qui permet d'accélérer le processus.

Procédure

- Procédez à un fenêtrage manuel de la première page du document à l'aide des boutons de la barre d'outils Image.

  Pour plus d'informations, consultez la rubrique **Fenêtrage manuel des documents**.

- Dans le menu **Fichier**, cliquez sur la commande **Sauvegarder le fenêtrage**.
- Ouvrez ou numérisez les autres pages du document en cliquant sur le bouton **Numériser** ou **Ouvrir** dans la barre d'outils principale.
- Dans le menu **Fichier**, cliquez sur la commande **Charger un fenêtrage**.
- Sélectionnez le fichier de fenêtrage que vous avez enregistré.
- Pour appliquer le fenêtrage à **toutes** les pages ouvertes ou numérisées, sélectionnez l'option **Appliquer fenêtrage à toutes les pages** dans la boîte de dialogue **Fenêtrage**.
- Cliquez sur **Ouvrir** pour charger le fichier de fenêtrage.

  Remarque : lorsque vous ajoutez un document à Readiris, le fichier de fenêtrage doit être rechargé puisque l'analyse de page est activée par défaut.

Ignorer la zone extérieure

En guise d'alternative au fenêtrage automatique des documents, il est possible d'utiliser la fonction **Ignorer la zone extérieure.** De cette façon, vous pouvez définir la zone spécifique de la page qui doit être reconnue. Toutes les données à l'extérieur de la zone d'OCR sont exclues de la reconnaissance.

Procédure

- Sélectionnez **Ignorer la zone extérieure** en cliquant sur le bouton correspondant dans la barre d'outils Image.

- Dessinez un cadre autour de la partie de la page que Readiris doit reconnaître.

  La partie de la page sélectionnée est automatiquement analysée.

  Le système vous invite à préciser si vous souhaitez ignorer la même zone extérieure pour toutes les pages du document.

  Pour annuler cette fonction, réexécutez l'**analyse de page** en cliquant sur le bouton **Analyser la page** dans la barre d'outils Image.

- Cliquez sur **Reconnaître + Enregistrer** pour exécuter l'OCR.

CHAPITRE 9
RECONNAISSANCE DES DOCUMENTS

INTRODUCTION

Pour reconnaître des documents, Readiris applique une analyse linguistique au cours de l'étape de reconnaissance. En conséquence, il reconnaît du texte, des tableaux, des images, des codes à barres et du texte manuscrit dans tous les types de documents. Readiris traite même les documents disposés en colonnes complexes ou de piètre qualité, des télécopies, des impressions matricielles, des documents mal numérisés ou copiés contenant des caractères trop clairs ou trop foncés, etc.

Readiris prend en charge 128 **langues** : toutes les langues d'Amérique et d'Europe, y compris les langues d'Europe centrale, les langues baltes, le grec, le turc et le cyrillique. Grâce à des modules complémentaires, Readiris peut lire des documents en arabe, persan et hébreu et dans quatre langues asiatiques, à savoir le japonais, le coréen, le chinois simplifié et le chinois traditionnel. Readiris peut même traiter des alphabets mélangés : le logiciel détecte les mots « occidentaux » qui ressortent des documents en grec, en cyrillique, en langue asiatique, en hébreu et en arabe puisque beaucoup de noms propres intranscriptibles, des noms de marques etc., sont écrits à l'aide des symboles occidentaux.

Readiris repose sur les technologies de reconnaissance les plus évoluées. La reconnaissance du texte indépendante de la police est complétée par des techniques d'auto-apprentissage. Le système est

capable d'apprendre de nouveaux mots et caractères grâce à une analyse contextuelle et linguistique. En d'autres termes, la précision de l'OCR s'améliore avec le temps et l'utilisation.

En outre, Readiris offre une fonction de vérification utilisateur (facultative). Lorsqu'elle est activée, la fonction de vérification utilisateur (**Apprentissage interactif**) marque les caractères considérés comme douteux par le système de reconnaissance et permet également d'améliorer la précision du système. Toutes les solutions que vous validez sont enregistrées temporairement dans la mémoire du système, ce qui améliore la vitesse et la fiabilité du système et accroît son intelligence. Ce puissant outil d'apprentissage interactif permet également à Readiris de reconnaître des caractères spéciaux, notamment les symboles mathématiques et les signes typographiques et de traiter les polices déformées.

Les résultats de l'apprentissage interactif peuvent être enregistrés de façon permanente dans des **dictionnaires de polices** pour utilisation ultérieure.

Pour améliorer la précision de la reconnaissance, une autre méthode consiste à utiliser les lexiques utilisateur. Vous pouvez créer des lexiques utilisateur spécialisés contenant la terminologie spécifique que Readiris doit pouvoir reconnaître.

SÉLECTION DE LA LANGUE DU DOCUMENT

Readiris offre des fonctionnalités d'OCR dans 128 langues. Il prend en charge toutes les langues d'Amérique et d'Europe, y compris les langues d'Europe centrale, les langues baltes, le grec, le turc et le cyrillique.

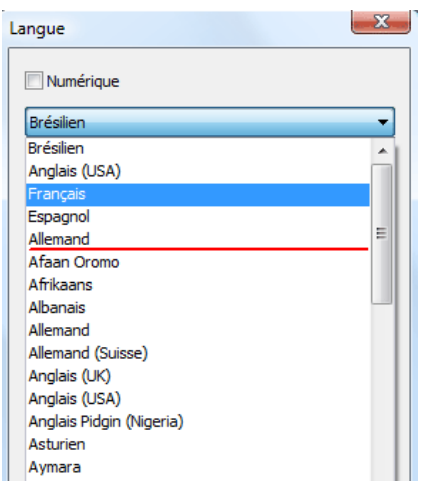
Readiris Pro Asian et Readiris Corporate Asian permettent en outre de reconnaître des documents en japonais, en chinois simplifié et traditionnel ainsi qu'en coréen.

Readiris Pro Middle-East et Readiris Corporate Middle-East reconnaissent des documents en arabe, en persan et en hébreu.

Pour que Readiris puisse reconnaître un document, vous devez préciser sa langue.

Pour ce faire, procédez comme suit :

Cliquez sur le bouton représentant un **globe** dans la barre d'outils principale et sélectionnez la langue de votre choix dans la liste déroulante.



Les 5 langues récemment sélectionnées sont placées en haut de la liste.

Important : sélectionnez la langue du document *avant* d'exécuter l'analyse de page lorsqu'il s'agit de documents asiatiques, hébreux ou arabes. Des routines d'analyse de page spécifiques sont utilisées pour ces documents.

La reconnaissance peut également être limitée à un jeu de caractères **numérique** pour optimiser la reconnaissance des tableaux et des chiffres. Readiris reconnaît alors uniquement les chiffres entre 0 et 9 ainsi que les symboles suivants :

+ signe plus	* astérisque	/ barre oblique	% pourcentage	, virgule
. point	(parenthèse d'ouverture	) parenthèse de fermeture	- tiret	= signe égal
$ dollar	£ livre sterling	€ euro	¥ yen	

Pour activer le mode numérique, sélectionnez **Numérique** dans la boîte de dialogue **Langue**.

Reconnaissance de documents multilingues

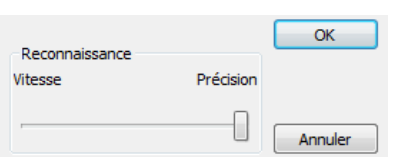
Readiris permet également d'activer la reconnaissance de jeux de caractères mixtes. De cette façon, Readiris change automatiquement de langue en milieu de phrase et reconnaît les mots anglais (noms propres, etc.) conservés dans des langues dites « exotiques ».

Cliquez sur le bouton représentant un **globe** dans la barre d'outils principale et sélectionnez la combinaison de langues de votre choix dans la liste déroulante.

Remarque : lors du traitement de documents en arabe ou en hébreu, les jeux de caractères mixtes sont automatiquement utilisés.

Vitesse/Précision

Choisissez le bon compromis entre **la vitesse d'OCR** et **la précision d'OCR**.



Ce compromis est disponible pour les alphabets latin, cyrillique et grec.

Conseil : privilégiez la précision plutôt que la vitesse lorsque l'image n'est pas de bonne qualité.

DÉFINITION DES CARACTÉRISTIQUES DU DOCUMENT

Outre la langue du document, d'autres caractéristiques de document, la **police** et le **pas de caractère**, jouent notamment un rôle important dans le processus de reconnaissance.

Police

Readiris distingue les documents dits « normaux » et les documents d'impression matricielle. Les symboles matriciels (du type 9 aiguilles) sont composés de points distincts.

Des techniques de segmentation et de reconnaissance spéciales sont nécessaires pour reconnaître les documents d'impression matricielle et doivent être activées.

Far out in the uncharted back

Pour sélectionner la police :

- Dans le menu **Configuration**, pointez sur **Police**.
- Par défaut, la police a la valeur **Automatique**.

De cette façon, Readiris reconnaît les impressions matricielles « 25 aiguilles » ou « NLQ » (qualité quasi-courrier) ou d'autres types d'impression « normale ».

- Pour reconnaître uniquement les documents d'impression matricielle, cliquez sur **A aiguilles**.

  Readiris reconnaît les documents d'impression matricielle de qualité « brouillon » ou à « 9 aiguilles ».

Pas des caractères

Le pas des caractères est le nombre de caractères par pouce dans un type de caractère. Le pas des caractères peut être **fixe**, auquel cas tous les caractères ont la même largeur, ou **proportionnel**, auquel cas leur largeur est différente.

Pour sélectionner le pas des caractères :

- Dans le menu **Configuration**, pointez sur **Pas des caractères**.
- Par défaut, le pas des caractères a la valeur **Automatique**.
- Cliquez sur **Fixe** si tous les caractères ont la même largeur. C'est souvent le cas dans les documents dactylographiés avec d'anciennes machines à écrire.
- Cliquez sur **Proportionnel** si les caractères ont une largeur différente. La vaste majorité des polices utilisées dans les journaux, magazines et livres est proportionnelle.

Important: ces caractéristiques de document ne s'appliquent pas aux documents asiatiques, hébreux ou arabes.

UTILISATION DE L'APPRENTISSAGE INTERACTIF

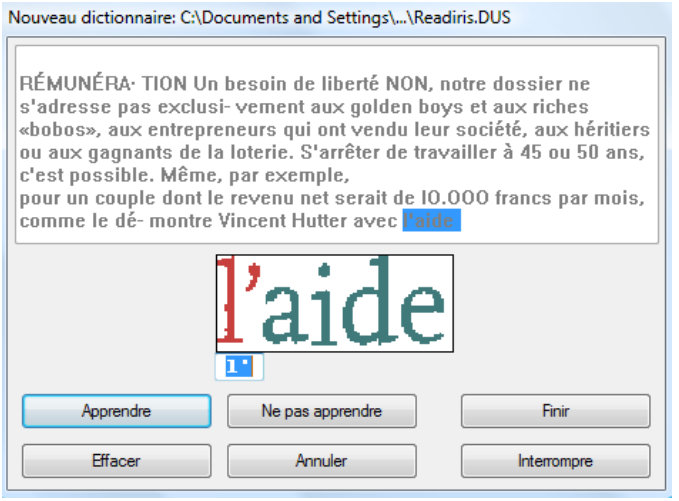
Readiris propose une fonction d'apprentissage interactif. Grâce à l'**apprentissage interactif**, vous pouvez « former » le système à reconnaître les polices et les formes des caractères et, le cas échéant, corriger les résultats de l'OCR. Lors de cette phase d'apprentissage interactif, tous les caractères non formellement reconnus par le système sont affichés dans une fenêtre d'aperçu, accompagnés de leur mot parent et de la solution proposée.

Cet apprentissage peut sensiblement améliorer la précision du système de reconnaissance et est particulièrement utile pour la reconnaissance des caractères dégradés et déformés. Il permet également d'apprendre à Readiris à reconnaître les symboles spéciaux qu'il n'arrive pas à identifier de prime abord, notamment les symboles mathématiques et scientifiques ainsi que les signes typographiques.

Pour activer l'apprentissage interactif :

- Dans le menu **Apprendre**, cliquez sur **Apprentissage interactif**.
- Définissez les paramètres requis puis cliquez sur **Reconnaître + Enregistrer** pour lancer la reconnaissance du document.
- A la fin de la reconnaissance, Readiris passe en phase d'apprentissage interactif.

  Les caractères jugés douteux par le système de reconnaissance sont affichés.



Si les résultats sont corrects :

- Cliquez sur le bouton **Apprendre** pour enregistrer les résultats confirmés.

  Les résultats de l'apprentissage sont stockés temporairement dans la mémoire de l'ordinateur pour la durée de la reconnaissance. Readiris n'affiche plus les caractères préalablement « appris » lorsqu'il effectue la reconnaissance du reste du document.

  Lorsqu'un nouveau document fait l'objet d'une reconnaissance, les résultats de l'apprentissage sont effacés.

  Pour enregistrer définitivement les résultats de l'apprentissage, utilisez un dictionnaire de polices. Pour plus d'informations, consultez la rubrique **Utilisation des dictionnaires de polices**.

- Cliquez sur **Finir** pour enregistrer toutes les solutions proposées par le logiciel.

Si les résultats ne sont pas corrects :

- Entrez les caractères corrects et cliquez sur le bouton **Apprendre**.

  ou

- Cliquez sur le bouton **Ne pas apprendre** pour enregistrer les résultats comme étant douteux.

  Utilisez cette commande pour les caractères endommagés qui pourraient être confondus avec d'autres caractères s'ils étaient enregistrés comme étant confirmés, par exemple, le nombre 1 et la lettre I, qui présentent une forme identique dans de nombreuses polices.

- Cliquez sur **Supprimer** pour supprimer les caractères de la sortie.

  Utilisez ce bouton pour éviter que du bruit n'apparaisse dans le fichier de sortie.

- Cliquez sur **Annuler** pour corriger des erreurs.

  Readiris peut assurer le suivi des 32 dernières opérations.

- Cliquez sur **Interrompre** pour interrompre l'apprentissage interactif.

Tous les résultats de l'apprentissage seront supprimés. La prochaine fois que vous cliquez sur **Reconnaître + Enregistrer**, l'**apprentissage interactif** est réexécuté.

UTILISATION DES DICTIONNAIRES DE POLICES

Lorsque vous numérisez de nombreux documents de même type, qualité de police et qualité d'impression, vous souhaitez parfois éviter de répéter le processus d'apprentissage. Dans un tel cas, l'emploi de dictionnaires de polices s'avère utile. Les dictionnaires de polices contiennent des informations sur les polices apprises lors de l'**apprentissage interactif** et peuvent améliorer sensiblement les résultats de la reconnaissance.

Notez que les dictionnaires de polices sont limités à 500 formes. Il est recommandé de créer des dictionnaires distincts pour des applications spécifiques.

Pour créer un nouveau dictionnaire de polices :

- Dans le menu **Apprendre**, cliquez sur la commande **Nouveau dictionnaire**.

  Readiris ouvre par défaut le dictionnaire **Readiris.dus**. Modifiez le nom de fichier et cliquez sur **Enregistrer** pour l'enregistrer comme un nouveau dictionnaire.

- Cliquez sur **Apprentissage interactif** dans le menu **Apprendre** pour l'activer.

- Cliquez sur **Reconnaître + Enregistrer** pour reconnaître le document.

- Readiris passe en phase d'apprentissage interactif. Utilisez les boutons de la boîte de dialogue pour enregistrer les caractères dans le dictionnaire de polices.

Pour utiliser un dictionnaire de polices existant :

- Dans le menu **Apprendre**, cliquez sur **Dictionnaire de polices**.

- Sélectionnez le dictionnaire de votre choix et cliquez sur **Ouvrir**.

- Dans le menu **Apprendre**, cliquez sur **Enrichir le dictionnaire** ou **Lire le dictionnaire**.

  Lorsque vous sélectionnez **Enrichir le dictionnaire**, veillez à activer **Apprentissage interactif**. Readiris reconnaît les formes des caractères stockés dans le dictionnaire et utilise l'apprentissage interactif, vous permettant ainsi d'enregistrer de nouvelles informations dans le dictionnaire.

  Lorsque vous sélectionnez l'option **Lire le dictionnaire**, Readiris reconnaît les formes des caractères stockés dans le dictionnaire mais n'ajoute pas de nouveau contenu à ce dernier, même si l'option **Apprentissage interactif** est activée. Notez qu'il est toujours utile d'avoir recours à l'**apprentissage interactif** pour vérifier et le cas échéant, corriger les résultats de la reconnaissance qui ne sont pas enregistrés dans le dictionnaire de polices.

Attention : ne cliquez pas sur **Dictionnaire de polices** dans le menu **Apprendre** pour ouvrir un dictionnaire existant lorsque le mode **Nouveau dictionnaire** est activé. Sans quoi, le contenu du dictionnaire existant sera effacé.

- Cliquez sur **Reconnaître + Enregistrer** pour reconnaître le document.

CHAPITRE 10
FORMATAGE ET ENREGISTREMENT DE DOCUMENTS

FORMATAGE DE DOCUMENTS

Les documents sélectionnés pour l'OCR dans Readiris peuvent être enregistrés dans divers formats de sortie. Readiris enregistre les résultats de l'OCR en tant que fichiers Adobe Acrobat PDF, Microsoft XPS, fichiers texte Word, WordML, RTF et OpenDocument, fichiers HTML et XML, feuilles de calcul SpreadsheetML et fichiers texte Ansi et Unicode.

Par ailleurs, Readiris peut exporter directement les résultats dans des applications de destination telle que Microsoft Word et Excel, Adobe Reader, Microsoft XPS Viewer, les navigateurs Web et applications de messagerie les plus connus, etc.

> Remarque : les fichiers RTF créés par Readiris peuvent être également ouverts dans des applications bureautiques Web, telles que AjaxWrite, ThinkFree, Zoho Writer et Google Writely (qui ouvre à la fois les documents RTF et HTML). Lorsque vous utilisez l'une de ces applications, pensez à sélectionner l'option de mise en page **Retenir le formatage des mots et des paragraphes**.

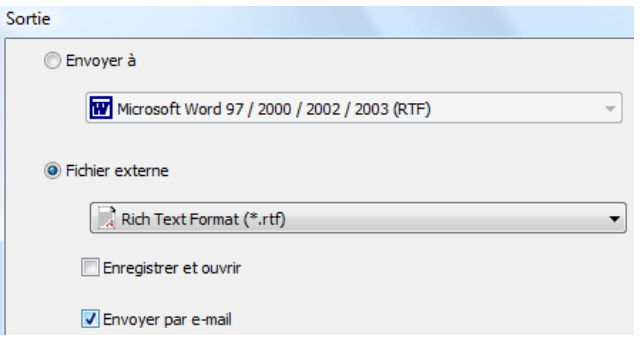
Procédure

- Cliquez sur le bouton **Format** dans la barre d'outils principale pour sélectionner le mode de sortie.

  Readiris exécute l'une des deux actions suivantes :

- envoyer des documents à une application, laquelle s'ouvre automatiquement, ou
- enregistrer les documents en tant que fichiers externes.

L'option **Envoyer par e-mail** crée un nouveau message e-mail et insère le document reconnu en tant que pièce jointe.



- Cliquez sur les différents onglets afin de sélectionner les options à appliquer.

Les options indisponibles pour le format de sortie sélectionné sont affichées en grisé.

Les formats de fichier de sortie les plus fréquemment utilisés ainsi que leurs options sont présentés en détail dans les sections suivantes.

- Cliquez sur **Reconnaître + Enregistrer** pour reconnaître et enregistrer le document.

Les résultats d'OCR peuvent être exportés à plusieurs reprises sans devoir répéter la reconnaissance. Cliquez de nouveau sur le bouton **Format** et modifiez le format du texte ainsi que les options de formatage. Après quoi, cliquez à nouveau sur **Reconnaître + Enregistrer**.

Pour des raisons de tri et de recherche, Readiris vous permet de définir les **propriétés de document** des fichiers de sortie PDF, XPS, Word, RTF, WordML, SpreadsheetML et HTML. Pour définir les propriétés

d'un document, cliquez sur l'option **Propriétés du document** du menu **Fichier**. Remarques : les options relatives aux propriétés de document sont également accessibles dans la boîte de dialogue **Fichier de sortie** qui est ouvert lorsque vous cliquez sur l'option **Reconnaître + Enregistrer**.

Lorsque vous enregistrez un document de plusieurs pages en tant que **fichier externe**, vous pouvez créer, dans Readiris, un fichier de sortie distinct pour chaque page ou enregistrer toutes les pages d'un même document dans un fichier de sortie unique. Il suffit de cliquer sur les options correspondantes dans la boîte de dialogue **Fichier de sortie**, à savoir **Créer un fichier par page** ou **Créer un fichier par document**.

Notez toutefois que les options **Créer un fichier par page** et **Créer un fichier par document** sont disponibles uniquement lors de l'enregistrement des documents en tant que fichier externe mais pas lorsque vous ouvrez des documents dans une application cible.

FORMATAGE DE DOCUMENTS TEXTE

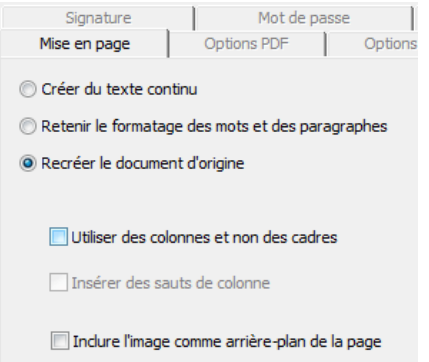
Readiris vous permet de générer plusieurs formats de sortie de type texte. Il peut notamment générer des sorties Word, WordML, RTF, txt et OpenDocument Text.

Pour générer des fichiers de sortie texte :

- Cliquez sur le bouton **Format** dans la barre d'outils principale et sélectionnez le format de sortie de votre choix dans la liste déroulante **Envoyer à** ou **Fichier externe**.
- Selon le format texte sélectionné, plusieurs options de formatage sont disponibles :

  Les options non disponibles pour le format de sortie sélectionné sont affichées en grisé.

Options de mise en page



- L'option **Créer du texte continu** empêche Readiris d'appliquer un formatage au texte.

  Readiris génère un texte continu.

- L'option **Retenir le formatage des mots et des paragraphes** représente une solution intermédiaire entre la génération d'un texte continu et le formatage automatique.

  Le type de police, la taille et le type de caractère sont préservés lors de la reconnaissance.

  Les tabulations et l'alignement de chaque bloc sont recréés.

  Les colonnes et les blocs de texte ne sont pas recréés, les paragraphes se suivent simplement.

  Les tableaux sont reproduits correctement.

- L'option **Recréer le document d'origine** permet de générer un facsimilé du document d'origine.

  Readiris génère une copie fidèle du document d'origine qui n'est plus une image numérisée.

Il recrée également les liens hypertexte vers les sites web ou les adresses e-mail.

- L'option **Utiliser des colonnes et non des cadres** crée des documents en colonnes.

  Le texte disposé en colonnes est plus facile à modifier que les documents contenant plusieurs cadres : le flux du texte est continu d'une colonne à l'autre.

  Remarque : lorsque le système n'est pas en mesure de détecter les colonnes dans le document d'origine, ce mode de formatage utilise des cadres comme solution alternative.

- L'option **Insérer des sauts de colonne** insère un saut de colonne forcé à la fin de chaque colonne.

  Tout texte que vous modifiez, ajoutez ou supprimez reste à l'intérieur de sa colonne. Il ne continue jamais automatiquement dans une autre colonne.

  Conseil : désactivez cette option lorsque vous avez du texte continu en colonne. Vous garantirez ainsi un flux continu du texte d'une colonne à l'autre.

- L'option **Inclure l'image comme arrière-plan de la page** place l'image numérisée en arrière-plan de la page derrière le texte reconnu.

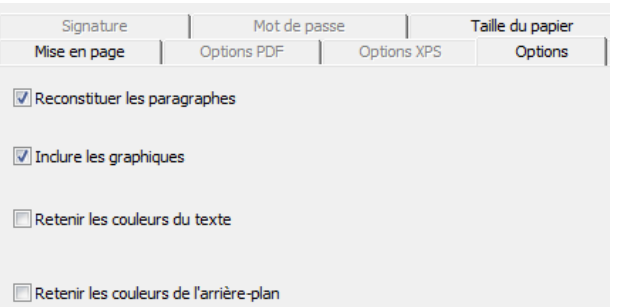
  Notez toutefois que cette option augmente sensiblement la taille des fichiers de sortie.

  Le format **PDF Texte-Image** offre un résultat identique pour les fichiers PDF.

  L'option **Retenir les couleurs de l'arrière-plan** de l'onglet **Options** offre une alternative plus compacte.

Options générales

Cliquez sur l'onglet **Options** pour sélectionner les options générales :



- L'option **Reconstituer les paragraphes** permet d'activer la détection automatique des paragraphes.

  Readiris exécute un retour à la ligne automatique du texte reconnu jusqu'au début d'un nouveau paragraphe et « recolle » les mots coupés à la fin d'une ligne.

- L'option **Inclure les graphiques** inclut les graphiques dans les fichiers formatés automatiquement.

  C'est indispensable pour créer une copie fidèle d'un document.

  Utilisez les options graphiques de l'onglet **Graphiques** pour déterminer la résolution et le mode couleurs des graphiques stockés dans les fichiers de sortie.

- L'option **Retenir les couleurs du texte** préserve les couleurs d'origine du texte lors de la reconnaissance.

- L'option **Retenir les couleurs de l'arrière-plan** préserve les couleurs d'accompagnement de l'arrière-plan de la page lors de la reconnaissance.

Une couleur d'arrière-plan uniforme est créée par paragraphe dans le fichier de sortie.

Tailles du papier

Selon le format sélectionné, vous pouvez spécifier les tailles de papier que vous préférez :

- Cliquez sur l'onglet **Taille du papier** et utilisez les flèches pour appliquer et exclure des tailles de papier.
- Readiris parcourt les tailles de papier actives dans l'ordre indiqué et utilise la première taille suffisamment grande pour contenir le document numérisé.

FORMATAGE DE DOCUMENTS COMPORTANT DES TABLEAUX

Readiris vous permet de générer des données tabulaires dans des feuilles de calcul, applications de traitement de texte et navigateurs web : les tableaux sont reconstitués cellule par cellule dans des feuilles de calcul et insérés en tant qu'objets tableau dans des fichiers de traitement de texte.

Readiris reconnaît les tableaux avec et sans quadrillage.

Performance test optical media	
CD-ROM Digital Versatile Disk	Average time (n
CD-ROM 32x speed	60
DVD	58
Tested on 333 MHz Pentium II	

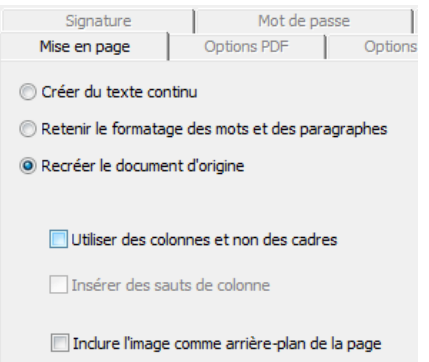
(avec quadrillage)

123.985	69	31
987.745	129	2
19	149	91
287.410	49	52

(sans quadrillage)

Pour générer des documents comportant des tableaux :

- Cliquez sur le bouton **Format** dans la barre d'outils principale et sélectionnez le format de sortie de votre choix dans la liste déroulante **Envoyer à** ou **Fichier externe**.
- Sélectionnez les options de mise en page requises.

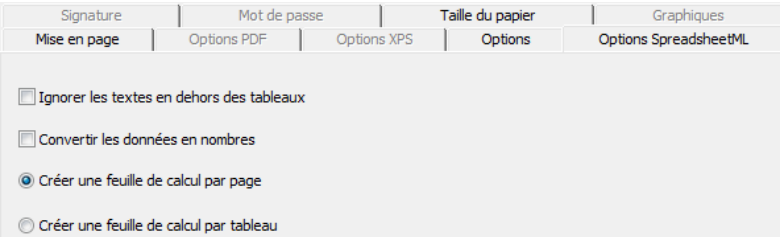


Pour plus d'informations sur les options de formatage, consultez la rubrique **Formatage de documents texte**.

Options SpreadsheetML

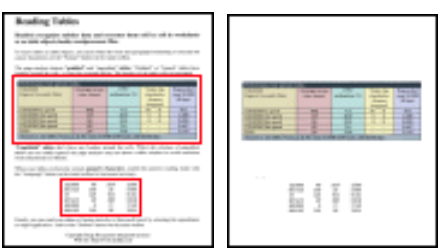
Lorsque vous sélectionnez **Microsoft Excel 2002/2003** en tant qu'application de destination, des options **SpreadsheetML** spécifiques sont disponibles. Cliquez sur l'onglet **Options SpreadsheetML** pour les afficher :

Notez que l'option de mise en page **Recréer le document d'origine** devient accessible lorsque ce format est sélectionné.



- L'option **Ignorer les textes en dehors des tableaux** enregistre les tableaux et ignore tous les autres résultats de la reconnaissance.

  Toutes les données figurant dans les tableaux sont recapturées mais pas celles en dehors des tableaux.



- L'option **Convertir les données en nombres** encode les données numériques reconnues en nombres.

  Par conséquent, vous pouvez effectuer des opérations arithmétiques sur ces cellules. Les cellules de texte (dans n'importe quel tableau) restent du texte.

  Notez que seules les données numériques au sein des tableaux sont encodées comme des nombres.

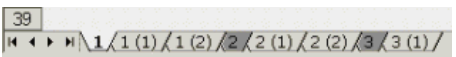
  Excel n'exécute des opérations mathématiques que sur les données encodées comme des nombres.

- Si l'option **Créer une feuille de calcul par page** est sélectionnée, une seule feuille de calcul est créée par page numérisée.

Si la page contient des tableaux et du texte, tout le contenu sera placé dans la même feuille de calcul.

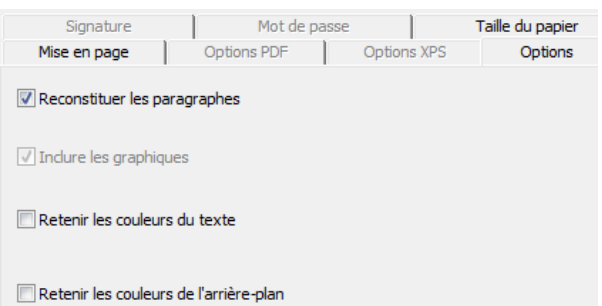
- L'option **Créer une feuille de calcul par tableau** place chaque tableau dans une feuille de calcul distincte et inclut le texte reconnu (en dehors des tableaux) dans une autre feuille de calcul.

  Si le document reconnu contient **plusieurs pages**,, cette structure est répétée par page.



Options générales

Cliquez sur l'onglet **Options** pour sélectionner les options générales :



- L'option **Reconstituer les paragraphes** permet d'activer la détection automatique des paragraphes.

  Readiris exécute un retour à la ligne automatique du texte reconnu jusqu'au début d'un nouveau paragraphe et « recolle » les mots coupés à la fin d'une ligne.

- L'option **Retenir les couleurs de l'arrière-plan** recrée la couleur d'arrière-plan de chaque cellule.

	A	B	C
1	Performance optical media		
2	CD-ROM	Average access	CPU
3	Digital Versatile Disk	time (msec)	utilization (%)
4	CD-ROM 24x speed	80	58.2
5	CD-ROM 32x speed	60	72.1
6	DVD	58	78.9
7	Tested on 333 MHz Pentium II with 64 MB RAM and 4 GB HD		

Tailles du papier

Selon le format sélectionné, vous pouvez spécifier les tailles de papier que vous préférez :

- Cliquez sur l'onglet **Taille du papier** et utilisez les flèches pour appliquer et exclure des tailles de papier.
- Readiris parcourt les tailles de papier actives dans l'ordre indiqué et utilise la première taille suffisamment grande pour contenir le document numérisé.

CRÉATION DE DOCUMENTS PDF

Readiris génère quatre types de sortie PDF : **texte**, **texte-image**, **image-texte** et **image**.

Pour générer des fichiers de sortie au format PDF :

Cliquez sur le bouton **Format** dans la barre d'outils principale et sélectionnez le type de PDF de votre choix dans la liste déroulante **Envoyer à** ou **Fichier externe** :

PDF Image

Lorsque vous sélectionnez PDF **Image**, Readiris génère des documents PDF comportant uniquement une image, il n'exécute pas l'OCR.

PDF Image-Texte

Lorsque vous sélectionnez l'option PDF **Image-Texte**, Readiris reconnaît le texte et crée des fichiers PDF indexés qui contiennent l'image de la page et le texte reconnu.

L'image de la page est placée au-dessus du texte.

Avec ce format, vous pouvez toujours voir le document d'origine (tel qu'il a été scanné) tout en ayant la possibilité de rechercher et de copier-coller le texte reconnu. Le texte est dissimulé derrière l'image. Ce format est utile dans le cadre de l'**archivage**.

PDF Texte

Lorsque vous sélectionnez l'option PDF **Texte**, Readiris reconnaît le texte et crée des fichiers PDF indexés.

L’image de la page n’est pas contenue dans ces fichiers PDF monocouches.

PDF Texte-Image

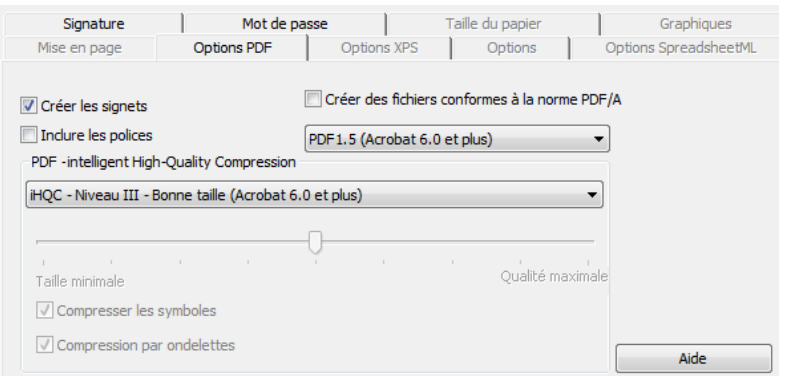
Lorsque vous sélectionnez l'option PDF **Texte-Image**, Readiris reconnaît le texte et crée des fichiers PDF indexés qui contiennent l'image de la page et le texte reconnu.

L'image de la page est placée derrière le texte.

SÉLECTION DES OPTIONS PDF

Pour sélectionner les options PDF :

- Cliquez sur le bouton **Format** dans la barre d'outils principale et sélectionnez le type de format PDF de votre choix dans la liste déroulante **Envoyer à** ou **Fichier externe**.
- Selon le **type de PDF** sélectionné, plusieurs options sont disponibles : cliquez sur l'onglet **Options PDF** pour y accéder.



Créer des signets

L'option **Créer des signets** permet de créer des signets pour chaque bloc de texte, graphique et tableau dans les fichiers PDF Adobe Acrobat.

Inclure les polices

Sélectionnez l'option **Inclure les polices** pour incorporer les polices dans les fichiers PDF Adobe Acrobat.

L'incorporation des polices permet d'empêcher toute substitution des polices et garantit que les lecteurs, indépendamment de la configuration de leur ordinateur, voient le texte avec ses polices d'origine.

Notez toutefois que cette option augmente quelque peu la taille de fichier des documents reconnus.

DOCUMENTS PDF COMPRESSÉS AVEC IHQC

Outre les quatre types de sortie PDF « normale », Readiris propose une sortie PDF compressée avec iHQC : les documents PDF de type **Image-Texte** et **Image** peuvent être hypercompressés à l'aide de la technologie iHQC sans dégradation aucune de la qualité de l'image.

iHQC est l'acronyme d'**intelligent High-Quality Compression**, la technologie d'hypercompression propriétaire d'I.R.I.S. iHQC est aux images ce que le MP3 est à la musique et ce que le DivX est aux films.

Pour générer des fichiers de sortie PDF compressés avec iHQC :

- Cliquez sur le bouton **Format** dans la barre d'outils principale pour sélectionner l'un de deux modes de sortie.
- Dans la liste **Envoyer à** ou **Fichier externe**, sélectionnez le type de format PDF de votre choix : **PDF Image-Texte** ou **PDF Image**.
- Dans l'onglet **Options PDF**, sélectionnez le niveau de compression requis.

  Readiris Pro prend en charge la compression **Niveau I - Bonne taille** et **Niveau I - Bonne qualité**.

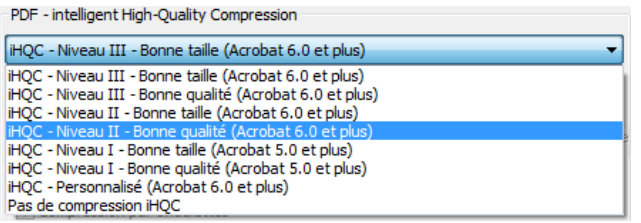
  Readiris Corporate prend également en charge la compression **Niveaux II et III Bonne taille** et **Bonne qualité** ainsi que l'option de compression **Personnalisé**.

  Dans la compression Niveau II, l'option **Compresser les symboles** est activée automatiquement pour compresser le texte de façon compacte.

  Dans la compression Niveau III, l'option **Compresser les ondelettes** est activée automatiquement pour compresser les

images de façon compacte.

Lorsque vous sélectionnez l'option de compression **Personnalisé**, vous pouvez activer ou désactiver ces options de façon indépendante. Vous pouvez également utiliser le curseur pour définir le taux de compression **Bonne taille - Bonne qualité**.



- Cliquez sur **Reconnaître + Enregistrer** pour reconnaître le document et l'enregistrer en tant que fichier PDF compressé iHQC.

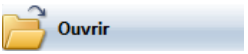
ADAPTATION DES DOCUMENTS PDF

En plus de générer des documents PDF, Readiris peut également adapter les fichiers PDF : Readiris convertit des PDF **image** en PDF **texte** ou tout autre format texte pris en charge et **déverrouille** le contenu PDF en lecture seule.

> **Avertissement** : Readiris n'ouvre pas les documents PDF protégés par mot de passe.

Procédure

- Cliquez sur le bouton **Ouvrir** dans la barre d'outils principale et sélectionnez le fichier PDF que Readiris doit adapter.



Dans la zone **Pages à convertir** de la boîte de dialogue, sélectionnez **Pages** et spécifiez les pages que Readiris doit ouvrir.

- Cliquez sur le bouton **Ouvrir** dans la boîte de dialogue pour ouvrir le fichier PDF de votre choix.
- Cliquez sur le bouton **Format** de la barre d'outils principale et sélectionnez le type de PDF requis.

  Pour plus d'informations sur les types de fichier PDF, consultez la rubrique **Création de documents PDF**.

- Cliquez sur le bouton **Reconnaître + Enregistrer** pour adapter le document.

CRÉATION DE DOCUMENTS XPS

Readiris génère quatre types de fichiers XPS : **texte**, **texte-image**, **image-texte** et **image**.

XPS est l'acronyme de XML Paper Specification. Il s'agir d'un format de document à mise en page fixe développé par Microsoft.

Pour générer des fichiers de sortie au format XPS :

Cliquez sur le bouton **Format** dans la barre d'outils principale et sélectionnez le type de fichier XPS PDF de votre choix dans la liste déroulante **Envoyer à** ou **Fichier externe** :

XPS Image

Lorsque vous sélectionnez XPS **Image**, Readiris génère des documents XPS comportant uniquement une image, il n'exécute pas l'OCR.

XPS Image-Texte

Lorsque vous sélectionnez l'option XPS **Image-Texte**, Readiris reconnaît le texte et crée des fichiers PDF indexés qui contiennent l'image de la page et le texte reconnu.

L'image de la page est placée au-dessus du texte.

Avec ce format, vous pouvez toujours voir le document d'origine (tel qu'il a été scanné) tout en ayant la possibilité de rechercher et de copier-coller le texte reconnu. Le texte est masqué par l'image. Ce format est utile dans le cadre de l'**archivage**.

XPS Texte

Lorsque vous sélectionnez l'option XPS **Texte**, Readiris reconnaît le texte et crée des fichiers XPS indexés.

L’image de la page n’est pas contenue dans ces fichiers XPS monocouches.

XPS Texte-Image

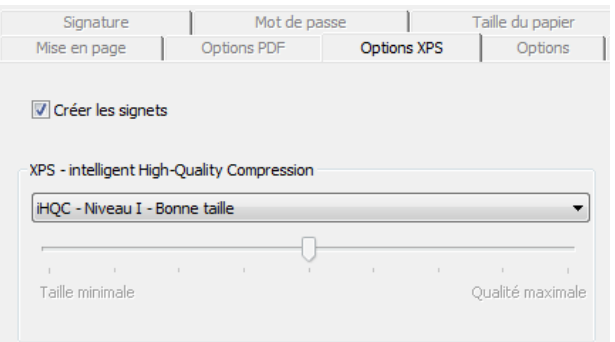
Lorsque vous sélectionnez l'option XPS Texte-Image, Readiris reconnaît le texte et crée des fichiers XPS indexés qui contiennent l'image de la page et le texte reconnu.

L'image de la page est placée derrière le texte.

SÉLECTION DES OPTIONS XPS

Pour sélectionner les options XPS :

- Cliquez sur le bouton **Format** dans la barre d'outils principale et sélectionnez le type de format XPS de votre choix dans la liste déroulante **Envoyer à** ou **Fichier externe**.
- Selon le **type de format XPS** sélectionné, plusieurs options sont disponibles : cliquez sur l'onglet **Options XPS** pour y accéder.



Créer des signets

L'option **Créer des signets** permet de créer des signets pour chaque bloc de texte, graphique et tableau dans les fichiers Microsoft XPS.

DOCUMENTS XPS COMPRESSÉS AVEC IHQC

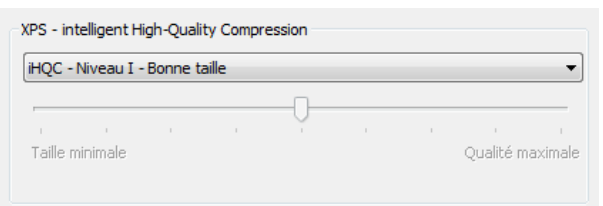
Outre les quatre types de sortie PDF « normale », Readiris propose une sortie XPS compressée avec iHQC : les documents XPS de type **Image-Texte** et **Image** peuvent être hypercompressés à l'aide de la technologie iHQC.

iHQC est l'acronyme d'**intelligent High-Quality Compression**, la technologie d'hypercompression propriétaire d'I.R.I.S. iHQC est aux images ce que le MP3 est à la musique et ce que le DivX est aux films.

Pour générer des fichiers de sortie XPS compressés avec iHQC :

- Cliquez sur le bouton **Format** dans la barre d'outils principale pour sélectionner l'un de deux modes de sortie.
- Dans la liste **Envoyer à** ou **Fichier externe**, sélectionnez le type de format PDF de votre choix : **XPS Image-Texte** ou **XPS Image**.
- Dans l'onglet **Options XPS**, sélectionnez le niveau de compression requis.

  Readiris prend en charge la compression **Niveau I - Bonne taille** et **Niveau I - Bonne qualité**.



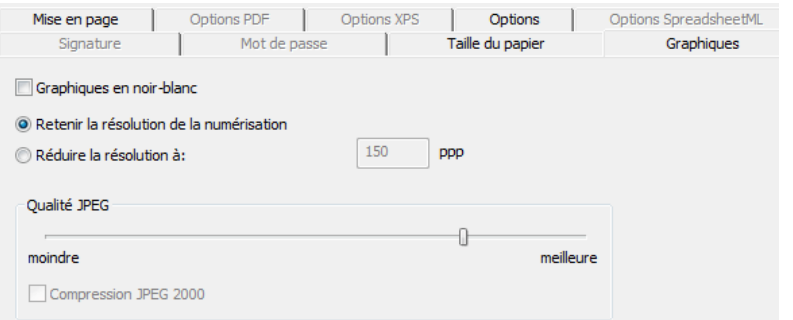
SÉLECTION DES OPTIONS DES GRAPHIQUES

Selon le format de sortie et l'application cible sélectionnés, vous pouvez accéder à certaines options de graphiques avancées. Les options des graphiques peuvent servir à modifier la qualité et la résolution de l'image.

Pour accéder aux options des graphiques :

- Cliquez sur le bouton **Format** dans la barre d'outils principale et sélectionnez le format de sortie de votre choix dans la liste déroulante **Envoyer à** ou **Fichier externe**.

- Cliquez sur l'onglet **Graphiques** pour en afficher les options :



Mode couleur

Par défaut, Readiris enregistre les graphiques en **couleur**.

Sélectionnez l'option **Noir-blanc** pour enregistrer les graphiques en noir et blanc.

Résolution

Readiris **conserve la résolution de numérisation** par défaut.

Toutefois, vous pouvez décider de **diminuer la résolution**.

Conseil : lors de l'enregistrement de documents au format HTML en vue de leur publication sur un site Web, réduisez la résolution à 70 ppp (résolution d'écran).

Qualité JPEG

Les images stockées à l'intérieur des documents PDF, XPS, Word et RTF sont enregistrées au format JPEG.

Utilisez le curseur pour ajuster la qualité JPEG.

Compression JPEG 2000

Lors de l'enregistrement des fichiers au format PDF ou XPS, Readiris peut appliquer une compression JPEG 2000 aux images en couleur et à niveaux de gris stockées dans ces fichiers.

JPEG 2000 est la nouvelle version plus compacte de la norme JPEG.

Sélectionnez l'option **Compression JPEG 2000** pour l'appliquer.

CHAPITRE 11
ENREGISTREMENT ET CHARGEMENT DES PARAMÈTRES

Tous les paramètres définis dans Readiris sont automatiquement enregistrés à des fins d'utilisation ultérieure lorsque vous fermez l'application. Pour rétablir la configuration d'usine, cliquez sur la commande **Rétablir la configuration d'usine** dans le menu **Fichier**.

Lorsque vous numérisez plusieurs groupes de documents qui exigent une configuration différente, il est utile d'enregistrer un **fichier de configuration** distinct pour chaque groupe.

Procédure

- Sélectionnez les paramètres à utiliser pour un groupe de documents donné.
- Dans le menu **Fichier**, cliquez sur la commande **Enregistrer la configuration**.
- Lorsque vous numérisez ou ouvrez par la suite un document du même groupe, cliquez sur la commande **Charger une configuration** du menu **Fichier**.
- Sélectionnez le fichier de configuration approprié et cliquez sur **Ouvrir**.
- Cliquez sur **Reconnaître + Enregistrer** pour reconnaître les documents à l'aide des paramètres adéquats.

Remarque : la commande **Infos** du menu **Fichier** permet d'afficher un résumé des principaux paramètres sélectionnés.

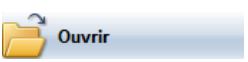
CHAPITRE 12
RECONNAISSANCE DES DOCUMENTS DE PLUSIEURS PAGES

OUVERTURE ET RECONNAISSANCE DE PLUSIEURS FICHIERS IMAGE

Readiris est conçu pour traiter plusieurs fichiers image à la fois.

Pour ouvrir plusieurs fichiers image :

- Cliquez sur **Ouvrir** dans la barre d'outils principale.



- Sélectionnez les fichiers image de votre choix :
    - Sélectionnez le premier fichier image puis, tout en maintenant la touche **Ctrl** enfoncée, sélectionnez d'autres images, ou ;
    - Sélectionnez une plage contiguë de fichiers image en sélectionnant la première image et la dernière tout en maintenant la touche **Maj** enfoncée.

Remarque : lorsque vous ouvrez un seul fichier de plusieurs pages, par exemple un fichier TIFF multipage ou un document PDF, vous pouvez spécifier la plage de pages. Dans la zone **Pages à convertir** de la boîte de dialogue, sélectionnez **Pages** et spécifiez les pages que vous souhaitez reconnaître.

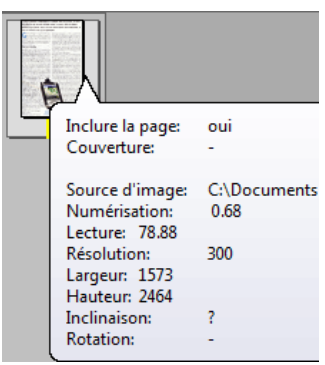
- Cliquez sur le bouton **Ouvrir** pour ouvrir l'image ou les images.

  Remarque : vous pouvez également faire glisser-déposer les fichiers image de l'Explorateur Windows vers la fenêtre des images de Readiris pour les ouvrir.

- La barre d'outils des pages affiche les fichiers image ouverts.

  Conseil : faites passer le curseur de la souris sur les miniatures des pages afin d'afficher leurs informations de configuration.

  La barre d'outils des pages permet de modifier des documents de plusieurs pages. Pour plus d'informations, consultez la rubrique **Modification des documents de plusieurs pages**.



- Spécifiez les paramètres de reconnaissance puis cliquez sur **Reconnaître + Enregistrer** pour exécuter la reconnaissance.

- Si vous souhaitez ouvrir ou numériser d'autres images dans le document actif, cliquez sur le bouton **Numériser** ou **Ouvrir** dans la barre d'outils principale.

  Le système vous invite à préciser si vous souhaitez supprimer le document actif ou non. Cliquez sur **Oui** pour supprimer le document actif et en créer un nouveau ou cliquez sur **Non** pour ajouter d'autres numérisations au document actif.

NUMÉRISATION ET RECONNAISSANCE DES DOCUMENTS DE PLUSIEURS PAGES

Readiris est conçu pour traiter des documents comportant plusieurs pages.

> Readiris Pro est en mesure de traiter des documents pouvant compter jusqu'à 50 pages.
>
> Readiris Corporate peut traiter des documents comptant un nombre illimité de pages.

Pour numériser des documents de plusieurs pages dans Readiris, vous pouvez utiliser la fonction the **Alimentation automatique de document** (avec un scanner à feuilles) ou la fonction **Numérisation avec pause** (avec un scanner à plat).

Numérisation de documents de plusieurs pages avec un scanner à feuilles

- Cliquez sur le bouton **Numériseur** de la barre d'outils principale et sélectionnez l'option **ADF** (alimentation automatique de documents).
- Placez les pages dans le bac d'alimentation du numériseur et cliquez sur **Numériser** pour lancer la numérisation.
- Cliquez sur **Reconnaître + Enregistrer** pour reconnaître les documents.

Numérisation de documents de plusieurs pages à l'aide de l'option de numérisation avec pause (scanner à plat)

- Cliquez sur le bouton **Numériseur** dans la barre d'outils principale.
- Sélectionnez **Numériser la page suivante après** et spécifiez un délai à l'aide des flèches.

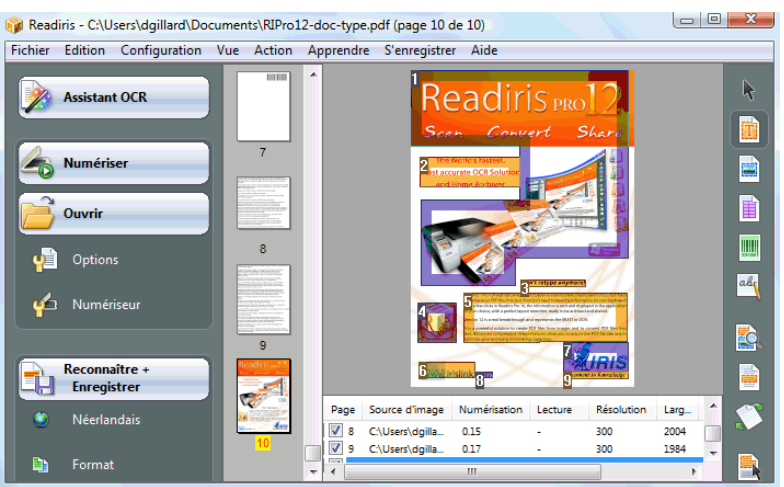
Le scanner numérise automatiquement une autre page après le nombre de secondes spécifié, sans que vous deviez cliquer chaque fois sur le bouton **Numériser**.

Cliquez sur **Interrompre** dans la boîte de dialogue **Numérisation avec pause** pour mettre un terme à la numérisation automatique ou appuyez sur la touche **ECHAP** du clavier.

Cliquez sur **Suspendre** dans la boîte de dialogue **Numérisation avec pause** pour suspendre la numérisation ou appuyez sur la barre d'espacement du clavier. Cliquez sur **Poursuivre** lorsque vous êtes prêt à continuer.

MODIFICATION DES DOCUMENTS DE PLUSIEURS PAGES

Lorsque plusieurs documents sont ouverts ou numérisés dans Readiris, la barre d'outils des pages affiche leurs miniatures. Les miniatures de la barre d'outils des pages permettent de modifier les documents de plusieurs pages.



Pour déplacer une page dans un document :

- Cliquez avec le bouton droit sur la page à déplacer et cliquez sur **Sélectionner page**.
- Faites glisser la page à l'emplacement voulu.
- Vous pouvez également cliquer avec le bouton droit sur une page et cliquer sur **Déplacer page en arrière** ou **Déplacer page en avant**.

Pour supprimer une page :

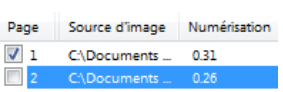
- Cliquez avec le bouton droit sur la page à supprimer et cliquez sur **Supprimer page**.
- Vous pouvez également sélectionner la page et appuyer sur la touche **Suppr** du clavier.

Pour exclure une page du processus de reconnaissance :

- Cliquez avec le bouton droit sur la page à déplacer et cliquez sur **Exclure la page**.
- Vous pouvez aussi effacer sa zone de numéro de page dans le panneau du document.

Les pages exclues sont barrées dans la barre d'outils des pages.

Les pages exclues ne sont pas prises en compte lors de l'impression des images numérisées et de l'enregistrement des numérisations dans des fichiers image multipages.



Conseil : les commandes **Inclure toutes les pages** et **Exclure toutes les pages** du menu **Edition** s'appliquent simultanément à toutes les pages.

CHAPITRE 13
RECONNAISSANCE DE TEXTE MANUSCRIT

En plus du texte dactylographié, des tableaux, des images et des codes à barres, Readiris reconnaît le texte **manuscrit**. Un texte manuscrit comporte des lettres capitales distinctes.

CELL PHONE

Il faut un logiciel ICR (intelligent character recognition, reconnaissance intelligente des caractères) hautement spécialisé pour reconnaître les caractères manuscrits.

Pour reconnaître du texte manuscrit :

- Cliquez sur le bouton texte manuscrit de la barre d'outils Image.

- Dessinez un cadre autour du texte manuscrit.
- Cliquez sur **Reconnaître + Enregistrer** dans la barre d'outils principale.

Le document sera entièrement reconnu, y compris le texte manuscrit.

Remarque : cliquez avec le bouton droit sur une zone de texte manuscrit et cliquez sur **Copier comme texte** pour reconnaître uniquement la zone de texte manuscrit et l'envoyer vers le Presse-papiers.

Symboles reconnus :

La reconnaissance de l'écriture manuscrite est limitée à l'alphabet latin et prend en charge les chiffres, les lettres majuscules (A – Z) et certains signes de ponctuation (virgule, point, signe plus et tiret).

Les accents, les trémas et autres caractères spéciaux ne sont pas pris en charge.

Remarques

- Readiris prend en charge **l'écriture manuscrite en lettres capitales**, pas l'écriture manuscrite liée.

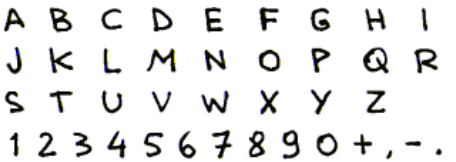
  Pour plus d'informations, consultez la rubrique **Règles d'écriture**.

- Les caractères en majuscules sont remplacés par des minuscules après la reconnaissance, à moins qu'ils ne soient en début de phrase.
- Les caractéristiques du document tel que la langue, le type de police et le pas des caractères ne s'appliquent pas au texte manuscrit.
- L'apprentissage interactif n'est pas possible non plus. La technologie ICR a été développée à partir de plus d'un million d'échantillons d'écriture.

RÈGLES D'ÉCRITURE

Pour que Readiris soit en mesure de reconnaître l'écriture manuscrite, vous devez respecter plusieurs règles :

- Ecrivez des caractères réguliers et bien espacés.



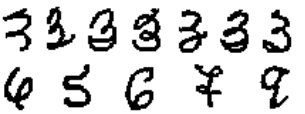
Notez l'écriture des lettres A, G et Q dans l'exemple.

- Utilisez un stylo à bille suffisamment épais.

  Les stylos noirs donnent de meilleurs résultats que les bleus.

  N'utilisez de crayons.

- Evitez les caractères stylisés.



Des caractères trop stylisés augmentent le risque d'erreur d'OCR.

- N'ouvrez pas de boucles qui devraient être fermées, et inversement.

- Evitez les caractères incomplets ou fragmentés.

- Evitez de repasser sur vos traits.

Le retraçage dégrade la qualité de l'image et la netteté des symboles manuscrits.

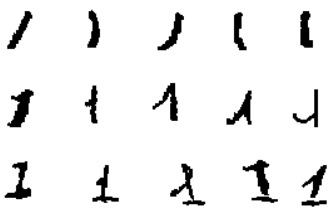
Les caractères complètement barrés ne seront pas reconnus.

- Ecrivez correctement le chiffre 1.

Le chiffre 1 peut être écrit dans un style « européen » et « anglicisé ».

Il peut être accompagné d'un trait horizontal ou non.

Le trait horizontal ne doit pas nécessairement toucher le reste du chiffre.



Conseil : lorsque vous n'obtenez pas de bons résultats, utilisez le **formulaire d'écriture** d'I.R.I.S. afin d'adapter votre style d'écriture. Le formulaire d'écriture vide d'I.R.I.S. sert de modèle de page complète sur lequel vous pouvez écrire en lettres capitales en respectant la taille et l'espacement corrects. Ce formulaire figure sur le CD-ROM de Readiris et dans le dossier d'installation de Readiris.

CHAPITRE 14
RECONNAISSANCE DES CODES À BARRES

PRÉSENTATION DE LA LECTURE DES CODES À BARRES

En plus de la reconnaissance optique des caractères dans 128 langues, Readiris propose également la lecture des codes à barres.

Toutes les symbologies les plus courantes des codes à barres sont prises en charge : Codabar, Code 128, Code 39, Code 39 extended, Code 39 HIBC, Code 93, Datalogic 2 of 5, Discrete 2 of 5, EAN-13, EAN-8, Interleaved 2 of 5, MSI Pharmaceutical, MSI-Plessey, code patch Kodak, PDF-417, PostNet, UCC-128, UPC-A et UPC-E.

Remarque : les codes à barres doivent être imprimés avec des imprimantes laser ou à jet d'encre pour que Readiris puisse effectuer l'OCR. Les codes à barres à impression matricielle ne sont pas pris en charge dans la mesure où le contraste obtenu n'est pas suffisant et où leur résolution est généralement limitée à 60 ppp.

Lecture manuelle des codes à barres

- Déterminez les codes à barres que Readiris doit reconnaître :
    - Dans le menu **Configuration**, cliquez sur **Codes à barres**.
    - Sélectionnez les symbologies que Readiris doit reconnaître :
    - Précisez si Readiris doit vérifier ou supprimer les chiffres de contrôle.
    - Cliquez sur le bouton code à barres dans la barre d'outils Image et tracez un cadre autour des zones de codes à barres dans le document.

    - Cliquez sur **Reconnaître + Enregistrer** dans la barre d'outils principale.

Le document sera entièrement reconnu, y compris les codes à barres qui s'y trouvent.

Remarque : cliquez avec le bouton droit d'une zone de codes à barres et cliquez sur **Copier comme données** pour en copier le contenu dans le Presse-papiers.

INDEX

A

ADF 99

alimentation automatique des documents 99

analyse de page 32

appareil photo numérique 38

apprentissage interactif 65

assistance produits 14

assistant..... 17

B

barre d'outils des images 20

barre d'outils principale..... 20

C

caractéristiques du document 63

chargement des paramètres .. 95

code à barres 107

compression JPEG 2000 93

configuration du numériseur 35

configuration d'usine..... 95

configuration requise 11

contraste 36, 44

couleur d'arrière-plan..... 77

couleur d'arrière-plan des cellules de tableau 80

D

désalignement..... 39

désinstallation de Readiris 13

dictionnaires de polices 67

documents arabes 4, 8, 60

documents asiatiques 4, 7, 60

documents avec données tabulaires 78

documents de plusieurs pages 97, 99

documents hébreux 4, 8, 60

documents texte..... 74

E

écriture manuscrite 103, 104

élimination du bruit 45

enregistrement..... 14

enregistrement de paramètres95

envoi de documents par e-mail 72

F

fenêtrage automatique..... 49

fenêtrage manuel..... 51

feuilles de calcul 78

fichier de configuration..... 95

fichier de fenêtrage 56

formats d'image pris en charge 34

formats de sortie 71

I

image couleur..... 36, 41

image en niveaux de gris..... 36

image en noir et blanc..... 36

images inversées 37

installation 12

interface utilisateur 19

L

langue 60

langue de l'interface utilisateur 22

langues asiatiques 5

Langues asiatiques.....7

langues moyen-orientales ..4, 8, 61

lissage des images couleurs .39, 42

luminosité 36, 42

M

matriciel.....63

mélange de langues 62

mode couleur 36

modèles de fenêtrage 56

modification de l'interface utilisateur.....22

modification des documents de plusieurs pages 100

N

numérique.....62

numérisation avec pause.....99

O

optimisation des documents numérisés.....41

options.....31

options de mise en page.....75

options des graphiques 91

options PDF 84

options XPS 89

P

pages 101

déplacement 101

exclusion 102

sélection 101

suppression..... 101

panneau de document..... 21

pas des caractères..... 64

précision ou vitesse..... 62

propriétés des documents..... 71

R

recréation des documents d'origine 75

redressement 31

redressement de la page 31

réorientation des données dans les documents PDF 87

résolution 36

restauration de la configuration d'usine 95

rotation..... 32

S

SmartTasks.....20, 25

sortie Excel.....71

sortie HTML.....71

sortie OpenDocument.....71

sortie PDF.....71, 82

sortie PDF iHQC 85

sortie RTF.....71

sortie SpreadsheetML.....71

sortie Unicode 71

sortie Word.....71

sortie WordML.....71

sortie XML 71

sortie XPS.....71, 87

sortie XPS iHQC 90

T

tableaux 78

tailles du papier 78

type de police 63

V

vitesse ou précision 63