

MANUEL D'INSTALLATION SOUS WINDOWS

SERVEUR GARGANTUA OCR

Ce document est la propriété de SIATEL.

Il ne peut être reproduit ou communiqué sans l'autorisation écrite de SIATEL.

TABLE DES MATIÈRES

INTRODUCTION	5
PRÉREQUIS ET PRÉPARATION	6
Configuration minimale requise	6
Contenu du kit d'installation	7
INSTALLATION DU LOGICIEL SERVEUR OCR.....	8
Installer le logiciel.....	8
Service Gargantua 8 OCR.....	9
Démarrage manuellement du service	10
Arrêter le service manuellement	10
Désinstaller le logiciel.....	11
PARAMÉTRAGE DU SERVEUR ET DU MOTEUR OCR.....	12
Paramétrage du serveur OCR	12
Paramétrage du moteur OCR	16
ABBYY FineReader 10	16
ABBYY FineReader 12	24
SERVEUR GARGANTUA	36
Installation du plugin 'OCR Serveur'	36
Paramétrage du plugin 'OCR Serveur'	39
Les extensions 'OCR Serveur'	41
L'extension 'Connexion serveur OCR'	42
L'extension 'Tâche asynchrone OCR'	47

INTRODUCTION

Le module d'extension serveur OCR Gargantua est un outil logiciel destiné à transformer en mode texte des documents de type image de différents formats. Il en résulte, selon les options choisies pour la reconnaissance, la génération de fichiers textes divers : textes séparés des originaux images, textes attachés aux documents d'origine (sous forme d'une couche par exemple) ou textes incrustés directement dans les originaux, ...

Les volumes de document à reconnaître peuvent être importants : il conviendra dès lors d'ajuster la puissance du serveur traitant les fonctions d'OCR en conséquence et en fonction des volumes à traiter.

Les volumes de document à reconnaître peuvent être limités mais les fichiers très volumineux : là aussi il conviendra d'ajuster la puissance de traitement à ce type de document.

Dans les deux cas le type de licence acquise devra être ; nombre de documents traités par mois, nombre de processeurs pouvant travailler en parallèle...

PRÉREQUIS ET PRÉPARATION

CONFIGURATION MINIMALE REQUISE

Assurez-vous que la machine sur laquelle vous souhaitez installer le serveur OCR Gargantua dispose de la configuration minimale requise :

1. Moteur OCR ABBYY FineReader Engine 12 :

- Prérequis matériels :
 - pour le traitement de documents d'une page - 400 Mo de RAM minimum, 1 Go de RAM recommandé
 - pour le traitement de documents de plusieurs pages - RAM minimale de 1 Go, et au moins 1,5 Go recommandé
 - pour le traitement en mode multiprocesseurs ou processeur multicœurs - 450 Mo + (nombre de cœurs) × 350 Mo de RAM
 - 1,5 Go pour l'installation de la bibliothèque et 15 Mo supplémentaires pour chaque page de traitement d'un document multipages
- Prérequis systèmes d'exploitation et framework :
 - Windows Server 2016 ou 2019
 - Windows 10
 - Windows Server 2012 R2 with the April 2014 update rollup ([KB2919355](#))
 - Windows 8.1 with the April 2014 update rollup ([KB2919355](#))
 - Windows Server 2012
 - Windows 8
 - Windows Server 2008 R2 SP1
 - Windows 7 SP1
 - .NET Framework 4.6.1 avec siatel-mfc140-package.msi (Visual C++ Redistributable 2014+)

2. Moteur OCR ABBYY FineReader Engine 10 :

- Prérequis matériels :
 - PC avec processeur 1 GHz ou supérieur
 - 650 Mo d'espace disque pour une installation standard et 650 Mo pour le fonctionnement du programme
 - 512 Mo de RAM, plus 256 Mo de RAM pour chaque cœur de processeur supplémentaire
- Prérequis systèmes d'exploitation et framework :
 - Windows Server 2008
 - Windows Server 2003
 - Windows 10
 - Windows 8

- Windows 7
- Windows XP
- .NET Framework 3.5 avec siatel-mfc-package.msi (Visual C++ Redistribuable 2008)

CONTENU DU KIT D'INSTALLATION

Le kit d'installation du serveur OCR GARGANTUA contient :

- **Gargantua 8 OCR Server Installation - Windows** : fichier PDF décrivant les étapes d'installation ;
- **Kit** : le setup OCR organisé en deux parties (serveur et paramétrage OCR) et le moteur OCR ABBYY;

Le setup d'installation Gargantua OCR serveur est disponible pour une architecture 32 ou 64 bits.

Le setup d'installation FineReader Engine ABBYY 10 est disponible uniquement pour une architecture 32 bits .

Le setup d'installation FineReader Engine ABBYY 12 est disponible pour une architecture 32 ou 64 bits.



Il faut que le serveur Gargantua OCR et le module FineReader Engine ABBYY 12, soient installés pour la même architecture (soit les deux pour 32 bits, soit pour 64 bits)



Pour une licence de type « runtime licence » il faut supprimer/renommer le fichier 'Protection.Developer.dll', localisé dans le chemin d'installation ABBYY (pour l'architecture de 64 bits en dossier « Bin64 » et pour 32 bits en dossier « Bin »)

INSTALLATION DU LOGICIEL SERVEUR OCR

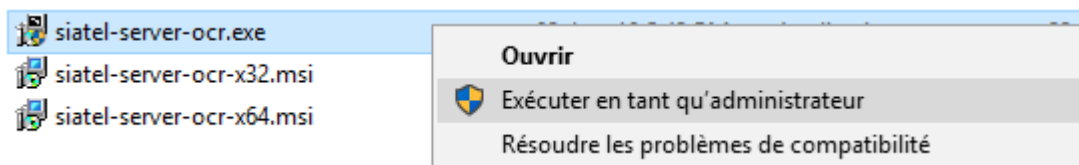
Pour installer le logiciel serveur OCR, procédez comme suit :

- Lancez l'installeur en tant qu'administrateur (applicable sur le fichier siatel-server-ocr.exe)
- Patientez jusqu'à ce que l'assistant d'installation s'affiche, puis cliquez sur Suivant
- Les paramètres par défaut conviennent généralement : à moins que vous n'ayez des raisons particulières de modifier certains d'entre eux, cliquez sur Suivant pour chacune des étapes, jusqu'à ce que l'installation démarre.
- Patientez jusqu'à la fin de l'installation.
- Une fois l'installation terminée, cliquez sur Fermer.

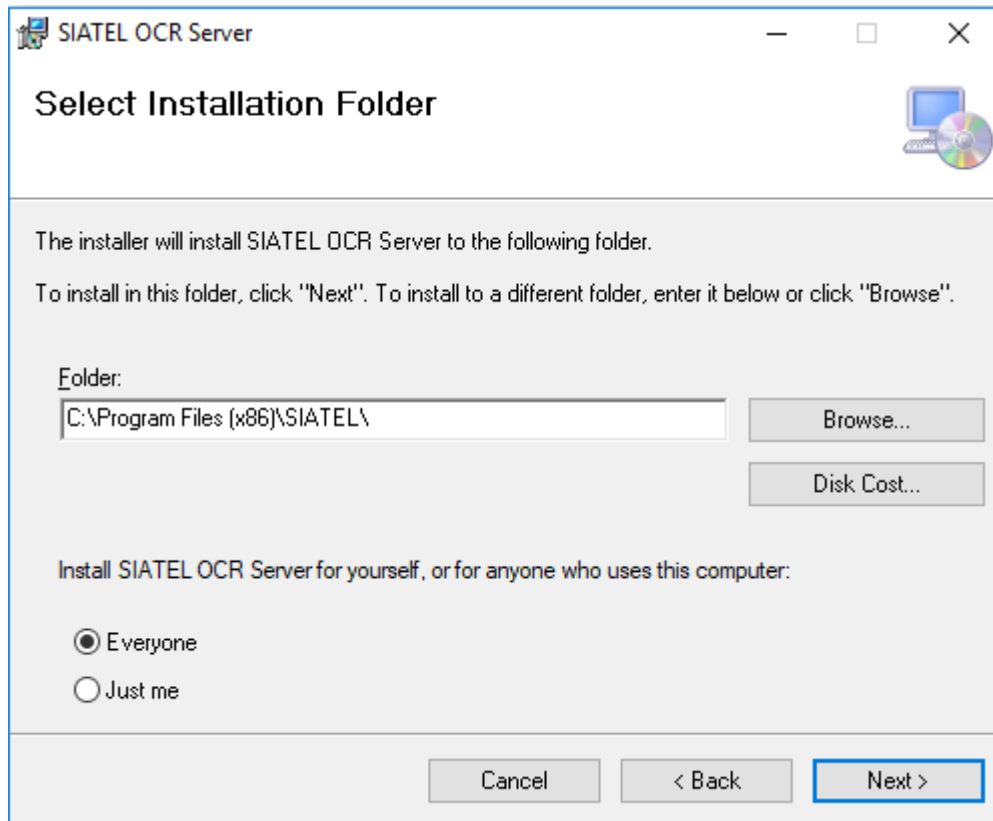
INSTALLER LE LOGICIEL

Pour installer le logiciel serveur OCR, procédez comme suit :

Exécutez le setup d'installation en tant qu'administrateur (fichier siatel-server-ocr.exe)



- Patientez jusqu'à ce que l'assistant d'installation s'affiche, puis cliquez sur Suivant
- Les paramètres par défaut conviennent généralement : à moins que vous n'ayez des raisons particulières de modifier certains d'entre eux, cliquez sur Suivant pour chacune des étapes, jusqu'à ce que l'installation démarre



- Patientez jusqu'à la fin de l'installation.
- Une fois l'installation terminée, cliquez sur Fermer.

SERVICE GARGANTUA 8 OCR

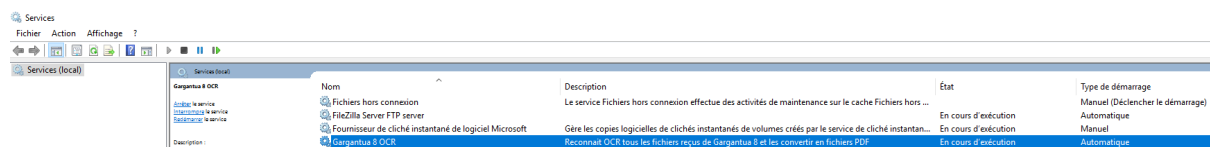
Après l'installation du logiciel, le service Gargantua 8 OCR s'installe et démarre en mode automatique.



Une fois que l'application OCR serveur est désinstallée, le service OCR est également désinstallé.

Pour accéder aux services de Windows, il suffit d'aller sur le menu « Démarrer » « Exécuter » et de taper « services.msc » ou « services ».

Vérifiez que le service est démarré en mode automatique, comme illustré sur la figure suivante :



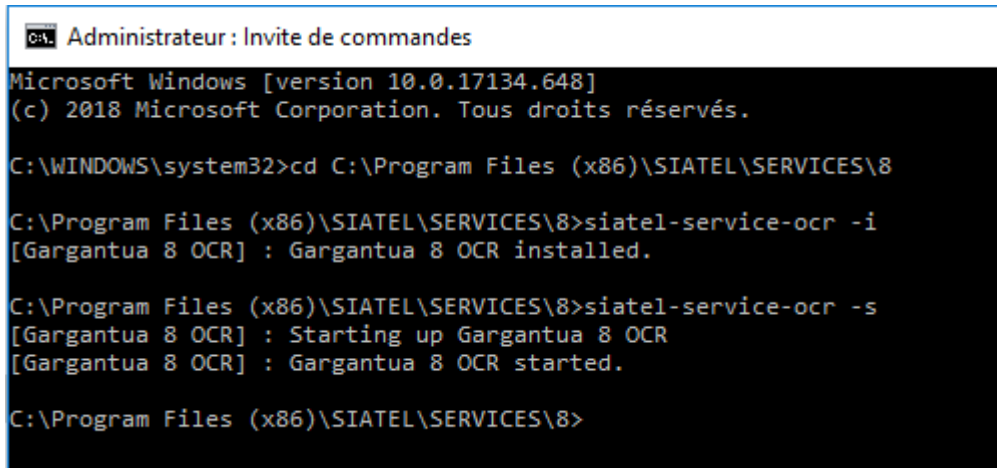
Si le serveur Gargantua OCR ne peut pas démarrer en mode automatique, vous pouvez le démarrer manuellement.

DÉMARRAGE MANUELLEMENT DU SERVICE

Exécuter l'application « Invite de commandes » en tant qu'administrateur et tapez le chemin d'installation (par défaut « C:\Program Files\SIATEL\SERVICES\8 »).

La procédure d'installation et de démarrage du service Gargantua 8 OCR, est la suivante :

- Installer le service : tapez dans la ligne de commandes « siatel-service-ocr -i »
- Démarrer le service : tapez dans la ligne de commandes « siatel-service-ocr -s »



```
Administrateur : Invite de commandes
Microsoft Windows [version 10.0.17134.648]
(c) 2018 Microsoft Corporation. Tous droits réservés.

C:\WINDOWS\system32>cd C:\Program Files (x86)\SIATEL\SERVICES\8

C:\Program Files (x86)\SIATEL\SERVICES\8>siatel-service-ocr -i
[Gargantua 8 OCR] : Gargantua 8 OCR installed.

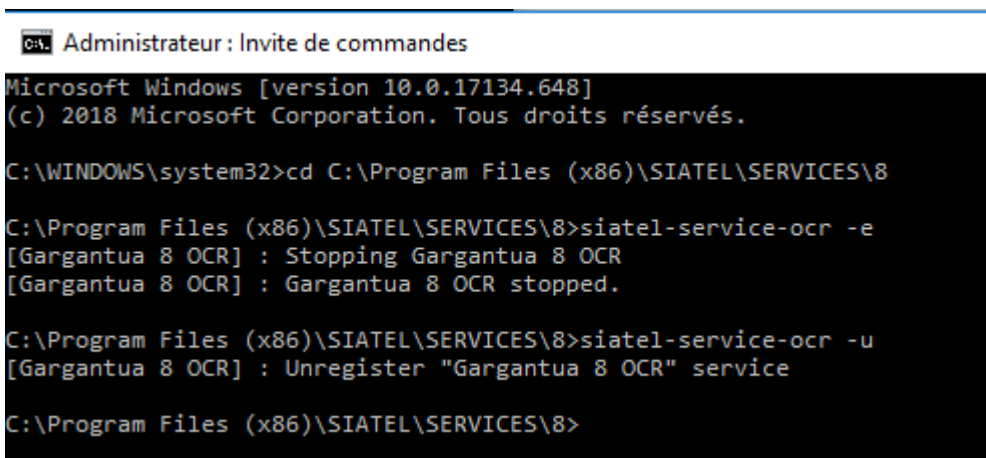
C:\Program Files (x86)\SIATEL\SERVICES\8>siatel-service-ocr -s
[Gargantua 8 OCR] : Starting up Gargantua 8 OCR
[Gargantua 8 OCR] : Gargantua 8 OCR started.

C:\Program Files (x86)\SIATEL\SERVICES\8>
```

ARRÊTER LE SERVICE MANUELLEMENT

Les procédures d'arrêt et de désinstallation du service Gargantua 8 OCR sont les suivantes :

- Arrêter le service : tapez dans la ligne de commandes « siatel-service-ocr -e »
- Désinstaller le service : tapez dans la ligne de commandes « siatel-service-ocr -u »



```
Administrateur : Invite de commandes
Microsoft Windows [version 10.0.17134.648]
(c) 2018 Microsoft Corporation. Tous droits réservés.

C:\WINDOWS\system32>cd C:\Program Files (x86)\SIATEL\SERVICES\8

C:\Program Files (x86)\SIATEL\SERVICES\8>siatel-service-ocr -e
[Gargantua 8 OCR] : Stopping Gargantua 8 OCR
[Gargantua 8 OCR] : Gargantua 8 OCR stopped.

C:\Program Files (x86)\SIATEL\SERVICES\8>siatel-service-ocr -u
[Gargantua 8 OCR] : Unregister "Gargantua 8 OCR" service

C:\Program Files (x86)\SIATEL\SERVICES\8>
```

DÉSINSTALLER LE LOGICIEL

Avant de désinstaller le logiciel, assurez-vous que le service Gargantua 8 OCR est arrêté, comme indiqué dans le chapitre « [Arrêt manuel du service](#) ».

Désinstallez le serveur Gargantua 8 OCR dans le « Panneau de configuration\Programmes\Programmes et fonctionnalités », comme illustré sur la figure suivante :

Désinstaller ou modifier un programme

Pour désinstaller un programme, sélectionnez-le dans la liste et cliquez sur Désinstaller, Modifier ou Réparer.

Organiser ▼	Désinstaller	Modifier	Réparer
Nom	Éditeur	Taille	Version
 SIATEL ABBYY FineReader 12 Engine	SIATEL SA	1.33 Go	12.2.27
 SIATEL OCR Server	SIATEL SA	1.84 Mo	8.0.1
 SIATEL Office enhancements for x64 architecture	SIATEL SA	4.54 Mo	8.0.7745

PARAMÉTRAGE DU SERVEUR ET DU MOTEUR OCR

Dans le chemin d'installation du serveur OCR, le fichier « siatel.ini » permet de paramétrer le service OCR et le moteur OCR.

Il peut être modifié manuellement ou via l'application de paramétrage « siatel-settings-ocr.exe » qui se trouve dans le chemin d'installation.

L'application de paramétrage OCR ouvre la boîte de dialogue suivante :

Paramètres de service OCR

Sélectionnez les paramètres du moteur OCR.
Sélectionnez la langue, le format du texte et les autres options.

Paramètres généraux

☒ Activer le fichier journal ☐ Activer le journal de la console Niveau d'erreur _____

☒ Déplacer le fichier non reconnu dans le dossier "ERROR" Afficher toutes les informations ▼

Chemin serveur Browse...

Paramètres du moteur OCR FineReader

Chemin ABBYY

Langage de détection

☒ Anglais ☒ Français ☒ Allemand ☒ Espagnol ☐ Grecque ☐ Russe ☐ Arabe

Type du texte

☒ Normal ☒ Typewriter ☒ Matrix ☐ Gothic ☐ Manuscrit Paramètres manuscrit...

Prétraitement de l'image

☒ Redressement ☒ Détection de l'orientation de l'image Resolution _____

☒ Nettoyage ☐ Conversion de couleur ou niveau de gris en noir et blanc

☐ Miroir ☒ Détection de l'image inversée (pendant la reconnaissance)

Recognized parameters

☐ Interdire en exposant ☐ Interdire le subscript ☐ Interdire la césure ☐ Orientation automatique

☒ Détection des codes barres

☒ Autodetect ☐ Code39 ☐ Interleaved25 ☐ EAN13 ☐ EAN8

☐ PDF417 ☐ QRCode ☐ Codabar ☐ Code128

Paramètres d'export en PDF

Mode d'exportation

Compliance

OCR result format

Default OK Cancel

PARAMÉTRAGE DU SERVEUR OCR

Les options de paramétrage du serveur OCR se trouvent dans la section [SIATEL].

Toutes ces options seront uniquement prises en compte au démarrage du serveur.

Les listes d'options et leurs descriptions sont détaillées dans le tableau ci-dessous.

Options de configuration du serveur OCR

Clé	Détails
ocr.root.path	Chemin sur le disque dur où les documents seront traités. Ce chemin doit avoir le droit d'écriture et de lecture. Valeur par défaut : »C:\SIATEL\SERVICES\ « Dans ce chemin, au démarrage du serveur, une structure de dossier sera automatiquement créée comme suit : 'DOWNLOAD' - sous-dossier utilisé pour stocker les fichiers reconnus 'ERROR' - sous-dossier utilisé pour stocker les fichiers non reconnus (avec erreurs) et leur fichier journal (siatel.log) 'LOG' - sous-dossier utilisé pour stocker le fichier journal avec le nom « siatel.log »; ces fichiers ont une taille maximale de 50 Mo, après quoi le journal sera sauvegardé dans un fichier final qui aura le nom « siatel_<date actuelle> .log » 'TEMP' - sous-dossier utilisé pour stocker les fichiers temporaires dans l'étape de reconnaissance OCR 'UPLOAD' - sous-dossier utilisé pour stocker les fichiers qui proviennent du serveur Gargantua 8 ; ces fichiers seront déplacés dans le sous-dossier TEMP pour être traités ;
dir.upload	Nom du dossier dans lequel les documents à traiter sont mis à la disposition du serveur OCR. Par défaut, ce paramètre est initialisé avec la valeur 'UPLOAD' (dossier 'UPLOAD').
dir.download	Nom du dossier que le service Gargantua analyse pour trouver les documents convertis. Les documents convertis porteront le même nom que les documents source (au type de document près qui peut potentiellement être différent). Par défaut, ce paramètre est initialisé avec la valeur 'DOWNLOAD' (dossier 'DOWNLOAD')
dir.error	Nom du dossier que le service Gargantua analyse pour détecter les échecs de conversion. En cas d'échec, le document source sera mis dans ce dossier pour signaler que l'opération n'a pas pu être exécutée. Par défaut, ce paramètre est initialisé avec la valeur 'ERROR' (dossier 'ERROR').

Clé	Détails
dir.skip	Dossier dans lequel le serveur OCR déplace les documents détectés comme déjà reconnus (c'est possible pour les fichiers PDF). Ces fichiers sont ignorés par le serveur OCR s'il est paramétré pour détecter ce type de fichiers. Par défaut, ce paramètre est initialisé avec la valeur 'SKIP'.
dir.temp	Dossier tampon utilisé pour le transfert des documents vers le serveur OCR. Les documents à traiter sont d'abord transférés dans ce dossier tampon, puis ils sont déplacés dans le dossier spécifié par le paramètre 'dir.upload', ceci pour éviter que le serveur OCR ne commence à traiter un document qui ne serait pas complètement transféré. Par défaut, ce paramètre est initialisé avec la valeur 'TEMP' (dossier 'TEMP').
log.console	Affichage des messages sur la console Valeurs possibles : 0 - désactivé 1 - activé Valeur par défaut : 0
log.file	Enregistrer les messages du journal dans un fichier Valeurs possibles : 0 - désactivé 1 - activé Valeur par défaut : 1
message.error.level	Définit le niveau et le type de messages qui seront écrits dans le journal Valeurs possibles : 0 - les messages de ce type sont obligatoires et seront écrits dans le journal 1 - seuls les messages d'erreurs et ceux obligatoires seront écrits dans le journal 2 - seulement les messages d'erreurs, obligatoires et les avertissements seront écrits dans le journal 3 - les messages d'information seront également écrits dans le journal 4 - tous les messages seront écrits dans le journal

Clé	Détails
	Valeur par défaut : 1
<i>move.error.file</i>	Déplacer les fichiers sources non reconnus dans le dossier « ERROR » Valeurs possibles : 0 - désactivé 1 - activé Valeur par défaut : 1
<i>maxim.threads.count</i>	Définir le nombre maximal de threads à créer pour la phase de reconnaissance Valeurs possibles : Doit être une valeur comprise entre 1 et 10 Valeur par défaut : 5
<i>process.timeout</i>	Définir le délai d'expiration du processus d'OCR en minutes. Une fois ce délai écoulé, le processus se termine (fin forcée par l'application). Valeurs possibles : Doit être une valeur minimale de 10 minutes. Pour une valeur inférieure à 10, cette option est désactivée. Valeur par défaut : 60.
<i>files.clean.timeout</i>	Définir le délai d'expiration du nettoyage des fichiers orphelins en jours Une fois ce délai écoulé, le serveur OCR supprimera tous les fichiers orphelins restants qui ont été créés il y a plus de « x » jours. Valeurs possibles : Doit être une valeur minimale de 1 jour. Pour une valeur à '0', cette option est désactivée. Valeur par défaut : 30 jours.
<i>ocr.engine</i>	Définir le moteur disponible pour les opérations d'OCR Valeurs possibles: 'ABBY10' - pour le moteur ocr FineReader ABBYY 10 'ABBY12' - pour le moteur ocr FineReader ABBYY 12

Clé	Détails
	Valeur par défaut : 'ABBY12'
<i>socket.name</i>	Informations sur le serveur Gargantua pour la connexion avec socket. Si ces informations ne sont pas spécifiées, le serveur OCR ne peut pas communiquer avec le serveur Gargantua 8 Valeur par défaut : 'Gargantua'
<i>socket.port</i>	25555
<i>ftp.enable</i>	Activer / désactiver le serveur FTP intégré dans le serveur OCR Valeurs possibles : 0 - désactivé 1 - activé Valeur par défaut : 0
<i>ftp.port</i>	Port utilisé pour la communication FTP . Par défaut '21'.
<i>ftp.anonymous</i>	Option permettant d'activer/de désactiver l' authentification FTP . Par défaut le paramètre est à ' false ', ce qui signifie qu'il faut spécifier un nom et un mot de passe pour authentifier les sessions FTP vers le serveur OCR, dans les options ' ftp.user ' et ' ftp.pass '. Si ce paramètre est à ' true ', les options d'authentification sont ignorées.
<i>ftp.user</i> <i>ftp.pass</i>	Nom d'utilisateur et mot de passe pour authentifier les sessions FTP vers le serveur OCR.

PARAMÉTRAGE DU MOTEUR OCR

Dans le chemin d'installation du serveur OCR, le fichier « siatel.ini » permet de paramétrer le service OCR et le moteur OCR.

Les options de paramétrage du serveur OCR se trouvent dans la section [ABBY].

ABBY FINEREADER 10

Le setup d'installation de FineReader Engine ABBYY 10 est uniquement disponible pour une architecture en 32

bits.

Pour enregistrer "FREngine.dll", utilisez la ligne de commande suivante:



regsvr32.exe /c /n /i:[Path to the Inc folder] [Path to the FREngine.dll]

*E.g. regsvr32.exe /c /n /i:"C:\Program Files (x86)\SIATEL\ABBYY\FineReader\v10\Inc"
"C:\Program Files (x86)\SIATEL\ABBYY\FineReader\v10\Bin\FREngine.dll"*



Pour enregistrer "LicensingService.exe", utilisez la ligne de commande suivante::

"C:\Program Files (x86)\SIATEL\ABBYY\FineReader\v10\Bin\LicensingService.exe" /install

Les paramètres de configuration de l'OCR ABBYY 10 sont les suivants:

- **ocr.install.path** - Chemin d'installation du moteur ABBYY 10

Valeur par défaut : "C:\Program Files (x86)\SIATEL\ABBYY\FineReader\v10\Bin"

- **recognized.mode** - définir le mode de reconnaissance (charger chaque page ou toutes les pages pendant la phase de reconnaissance)

Valeurs possibles :

0 - par page

1 - par fichier

Valeur par défaut : 1

- **multi.processing.mode** - Spécifie le mode de traitement des documents multipages avec utilisation de processeurs multicœurs

Valeurs possibles :

0 - le nombre de processeur est détecté automatiquement. Si une seule page est traitée ou s'il n'y a qu'un seul processus dans le système, un seul processeur est utilisé. Sinon, le traitement est effectué en parallèle.

1 - les pages d'un document sont traitées de manière séquentielle par un seul processeur

Valeur par défaut : 0

- **multi.processes.count** - Spécifie le nombre maximal de processus pouvant être exécutés simultanément.

Par défaut, cette propriété est 0, ce qui signifie que le nombre de processus sera égal au minimum des valeurs suivantes:

- nombre de cœurs de processeur physique ou logique disponibles,
- nombre de cœurs de processeur disponibles dans la licence,
- nombre de pages du document en cours de traitement.

- ***recognized.profile*** - Profils de reconnaissance prédéfinis

Valeurs possibles :

"Default" - Définit tous les paramètres à leurs valeurs par défaut.

"DocumentConversion_Accuracy" - Convient pour convertir des documents dans un format éditable (par exemple RTF, DOCX). Favorise la précision.

"DocumentConversion_Speed" - Convient pour convertir des documents dans un format éditable (par exemple RTF, DOCX). Favorise la vitesse d'exécution.

" DocumentArchiving_Accuracy" - Convient pour générer une archive numérique (conversion en PDF, PDF/A, PDF et PDF/A avec MRC). Attention ! Ce profil n'est pas destiné à convertir en RTF, DOCX, PDF texte seulement. Favorise la précision.

"DocumentArchiving_Speed" - Convient pour créer une archive numérique (conversion en PDF, PDF/A, PDF et PDF/A avec MRC). Attention ! Ce profil n'est pas destiné à convertir en RTF, DOCX, PDF texte seulement. Favorise la vitesse d'exécution.

"BookArchiving_Accuracy" - Convient pour créer une bibliothèque électronique (conversion en PDF, PDF/A, PDF et PDF/A avec MRC). Favorise la précision.

" BookArchiving_Speed" - Convient pour créer une bibliothèque électronique (conversion en PDF, PDF/A, PDF et PDF/A avec MRC). Favorise la vitesse d'exécution.

"TextExtraction_Accuracy" - Convient pour extraire du texte d'un document. Les paramètres ont été optimisés pour la précision. Attention ! Ce profil n'est pas destiné à convertir en RTF, DOCX, PDF texte seulement.

"TextExtraction_Speed" - Convient pour extraire du texte d'un document. Les paramètres ont été optimisés pour la vitesse d'exécution. Attention ! Ce profil n'est pas destiné à convertir en RTF, DOCX, PDF texte seulement.

"FieldLevelRecognition" - Convient pour reconnaître de petites parties de textes. Ce profil conserve les paramètres par défaut.

"BarcodeRecognition_Accuracy" - Convient pour l'extraction de codes-barres. Ce profil n'extrait que les codes-barres (textes, images ou tableaux ne sont pas détectés). Favorise la précision.

"BarcodeRecognition_Speed" - Convient pour l'extraction de codes-barres. Ce profil n'extrait que les codes-barres (textes, images ou tableaux ne sont pas détectés). Favorise la vitesse.

"HighCompressedImageOnlyPdf" - Convient pour créer des fichiers PDF hautement compressés contenant des documents entiers sauvegardés sous forme d'images. L'intégralité du document est sauvegardé comme image (PEM_Mode ImageOnly).

"BusinessCardsProcessing" - Convient pour reconnaître les cartes de visite.

"EngineeringDrawingsProcessing" - Convient pour reconnaître les dessins techniques. Il prend en charge les grandes tailles et la complexité des schémas d'ingénierie, en tenant compte des différentes orientations du texte dans l'image. Ce profil est destiné pour convertir ces images en fichier PDF recherachable.

Attention ! Ce profil n'est pas destiné à convertir en RTF, DOCX, PDF texte seulement.

"Version9Compatibility" - Pour des raisons de compatibilité, définit les paramètres aux valeurs par défaut d'ABBYY FineReader Engine 9.0.

Valeur par défaut : *"Default"*

- **languages** - Langues de reconnaissance

Valeurs possibles : cela peut être une liste des langues suivantes séparées par ";" :

"French"

"OldFrench"

"English"

"OldEnglish"

"German"

"Spanish"

"Greek"

"Russian"

"Arabic"

"Romanian"

"Dutch"

"Digits"

"Czech"

"Hungarian"

"PortugueseStandard"

etc...

Valeur par défaut : French;English;German;Spanish;Digits;Romanian;Greek;

- **text.type** - Type de texte reconnu

Valeurs possibles : Peuvent être une liste des valeurs suivantes séparées par ";" :

0 - détection automatique

1 - texte normal

2 - texte imprimé

3 - matrice

4 - gothique

5 - écrit à la main

Valeur par défaut : 0;1;2;3;

- **handprint.text.type** : disponible uniquement lorsque le type de texte reconnu est écrit à la main (TextType contient la valeur 5)

Valeurs possibles :

0 - texte simple / normal

1 - texte souligné

2 - Texte encadré

3 - Précasé gris

4 - Précasé en cases séparées

5 - Peigne simple

6 - Peigne encadré

7 - Précasé continu encadré

Valeur par défaut : 0

- **resolution** - résolution de l'image en dpi

Valeur recommandée :

- 300 dpi pour des textes imprimés avec des polices de taille 10 points ou plus

- 400 à 600 dpi pour les textes imprimés avec des polices inférieures ou égales à 9 points

Valeurs possibles :

0 - conserve la résolution de l'image d'origine

100

200

300

400

500

600

Valeur par défaut : 0

- ***correct.skew*** - Redressement de l'image

Valeurs possibles :

0 - désactivé

1 - activé

Valeur par défaut : 1

- ***discard.color.image*** - Conversion en noir et blanc de documents initialement en couleur ou en niveaux de gris

Valeurs possibles :

0 - désactivé

1 - activé

Valeur par défaut : 0

- ***mirror.image*** - Miroir de l'image

Valeurs possibles :

0 - désactivé

1 - activé

Valeur par défaut : 0

- ***remove.garbage*** - Nettoyage de l'image

Valeurs possibles :

0 - désactivé

1 - activé

Valeur par défaut : 1

- ***detect.orientation*** - Détecte automatiquement l'orientation de l'image

Valeurs possibles :

0 - désactivé

1 - activé

Valeur par défaut : 1

- ***detect.invert*** - Détecter l'image inversée (pendant la reconnaissance)

Valeurs possibles :

0 - désactivé

1 - activé

Valeur par défaut : 1

- ***detect.barcodes*** - Détection des codes à barres

Valeurs possibles :

0 - désactivé

1 - activé

Valeur par défaut : 1

- ***barcode.type*** - Type de code à barres détecté

Valeurs possibles : Peuvent être une liste des valeurs suivantes séparées par ";" :

0 - détection automatique

1 - code à barre Code39

2 - type de code à barres Interleaved25

3 - type de code à barres EAN13

4 - type de code à barres Code128

5 - type de code à barres EAN8

6 - type de code à barres PDF417

7 - type de code à barres Codabar

8 - type de code à barres QRCode

Valeur par défaut : 0

- ***barcode.orientation*** - Orientation du code à barres

Valeurs possibles :

0 - toutes les directions

1 - détection automatique

Valeur par défaut : 1

- ***prohibit.superscript*** - Interdire les caractères "exposant"

Valeurs possibles :

0 - désactivé

1 - activé

Valeur par défaut : 0

- ***prohibit.subscript*** - Interdire les caractères "indice"

Valeurs possibles :

0 - désactivé

1 - activé

Valeur par défaut : 0

- ***prohibit.hyphenation*** - Interdire les césures de mot

Valeurs possibles :

0 - désactivé

1 - activé

Valeur par défaut : 0

- ***pdf.compliance*** - Niveaux de conformité et versions pour le fichier PDF

Valeurs possibles :

0 - Aucune

1 - PDF/A 1B

2 - PDF/A 1A

Valeur par défaut : 0

- ***pdf.export.mode*** - Mode d'exportation en PDF

Valeurs possibles :

- 0 - Texte et image seulement
- 1 - Texte superposé sur l'image de la page
- 2 - Texte sous l'image de la page (pdf consultable)
- 3 - Image de la page seulement
- 4 - Texte seulement

Valeur par défaut : 2

- ***pdf.scenario.mode*** - Etablit un scénario de reconnaissance

Valeurs possibles :

- 0 - Qualité maximale
- 1 - Équilibré
- 2 - Taille minimale
- 3 -Vitesse maximale

Valeur par défaut : 0

- ***export.file.type*** - Exporter le résultat OCR dans un format de type de fichier

Valeurs possibles :

- 0 - PDF
- 1 - RTF
- 2 - TEXT
- 3 - XLS
- 4 - HTML
- 5 - PPT
- 6 - DOCX
- 7 - EBOOK

Valeur par défaut : 0

ABBYY FINEREADER 12

Le setup d'installation FineReader Engine ABBYY 12 est disponible pour une architecture en 32 ou 64 bits.



Il faut que le serveur Gargantua OCR et le module FineReader Engine ABBYY 12, soient installés sur une même architecture (soit les deux en 32 bits, soit les deux en 64 bits)



Pour enregistrer "FREngine.dll", utilisez la ligne de commande suivante:

regsvr32.exe /c /n /i:[Path to the Inc folder] [Path to the FREngine.dll]

*E.g. regsvr32.exe /c /n /i:"C:\Program Files\SIATEL\ABBYY\FineReader\v12\Inc"
"C:\Program Files\SIATEL\ABBYY\FineReader\v12\Bin64\FREngine.dll"*



Pour enregistrer "LicensingService.exe", utilisez la ligne de commande suivante:

"C:\Program Files\SIATEL\ABBYY\FineReader\v12\Licensing\LicensingService.exe" /install



Pour enregistrer "CodeMeterRuntime.exe", utilisez la ligne de commande suivante:

*"C:\Program Files\SIATEL\ABBYY\FineReader\v12\Licensing\Wibu\CodeMeterRuntime.exe"
/ComponentArgs */n*

Les paramètres de configuration de l'OCR ABBYY 12 sont les suivants :

- **ocr.install.path** - Chemin d'installation du moteur OCR ABBYY 12

Valeur par défaut : "C:\Program Files (x86)\SIATEL\ABBYY\FineReader\v12\Bin" pour une architecture en 32 bits

"C:\Program Files\SIATEL\ABBYY\FineReader\v12\Bin64" pour une architecture en 64 bits

- **recognized.mode** - définir le mode de reconnaissance (charger chaque page ou toutes les pages pendant la phase de reconnaissance)

Valeurs possibles :

0 - par page

1 - par fichier

Valeur par défaut : 1

- **multi.processing.mode** - Spécifie le mode de traitement des documents multipages avec utilisation de processeurs multicoeurs

Valeurs possibles :

0 - le nombre de processeur est détecté automatiquement. Si une seule page est traitée ou s'il n'y a qu'un seul processeur dans le système, un seul processeur est utilisé. Sinon, le

traitement s'effectue en parallèle.

1 - les pages d'un document sont traitées de manière séquentielle par un seul processeur.

Valeur par défaut : 0

- **multi.processes.count** - Spécifie le nombre maximal de processus pouvant être exécutés simultanément.

Par défaut, cette propriété est 0, ce qui signifie que le nombre de processus sera égal au minimum des valeurs suivantes :

- nombre de cœurs des processeurs physiques ou logiques disponibles,
- nombre de cœurs des processeurs disponibles dans la licence,
- nombre de pages du document en cours de traitement.

- **recognized.profile** - Profils de reconnaissance prédéfinis

Valeurs possibles :

"Default" - Définit tous les paramètres à leurs valeurs par défaut.

"DocumentConversion_Accuracy" - Convient pour convertir des documents dans un format éditable (par exemple RTF, DOCX). Favorise la précision.

"DocumentConversion_Speed" - Convient pour convertir des documents dans un format éditable (par exemple RTF, DOCX). Favorise la vitesse d'exécution.

"DocumentArchiving_Accuracy" - Convient pour créer une archive numérique (conversion en PDF, PDF/A, PDF et PDF/A avec MRC). Attention ! Ce profil n'est pas prévu pour convertir en RTF, DOCX, PDF texte seulement. Favorise la précision.

"DocumentArchiving_Speed" - Convient pour créer une archive numérique (conversion en PDF, PDF/A, PDF et PDF/A avec MRC). Attention ! Ce profil n'est pas prévu pour convertir en RTF, DOCX, PDF texte seulement. Favorise la vitesse d'exécution.

"BookArchiving_Accuracy" - Convient pour créer une bibliothèque électronique (conversion en PDF, PDF/A, PDF et PDF/A avec MRC). Favorise la précision.

"BookArchiving_Speed" - Convient pour créer une bibliothèque électronique (conversion en PDF, PDF/A, PDF et PDF/A avec MRC). Favorise la vitesse d'exécution.

"TextExtraction_Accuracy" - Convient pour extraire du texte d'un document. Les paramètres sont optimisés pour favoriser la précision. Attention ! Ce profil n'est pas prévu pour convertir en RTF, DOCX, PDF texte seulement.

"TextExtraction_Speed" - Convient pour extraire du texte d'un document. Les paramètres sont optimisés pour favoriser la vitesse d'exécution. Attention ! Ce profil n'est prévu pour convertir en RTF, DOCX, PDF texte seulement.

"FieldLevelRecognition" - Convient pour reconnaître de petites parties de texte. Ce profil conserve les paramètres par défaut.

"BarcodeRecognition_Accuracy" - Convient pour l'extraction de codes-barres. Seuls les codes-barres sont extraits (textes, images ou tableaux ne sont pas identifiés). Favorise la précision.

"BarcodeRecognition_Speed" - Convient pour l'extraction de codes-barres. Seuls les codes-barres sont extraits (textes, images ou tableaux ne sont pas identifiés). Favorise la vitesse d'exécution.

"HighCompressedImageOnlyPdf" - Convient pour créer des fichiers PDF hautement compressés qui contiennent des documents entiers sauvegardés au format image. L'intégralité du document est sauvegardé sous forme d'image (PEM_Mode ImageOnly).

"BusinessCardsProcessing" - Convient pour reconnaître les cartes de visite.

"EngineeringDrawingsProcessing" - Convient pour reconnaître les dessins techniques. Prend en charge les grandes tailles et la complexité des schémas d'ingénierie en tenant compte autant que possible des différentes orientations du texte dans l'image. Ce profil est prévu pour générer ce genre d'images sous la forme de fichiers PDF recherchables.

Attention ! Ce profil n'est pas prévu pour convertir en RTF, DOCX, PDF texte seulement.

"Version9Compatibility" - Fourni pour la compatibilité, définit les paramètres aux valeurs par défaut de ABBYY FineReader Engine 9.0.

"Gargantua"

- Définit tous les paramètres aux valeurs par défaut.
- Définit les valeurs optimales des paramètres pour augmenter la précision de la reconnaissance, peut pénaliser la vitesse d'exécution.
- Augmente la résolution à 300 dpi
- Passe les documents pour lesquels la part de caractères non reconnus est supérieure à "recognition.accuracy" pourcent du total.
- Detecte et supprime les caractères douteux du texte

Valeur par défaut : *"Default"*

- **languages** - Langues de reconnaissance

Valeurs possibles : Peut être une liste des langues suivantes séparées par ";" :

"French"

"OldFrench"

"English"

"OldEnglish"

"German"

"Spanish"

"Greek"

"Russian"

"Arabic"

"Romanian"

"Dutch"

"Digits"

"Czech"

"Hungarian"

"PortugueseStandard"

etc...

Valeur par défaut : French;English;German;Spanish;Digits;Romanian;Greek

- **detect.languages** - Gère la détection automatique de la langue.

L'autodétection est destinée à être utilisée lors de la reconnaissance de documents dont vous ne connaissez pas la langue.

Il est sélectionné dans la liste des langues spécifiées dans la propriété 'languages'.

Valeurs possibles :

0 - désactivé

1 - activé

2 - auto

Valeur par défaut : 1

- **detect.languages.mode** - Détection automatique des langues pour la page entière ou la première page du document.

Disponible uniquement à partir du script/extension Gargantua.

Valeurs possibles :

0 - document entier

1 - première page

- **text.type** - Type de texte reconnu

Valeurs possibles : Peut être une liste des valeurs suivantes séparées par ";" :

0 - détection automatique

1 - texte normal

2 - texte imprimé

3 - matrice

4 - gothique

5 - écrit à la main

Valeur par défaut : 0;1;2;3;

- **handprint.text.type** : disponible uniquement lorsque le type de texte reconnu est Écrit à la main (TextType contient la valeur 5)

Valeurs possibles :

0 - texte simple / normal

1 - texte souligné

2 - Texte encadré

3 - Précasé gris

4 - Précasé en cases séparées

5 - Peigne simple

6 - Peigne encadré

7 - Précasé continu encadré

Valeur par défaut : 0

- **resolution** - résolution de l'image en dpi

Valeur recommandée :

- 300 dpi pour des textes imprimés avec des polices de taille 10 points ou plus

- 400 à 600 dpi pour les textes imprimés avec des polices inférieures ou égales à 9 points

Valeurs possibles :

0 - conserve la résolution de l'image d'origine

100

200

300

400

500

600

1200

Valeur par défaut : 0

- ***correct.skew*** - Redressement de l'image

Valeurs possibles :

0 - désactivé

1 - activé

Valeur par défaut : 1

- ***discard.color.image*** - Conversion en noir et blanc de documents initialement en couleur ou en niveaux de gris

Valeurs possibles :

0 - désactiver

1 - activer

Valeur par défaut : 0

- ***mirror.image*** - Miroir de l'image

Valeurs possibles :

0 - désactivé

1 - activé

Valeur par défaut : 0

- ***remove.garbage*** - Nettoyage de l'image

Valeurs possibles :

0 - désactivé

1 - activé

Valeur par défaut : *1*

- ***detect.orientation*** - Détecte automatiquement l'orientation de l'image

Valeurs possibles :

0 - désactivé

1 - activé

Valeur par défaut : *1*

- ***detect.inverted*** - Détecter l'image inversée (pendant la reconnaissance)

Valeurs possibles :

0 - désactivé

1 - activé

Valeur par défaut : *1*

- ***detect.barcodes*** - Détection des codes à barres

Valeurs possibles :

0 - désactivé

1 - activé

Valeur par défaut : *1*

- ***barcode.type*** - types de code à barres détectés

Valeurs possibles : Peut être une liste des valeurs suivantes séparées par ";" :

0 - détection automatique

1 - code à barre Code39

2 - type de code à barres Interleaved25

3 - type de code à barres EAN13

4 - type de code à barres Code128

5 - type de code à barres EAN8

6 - type de code à barres PDF417

7 - type de code à barres Codabar

8 - type de code à barres QRCode

Valeur par défaut : 0

- ***barcode.orientation*** - Orientation du code à barres

Valeurs possibles :

0 - toutes les directions

1 - détection automatique

Valeur par défaut : 1

- ***prohibit.superscript*** - Interdire les caractères "exposant"

Valeurs possibles :

0 - désactivé

1 - activé

Valeur par défaut : 0

- ***prohibit.subscript*** - Interdire les caractères "indice"

Valeurs possibles :

0 - désactivé

1 - activé

Valeur par défaut : 0

- ***prohibit.hyphenation*** - Interdire les césures de mot

Valeurs possibles :

0 - désactivé

1 - activé

Valeur par défaut : 0

- ***pdf.compliance*** - Niveaux de conformité et versions pour le fichier PDF

Valeurs possibles :

0 - Aucune

1 - PDF/A 1A

2 - PDF/A 1B

3 - PDF/A 2A

4 - PDF/A 2U

5 - PDF/A 2B

6 - PDF/A 3A

7 - PDF/A 3U

8 - PDF/A 3B

Valeur par défaut : 0

- ***pdf.export.mode*** - Mode d'exportation en PDF

Valeurs possibles :

0 - Texte et image seulement

1 - Texte superposé sur l'image de la page

2 - Texte sous l'image de la page (pdf consultable)

3 - Image de la page seulement

4 - Texte seulement

Valeur par défaut : 2

- ***pdf.scenario.mode*** - Etablit un scénario de reconnaissance

Valeurs possibles :

0 - Qualité maximale

1 - Équilibré

2 - Taille minimale

3 -Vitesse maximale

Valeur par défaut : 0

- ***export.file.type*** - Exporter le résultat OCR dans un format de type de fichier

Valeurs possibles :

0 - PDF

1 - RTF

2 - TEXT

3 - XLS

4 - HTML

5 - PPT

6 - DOCX

7 - EBOOK

Valeur par défaut : 0



Les paramètres qui suivent peuvent améliorer la précision de la reconnaissance mais peuvent pénaliser la vitesse d'exécution.



Les paramètres suivants sont forcés à leur valeur optimale si `recognized.profile = "Gargantua"`

- ***recognition.accuracy*** - Les documents pour lesquels les caractères non reconnus ou suspects représentent plus que "recognition.accuracy" pourcent du total, ne sont pas pris en compte.

Valeurs possibles :

0 - option désactivée

1 - valeur minimale

100 - valeur maximale

90 - valeur optimale

Valeur par défaut : 90



Il est possible de supprimer les caractères corrects du texte

- ***confidence.chars.level*** - Supprime les caractères ayant un niveau bas de confiance. Le niveau est le niveau bas de confiance pour les caractères.

Valeurs possibles :

0 - option désactivée

1 - valeur minimale

100 - valeur maximale

10 - valeur optimale

Valeur par défaut : 0

- ***order.zones*** - Lire le texte reconnu selon l'ordre des zones sélectionnées pour chaque page

Sinon, l'ordre ABBYY est utilisé (de gauche à droite, de haut en bas par page).

Valeurs possibles :

0 - désactivé

1 - activé

Valeur par défaut : 1

SERVEUR GARGANTUA

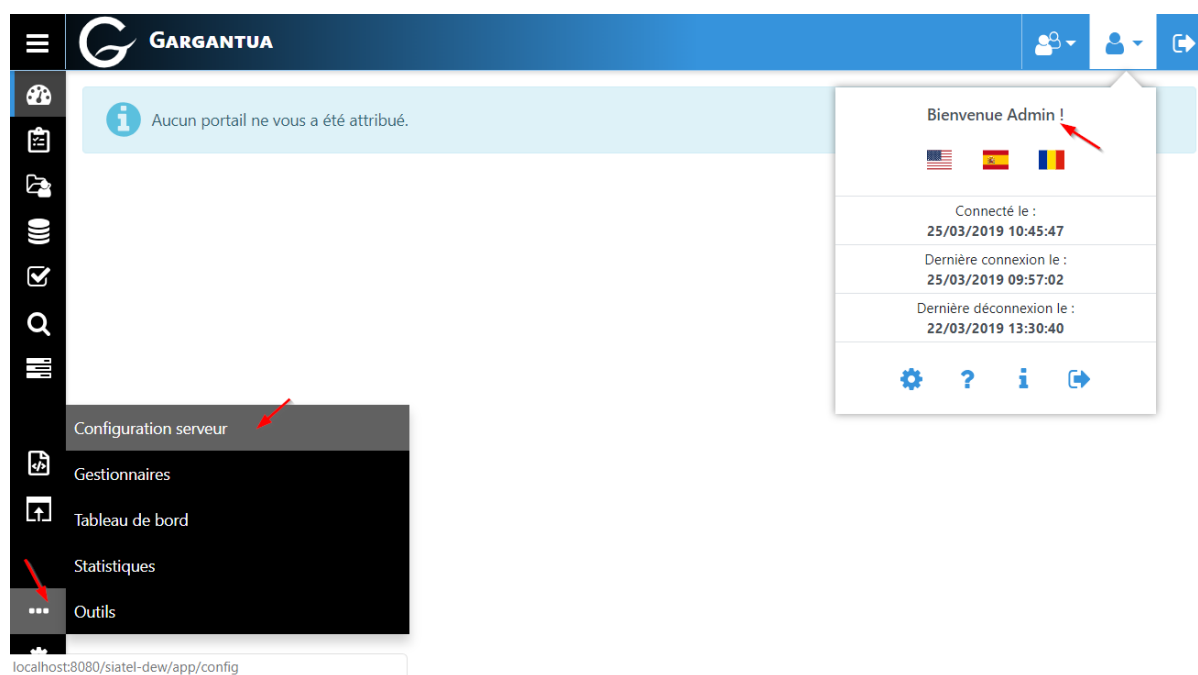
Le serveur OCR est un module additionnel à Gargantua (une extension de type serveur). Cette extension s'installe depuis un navigateur par un utilisateur ayant les droits d'administration sur Gargantua.

Le fonctionnement de cette extension est conditionné par la licence Gargantua.

INSTALLATION DU PLUGIN 'OCR SERVEUR'

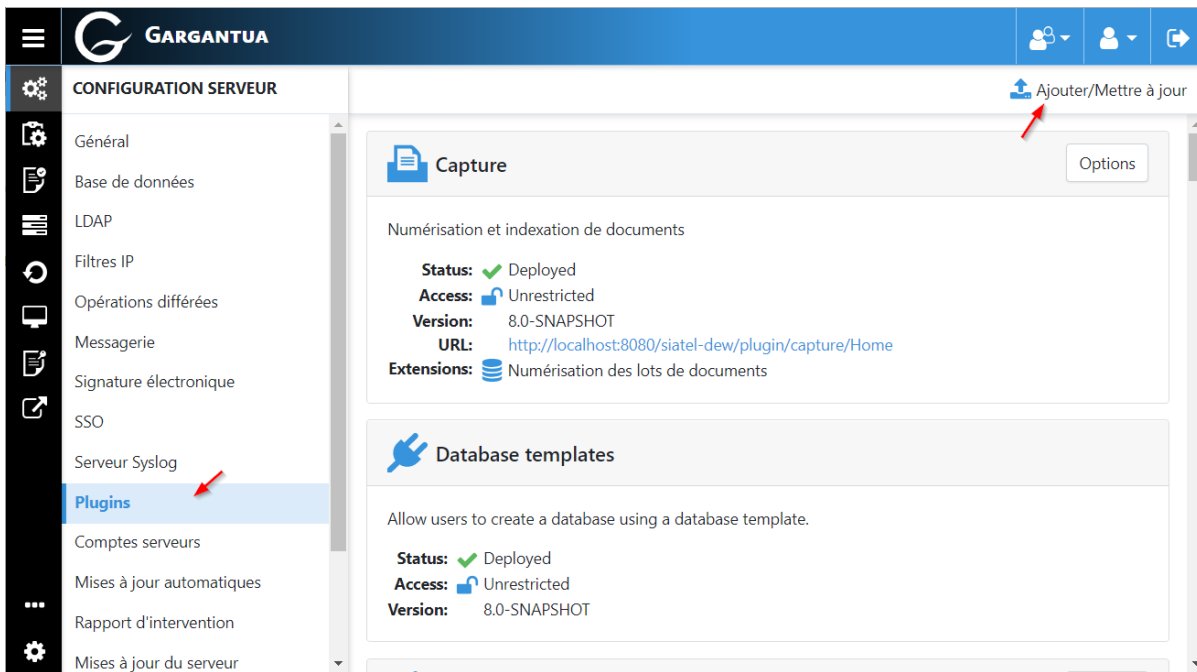
Procédure d'installation ou de mise à jour du plugin 'OCR Server':

1. Ouvrez dans un navigateur une session Gargantua en tant qu'Administrateur.
2. Sélectionnez l'écran de configuration serveur.

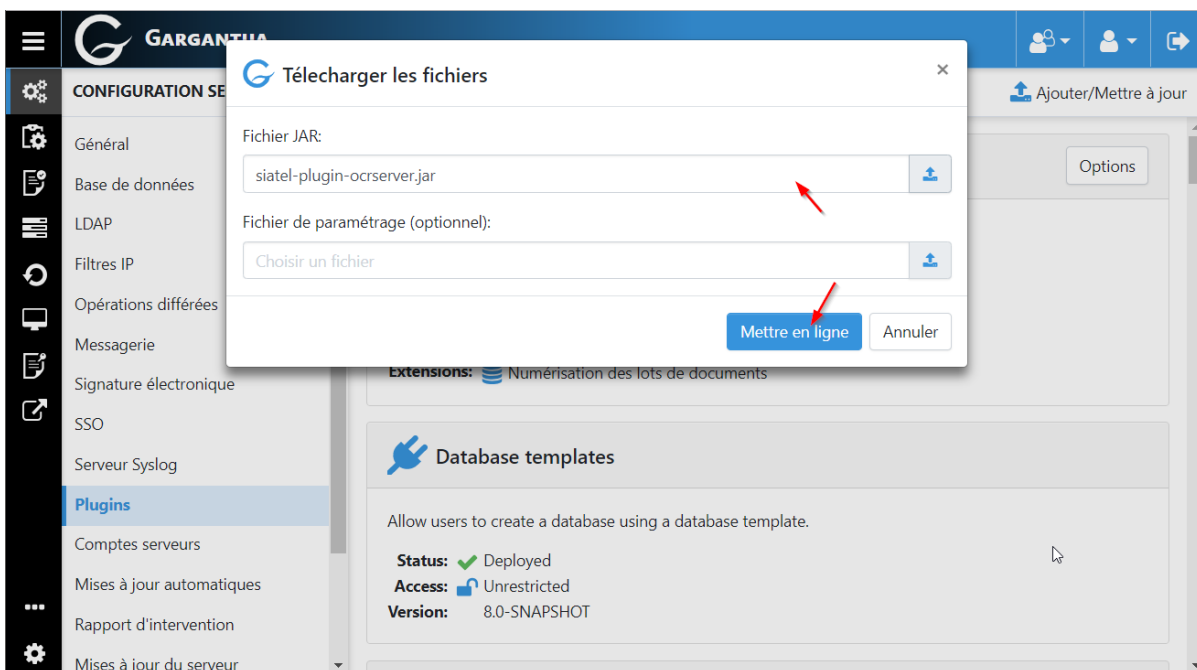


3. Sélectionnez l'entrée « Plugins »

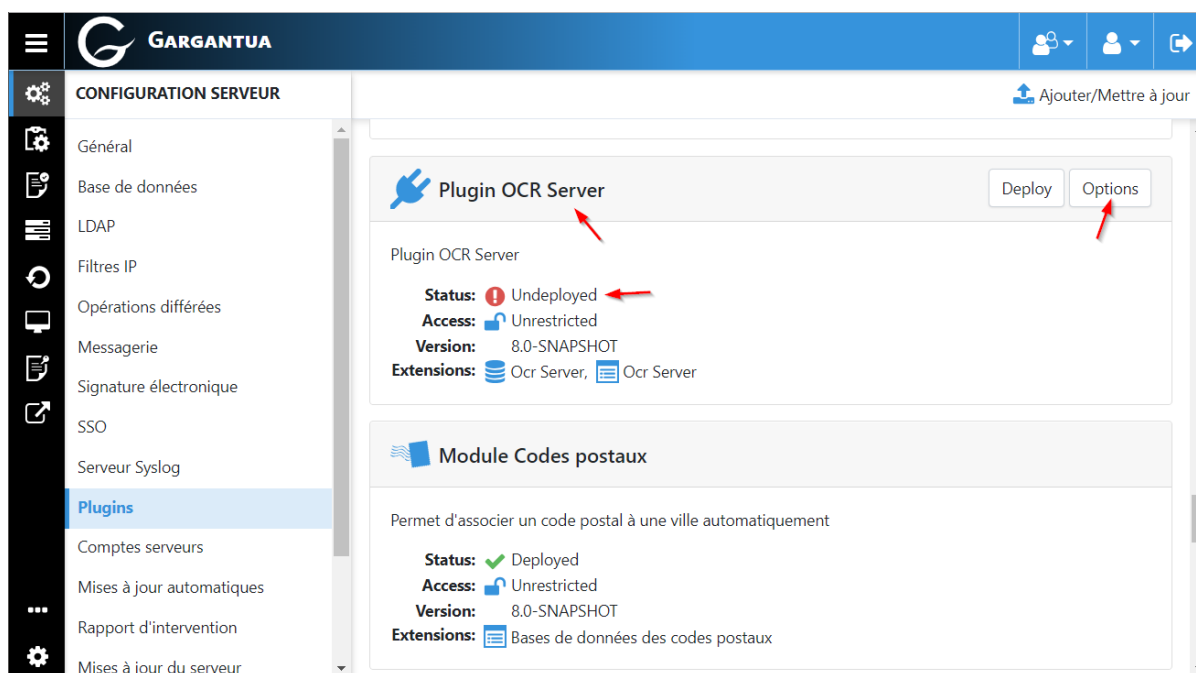
4. Choisissez la commande « Ajouter / Mise à jour » en haut et à droite de la fenêtre. La fenêtre de sélection des fichiers plugin à télécharger s'affiche.



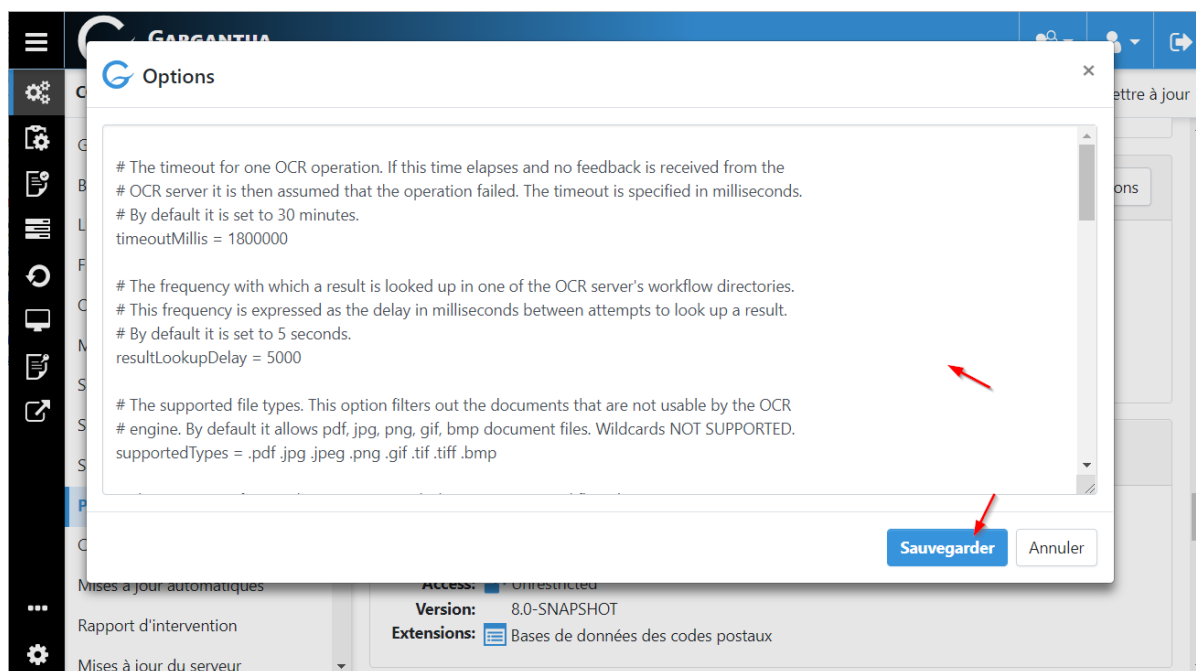
5. Sélectionnez le fichier contenant le plugin (et optionnellement le fichier de la configuration initiale) puis cliquez sur « Mettre en ligne ».



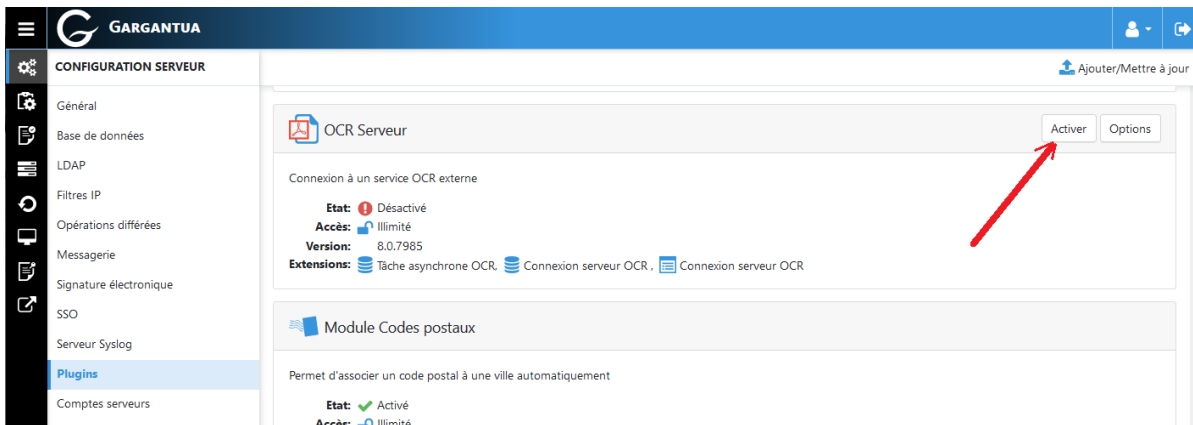
6. Vérifiez que le plugin est installé et ajustez les options de configuration en cliquant sur le bouton « Options ».



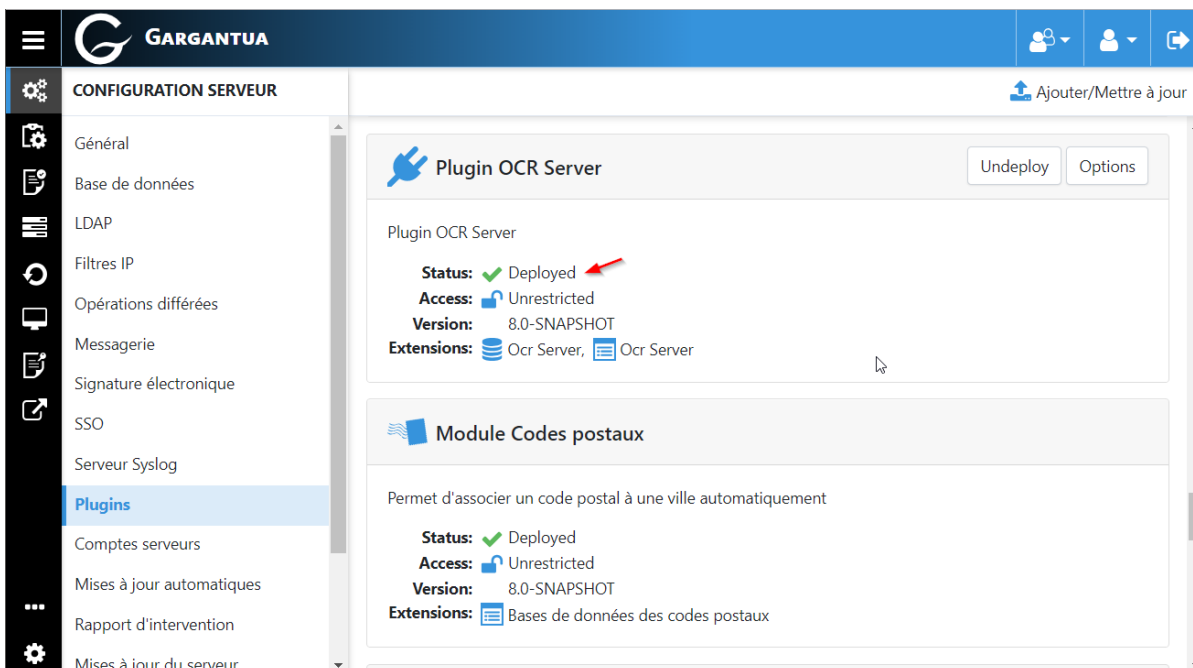
7. Une fenêtre popup s'ouvre



8. Cliquez sur 'Activer' pour démarrer le plugin.



9. Le résultat obtenu s'affiche indiquant que le plugin est déployé et actif



La publication et le fonctionnement de ce plugin sont conditionnés par la licence du serveur Gargantua. Le plugin ne pourra démarrer que si la licence le permet.

PARAMÉTRAGE DU PLUGIN 'OCR SERVEUR'

Le paramétrage du plugin de connexion de Gargantua vers le serveur OCR se trouve dans le fichier **siatel-plugin-ocrserver.config**. Ce fichier se trouve dans le dossier où les modules d'extension/plugins du service Gargantua ont été installés. Ce fichier est aussi accessible dans l'écran de gestion des plugins de service Gargantua (point 7 de la section "Installation de l'extension OCR Serveur"). Dans ce fichier, les options de configuration sont explicitées par des commentaires détaillant leurs valeurs, ainsi que l'impact d'un choix sur le fonctionnement de cette extension. Les listes d'options et leurs descriptions sont détaillées dans le tableau ci-dessous.

Tableau 1 - Options de configuration du plugin OCR Serveur

Clé	Détails
timeoutMillis	Indique le délai maximum d'attente pour l'obtention d'un résultat depuis le serveur OCR. Passé ce délai, l'opération d'océrisation s'arrête et est considérée échouée. Ce délai est spécifié en millisecondes . Si on le met par exemple à 60000, ça signifie que si un document est mis à la disposition du serveur OCR et qu'aucun résultat d'OCR n'est obtenu après 1 minute, le système ne produit pas de résultat et considère que l'opération a échoué. Par défaut, la valeur est positionnée à 1800000, soit 30 minutes.
resultLookupDelay	Fréquence de vérification des dossiers pour signaler un résultat produit par le serveur OCR. Cette fréquence est spécifiée en millisecondes . Si on la met par exemple à 5000, ça veut dire que toutes les 5 secondes, le système scrute le dossier dans lequel les fichiers océrisés sont déposés pour détecter si de nouveaux résultats ont été produits dans l'intervalle. Par défaut, cette valeur est positionnée à 5000 soit 5 secondes.
supportedTypes	Types de formats de documents acceptés par le serveur OCR. Par défaut, ce paramètre est initialisé avec les formats images les plus couramment rencontrés, soit ".pdf .jpg .jpeg .png .gif .tif .tiff .bmp". Il est possible de restreindre cette liste.
pdfPrecheck	Option permettant de vérifier au préalable le contenu d'un document pdf avant de l'envoyer au serveur OCR. Valeurs possibles : 'true' ou 'false' . Par défaut, l'option est initialisée à 'false' (par défaut, cette vérification n'est pas effectuée).
mode	Mode de fonctionnement en "local" ou en "ftp" . Si le mode choisi est "local" , les dossiers de travail sur lesquels s'appuie le serveur OCR doivent pouvoir être directement accédés via le système de fichiers de l'OS sur lequel tourne le service Gargantua. Si le mode "ftp" est choisi, ces dossiers doivent être accessibles à travers le protocole FTP (paramétrage spécifique décrit plus loin). Un service FTP doit être actif sur la machine qui héberge le service OCR et configuré pour accepter des sessions depuis la machine qui héberge le service Gargantua.
dir.upload	Nom du dossier dans lequel les documents à traiter sont mis à la disposition du serveur OCR. Par défaut, ce paramètre est initialisé avec la valeur 'UPLOAD' (dossier 'UPLOAD').
dir.download	Nom du dossier que le service Gargantua analyse pour trouver les documents convertis. Les documents convertis porteront le même nom que les documents source (au type de document près qui peut potentiellement être différent). Par défaut, ce paramètre est initialisé avec la valeur 'DOWNLOAD' (dossier 'DOWNLOAD').
dir.error	Nom du dossier que le service Gargantua analyse pour détecter les échecs de conversion. En cas d'échec, le document source sera mis dans ce dossier pour signaler que l'opération n'a pas pu être exécutée. Par défaut, ce paramètre est initialisé avec la

Clé	Détails
	valeur 'ERROR' (dossier 'ERROR').
dir.skip	Dossier dans lequel le serveur OCR déplace les documents détectés comme déjà reconnus (c'est possible pour les fichiers PDF). Ces fichiers sont ignorés par le serveur OCR s'il est paramétré pour détecter ce type de fichiers. Par défaut, ce paramètre est initialisé avec la valeur 'SKIP'.
dir.temp	Dossier tampon utilisé pour le transfert des documents vers le serveur OCR. Les documents à traiter sont d'abord transférés dans ce dossier tampon, puis ils sont déplacés dans le dossier spécifié par le paramètre 'dir.upload', ceci pour éviter que le serveur OCR ne commence à traiter un document qui ne serait pas complètement transféré. Par défaut, ce paramètre est initialisé avec la valeur 'TEMP' (dossier 'TEMP').
ftp.host	Hôte (nom ou IP) hébergeant le service OCR, accessible à travers le protocole FTP . Par défaut, ce paramètre est initialisé avec la valeur 'localhost'.
ftp.port	Port utilisé pour la communication FTP . Par défaut '21'.
ftp.basePath	Nom du dossier racine dans lequel se trouvent les dossiers de travail du service OCR (spécifiés par le options ' dir.XXX '). Par défaut '/'.
ftp.anonymous	Option permettant d'activer ou désactiver l' authentification FTP . Par défaut le paramètre est à ' false ', ce qui signifie qu'il faut spécifier un nom et un mot de passe pour authentifier les sessions FTP vers le serveur OCR, dans les options ' ftp.user ' et ' ftp.pass '. Si ce paramètre est à ' true ', les options d'authentification sont ignorées.
ftp.user ftp.pass	Nom d'utilisateur et mot de passe pour authentifier les sessions FTP vers le serveur OCR.
sock.host	Hôte (nom ou IP) hébergeant le service OCR, accessible à travers le protocole TCP , qui permet d'utiliser le serveur OCR depuis les outils d'imagerie de Gargantua. Par défaut, ce paramètre est initialisé avec la valeur 'localhost'.
sock.port	Port utilisé pour la communication TCP . Par défaut '25555'.

LES EXTENSIONS 'OCR SERVEUR'

Les fonctionnalités de l'extension « **OCR Serveur** » sont contrôlées via deux extensions qui permettent de définir les traitements qui seront effectués. Ces extensions se trouvent au niveau des bases documentaires elles-mêmes (onglet « Extensions ») et au niveau des formulaires (onglet « Extensions »). Il existe deux types d'extension :

1. L'extension '**Connexion serveur OCR**' : cette extension est accessible depuis le module

d'administration, elle s'applique à une ou à des base(s) documentaire(s) Gargantua (*voir paragraphe : niveau base de données*), à un ou des formulaire(s) (*voir paragraphe : niveau formulaire*), ou aux deux à la fois.

2. L'extension '**Tâche asynchrone OCR**' : cette extension s'applique aux bases documentaires Gargantua pour activer le lancement des tâches asynchrones de reconnaissance OCR (comme par exemple le fait de vouloir lancer la nuit la reconnaissance de volumes importants de documents)

L'EXTENSION 'CONNEXION SERVEUR OCR'

Cette extension peut être appliquée au **niveau d'une base documentaire** ou au **niveau d'un formulaire** dans une base Gargantua.

Sa mise en œuvre active l'**envoi automatique des documents ajoutés dans la base documentaire vers le serveur OCR**, ceci lors leur ajout dans la base. Les **filtres sont directement** paramétrés dans l'extension. Ils permettent :

- de spécifier l'**origine** des documents à reconnaître (uniquement ceux en provenance de la base documentaire, uniquement ceux contenus dans des processus, les deux) ;
- de spécifier les **types de documents** à reconnaître ce qui permet un niveau de contrôle supplémentaire. Deux options sont possibles : tous les types de fichiers (sous-entendu images), ou ceux dont les extensions sont de type tiff, jpeg, png....
- de définir les **actions à appliquer** aux documents convertis : **remplacer** le document original ou lui attacher une **version mineure ou majeure** ;
- de définir si le résultat de la reconnaissance sera indexé ou non en texte intégral. Attention, cela ne sera possible que si pour la base documentaire sélectionnée, l'option **Texte intégral** se trouvant dans les « Propriétés » est activée.

Quand on ajoute un document qui utilise un formulaire pour lequel l'extension « Connexion serveur OCR » est active, le paramétrage qui sera pris en compte pour l'envoi vers le serveur OCR sera celui spécifié dans le formulaire, il ne sera pas tenu compte du paramétrage de l'extension serveur OCR défini au niveau de la base ; il peut d'ailleurs ne pas être installé. On peut donc dire que le paramétrage de l'extension « serveur OCR » du formulaire est prioritaire par rapport à celui de la base (s'il existe).



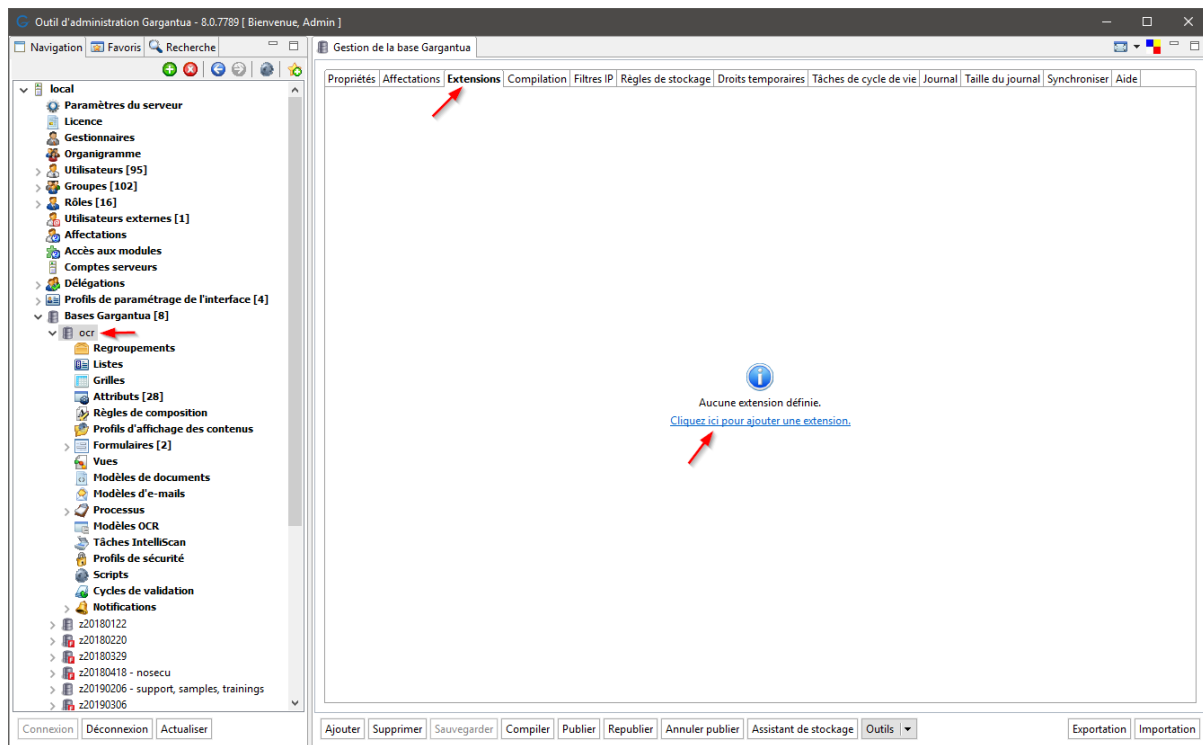
Le paramétrage d'extension OCR au niveau d'un formulaire (si défini) est prioritaire sur celui défini au niveau de la base documentaire, ceci uniquement pour les documents qui utilisent ce formulaire.

NIVEAU BASE DE DONNÉES

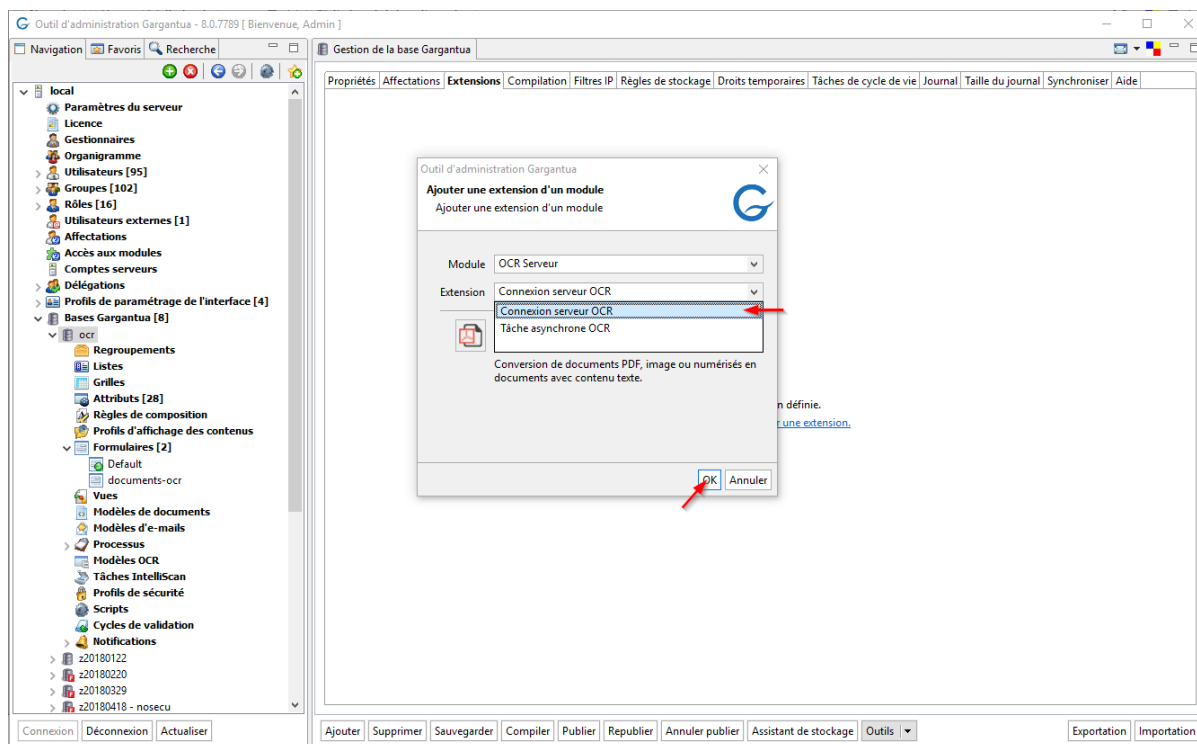
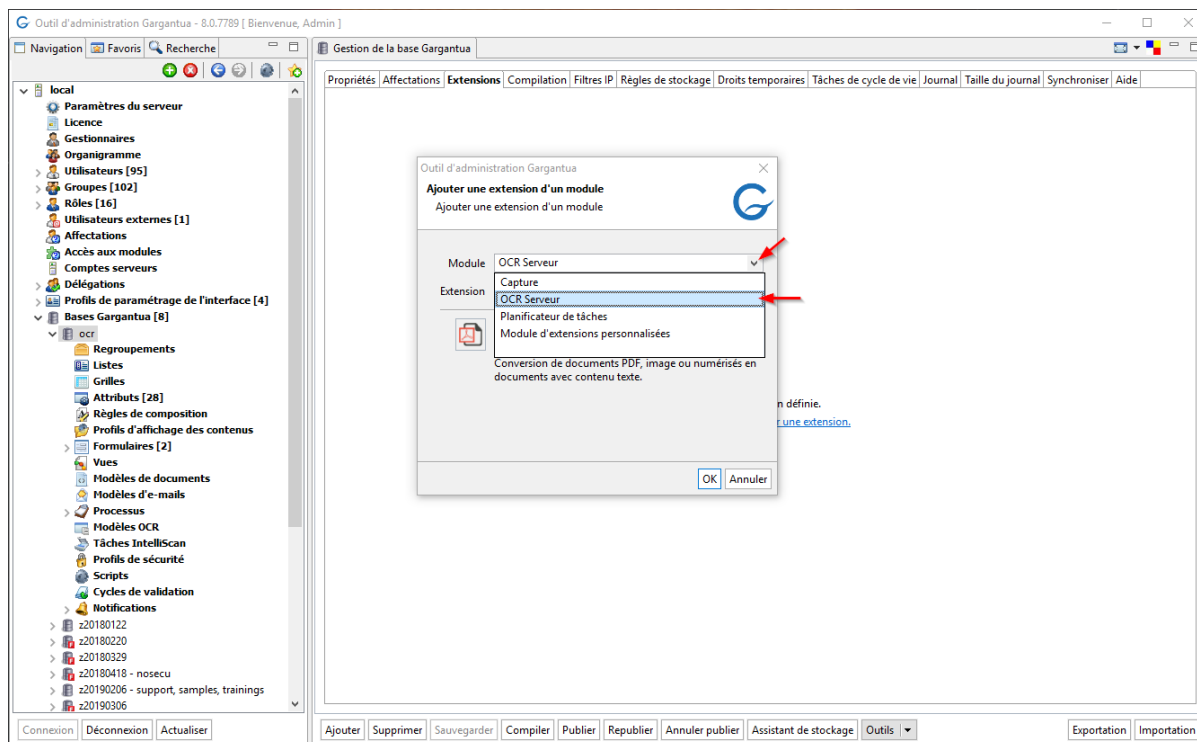
Procédure de paramétrage de l'extension « Connexion serveur OCR ».

1. Dans la console d'administration sélectionnez la base sur laquelle vous souhaitez ajouter et rendre active l'extension OCR. Pour cela, sélectionnez une base puis sélectionnez l'onglet « Extensions » et enfin cliquez sur

le bouton « Ajouter une extension ».

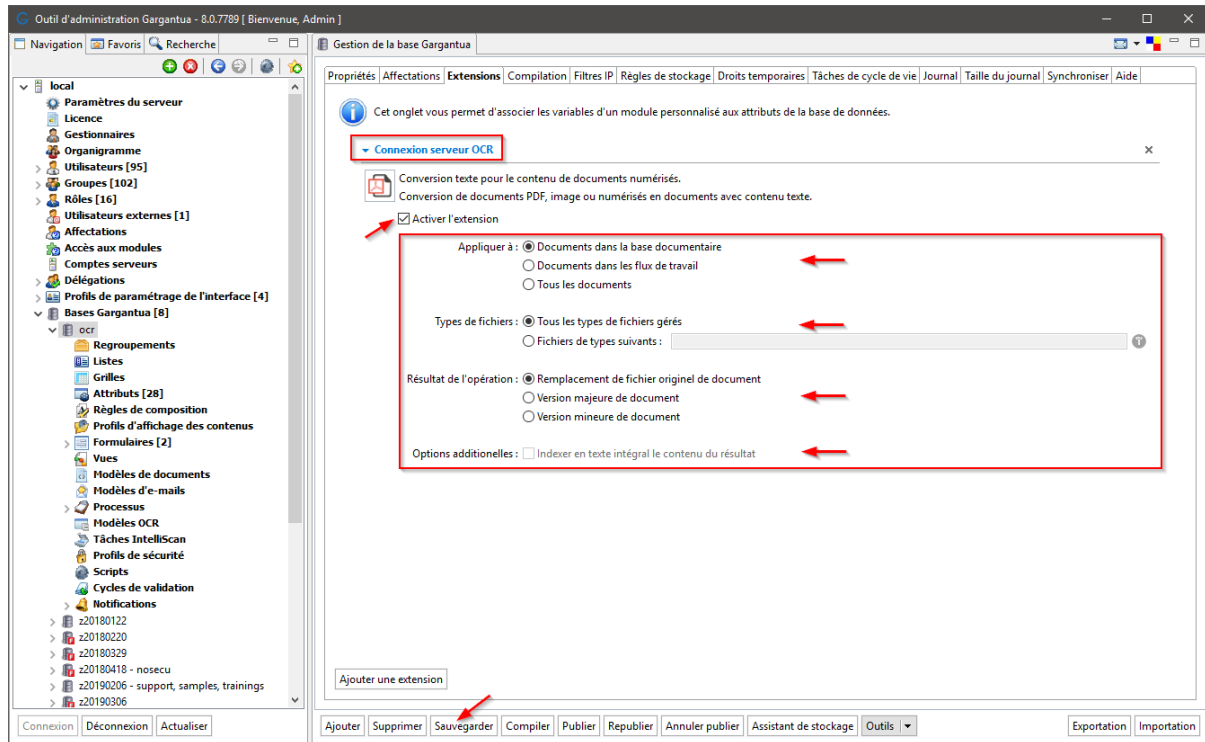


2. La boîte de sélection de l'extension s'affiche. Sélectionnez le plugin et l'extension. Cliquez 'OK' pour confirmer l'opération (ou annulez si vous ne souhaitez pas continuer).



3. Dès que l'extension est ajoutée, les options de paramétrage s'affichent. Par défaut l'extension est activée. Modifiez les options en fonction de votre objectif, puis sauvegardez les modifications. Ces options (ou les

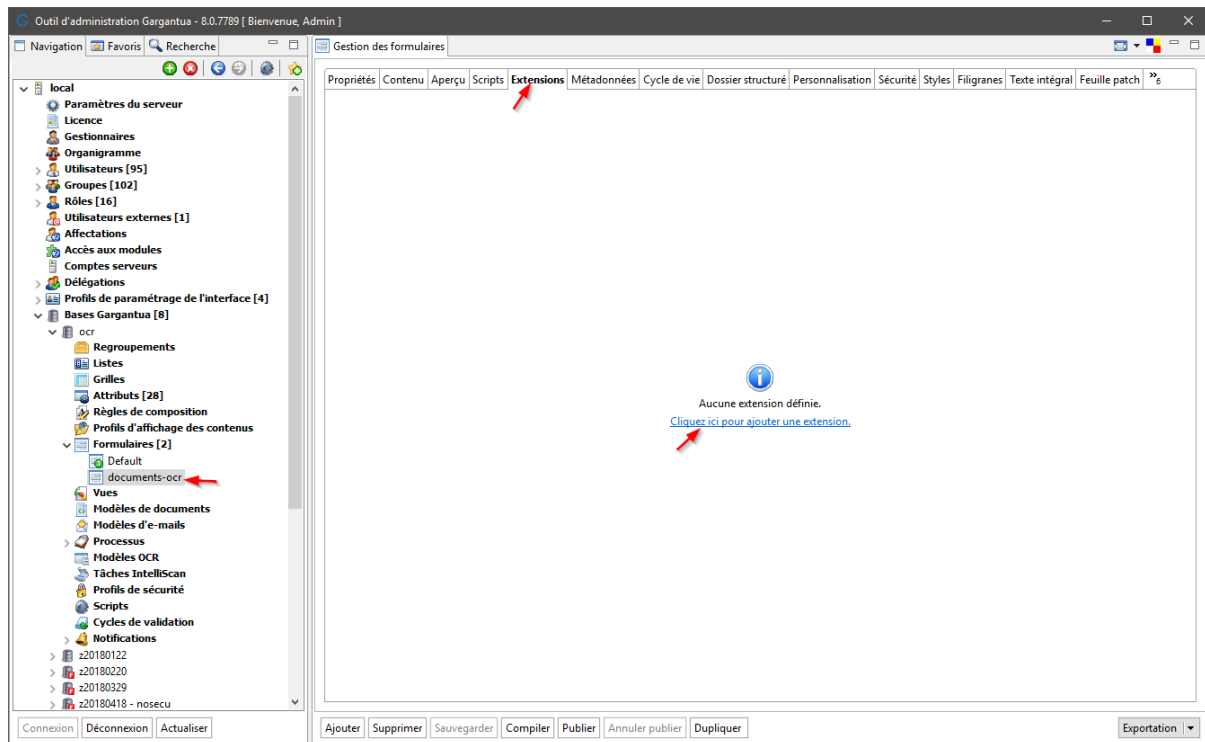
modifications) seront prises en compte avec un délai de l'ordre de 1 minute.



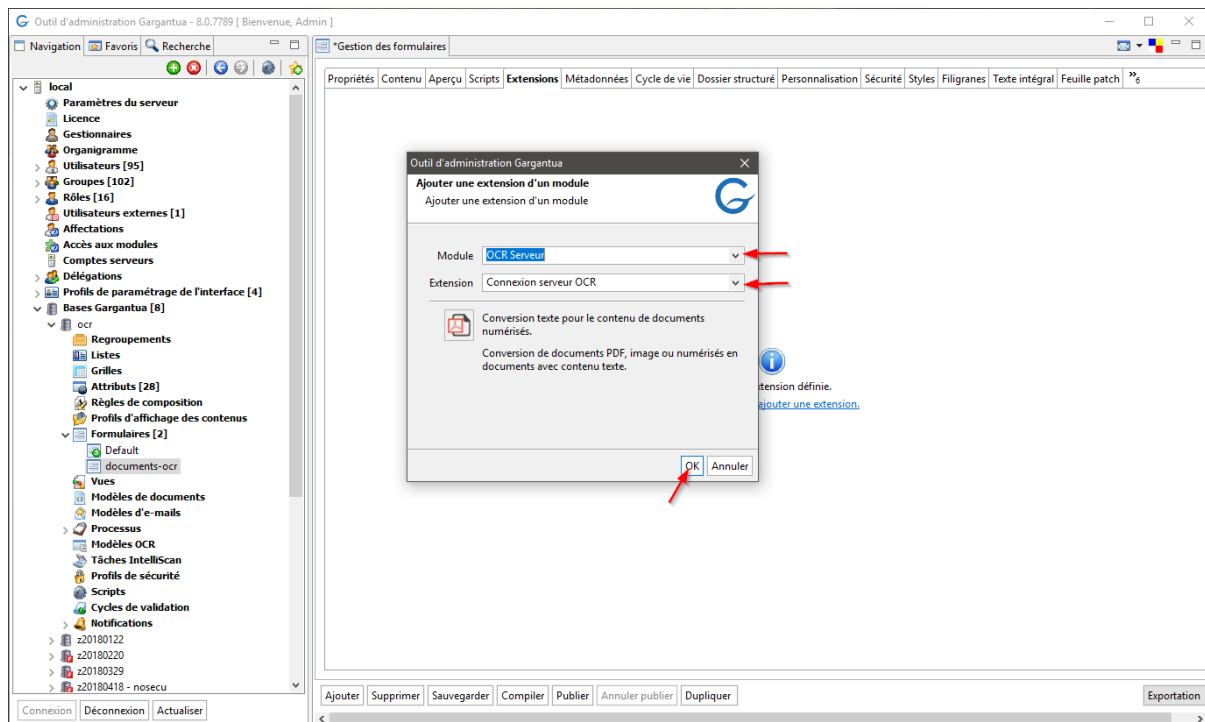
NIVEAU FORMULAIRE

Procédure de paramétrage de l'extension « Connexion serveur OCR ».

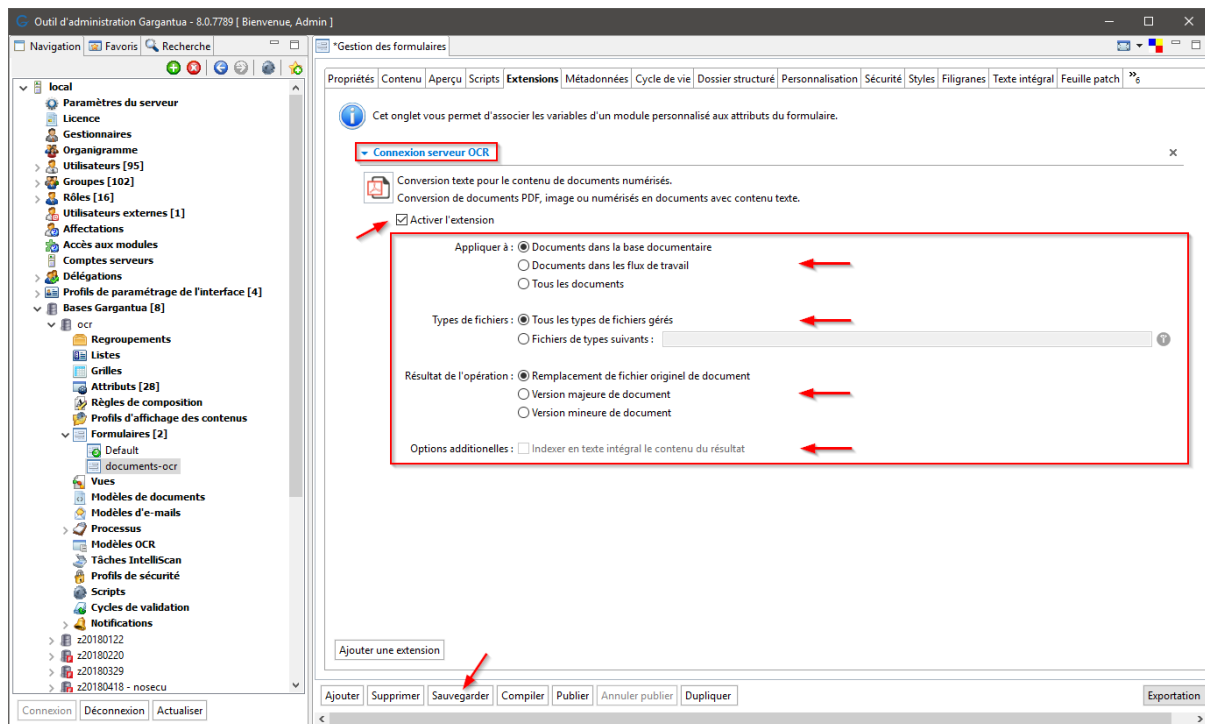
1. Dans la console d'administration, sélectionnez le formulaire sur lequel vous souhaitez ajouter et activer l'extension OCR, puis sélectionnez l'onglet 'Extensions' et enfin « Ajouter une extension ».



2. La boîte de sélection de l'extension s'affiche. Sélectionnez le plugin et l'extension. Cliquez sur 'OK' pour confirmer l'opération (ou annulez-le si vous ne souhaitez pas continuer).



3. L'extension est ajoutée et les options de paramétrage s'affichent. Par défaut l'extension est activée. Modifiez les options en fonction de votre objectif, puis sauvegardez les modifications. Ces options (ou les modifications) seront prises en compte avec un délai de maximum 1 minute.



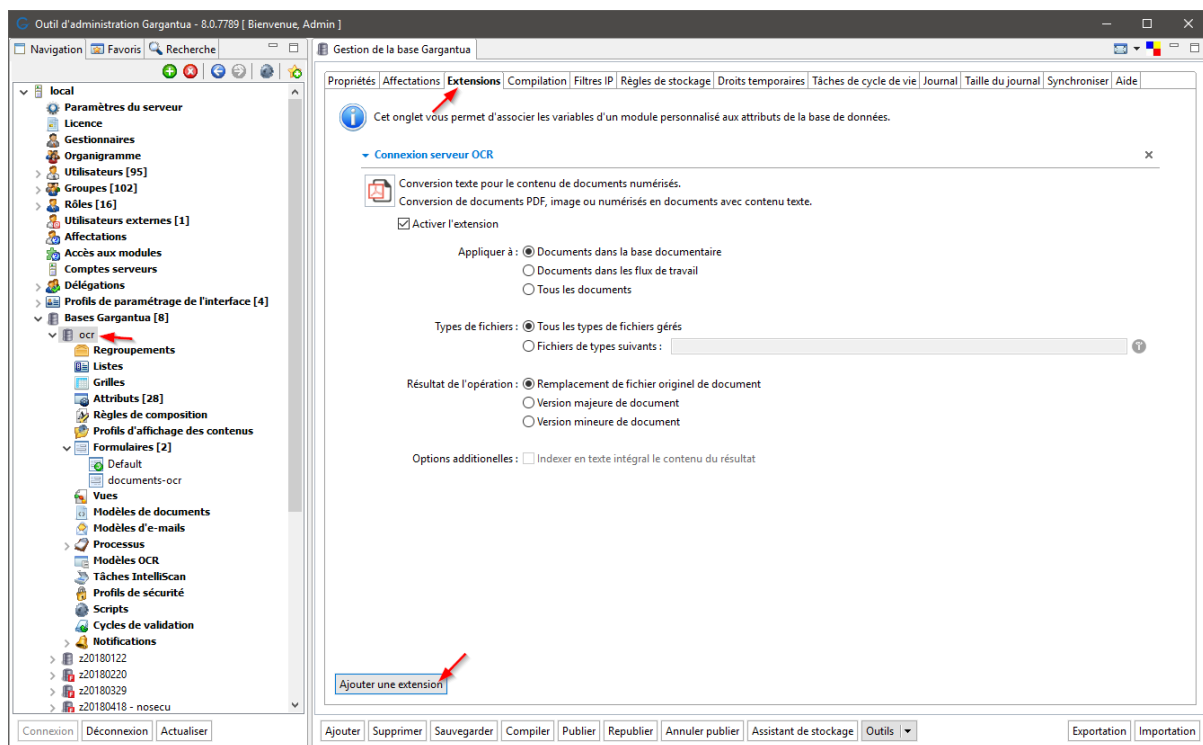
L'EXTENSION 'TÂCHE ASYNCHRONE OCR'

L'ajout de cette extension sans ajouter également l'extension 'Connexion serveur OCR' (au niveau de la base documentaire ou au niveau d'au moins un formulaire dans la base documentaire) ne va pas ajouter l'entrée correspondante dans la liste de tâches asynchrones disponibles dans l'écran d'exécution de tâches asynchrones. Les actions à appliquer aux résultats de l'opération de conversion OCR sont paramétrées par l'extension 'Connexion serveur OCR'.

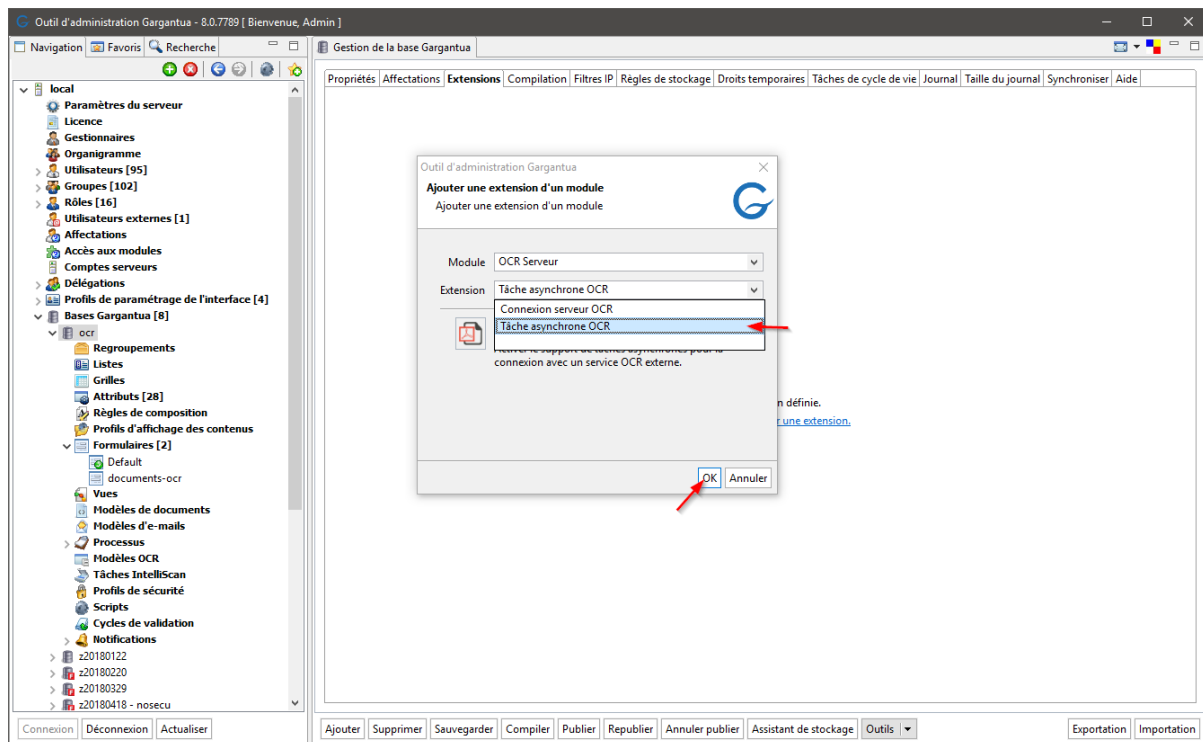
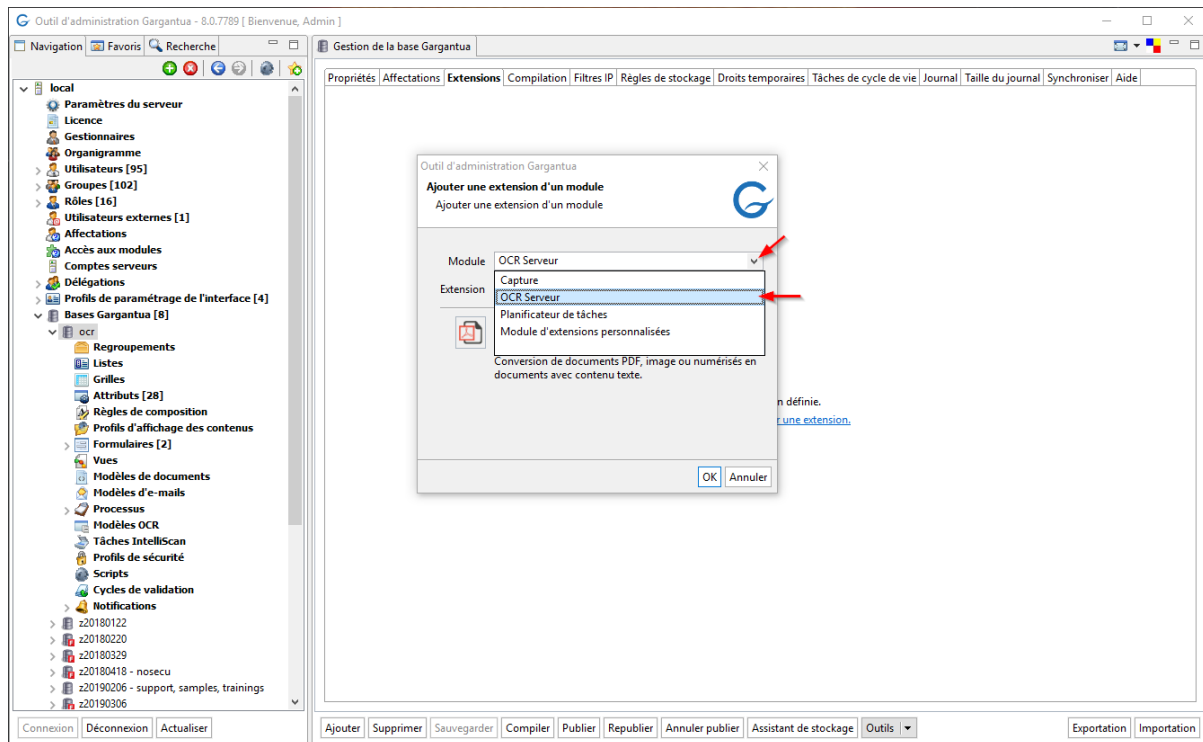
Procédure de paramétrage de l'extension 'Tâche asynchrone OCR'.

1. Dans la console d'administration, sélectionnez la ou les base(s) sur lesquelles vous souhaitez ajouter l'extension. L'écran des propriétés de la base étant affiché, sélectionnez l'onglet « Extensions », puis « Ajouter une extension ».

Cette opération se fait base par base.



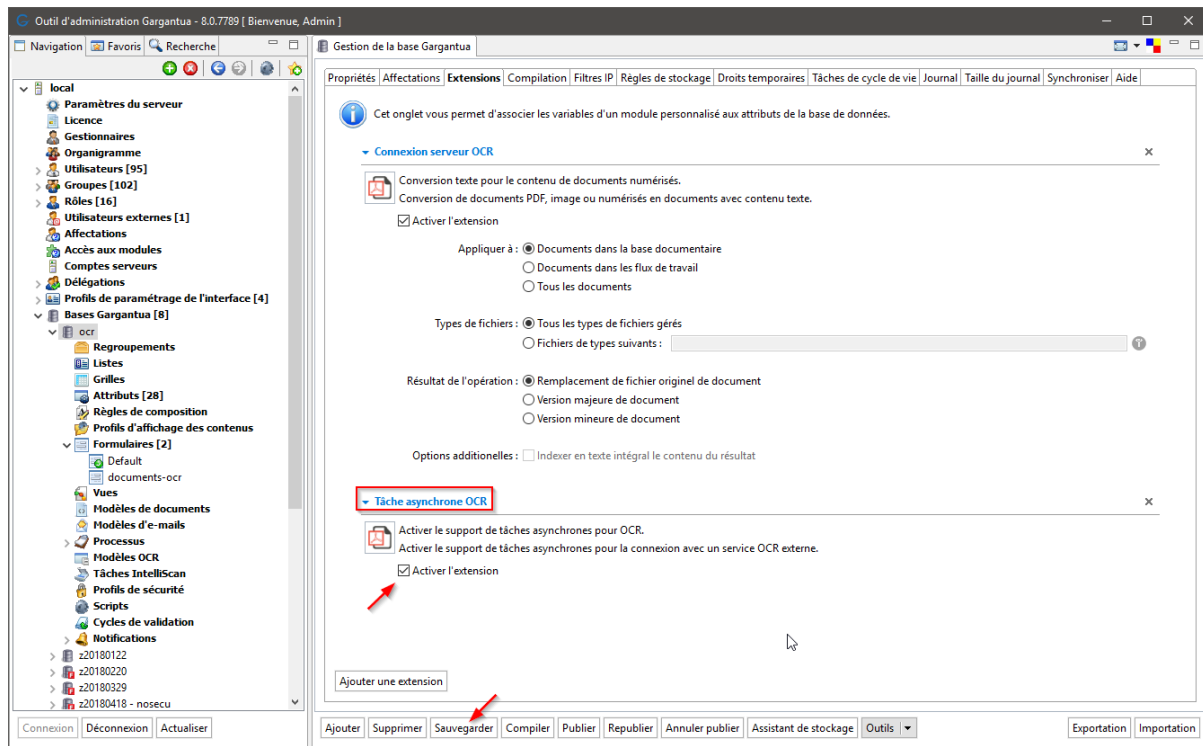
2. La boîte de sélection de l'extension s'affiche. Sélectionnez le plugin et l'extension. Cliquez sur 'OK' pour confirmer l'opération (ou annulez si vous ne souhaitez pas continuer).



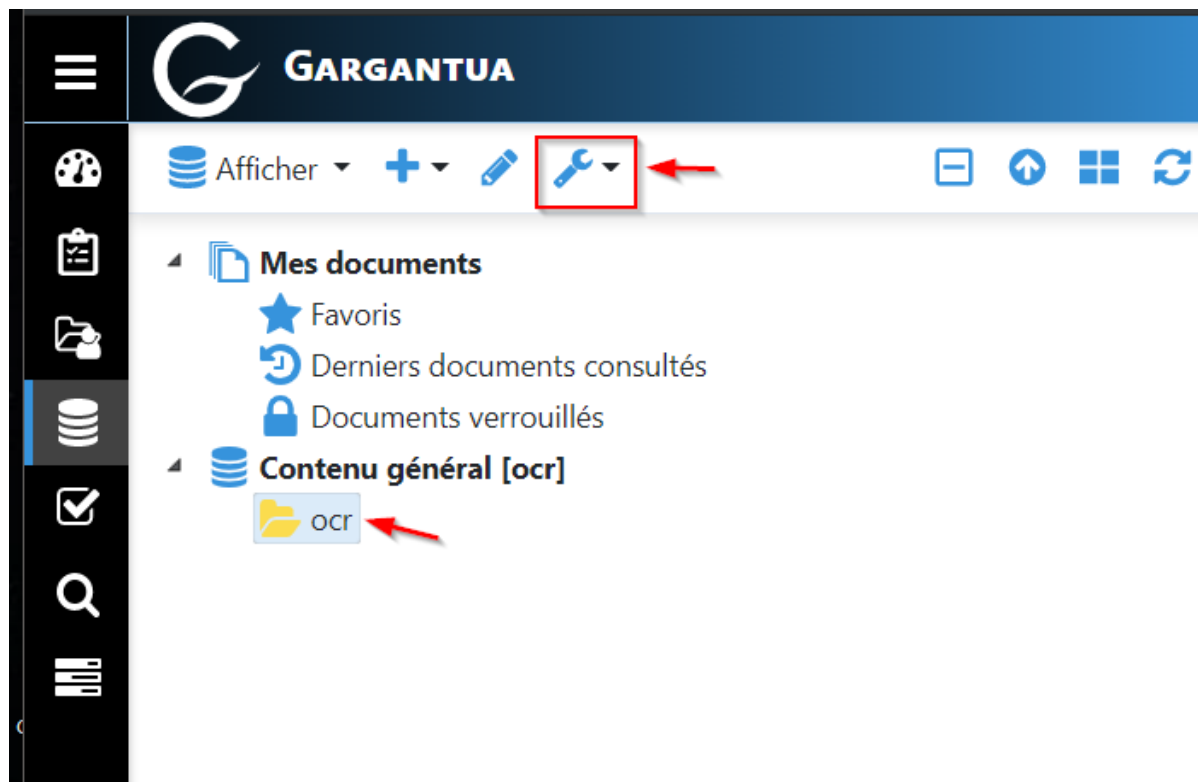
3. L'extension est ajoutée et les options de paramétrage s'affichent. Par défaut, l'extension est activée. Il n'y a pas d'autre paramétrage à effectuer, le paramétrage pour l'extension '**Connexion serveur OCR**' qui sera utilisé

dans l'exécution de tâches asynchrones est effectué par défaut.

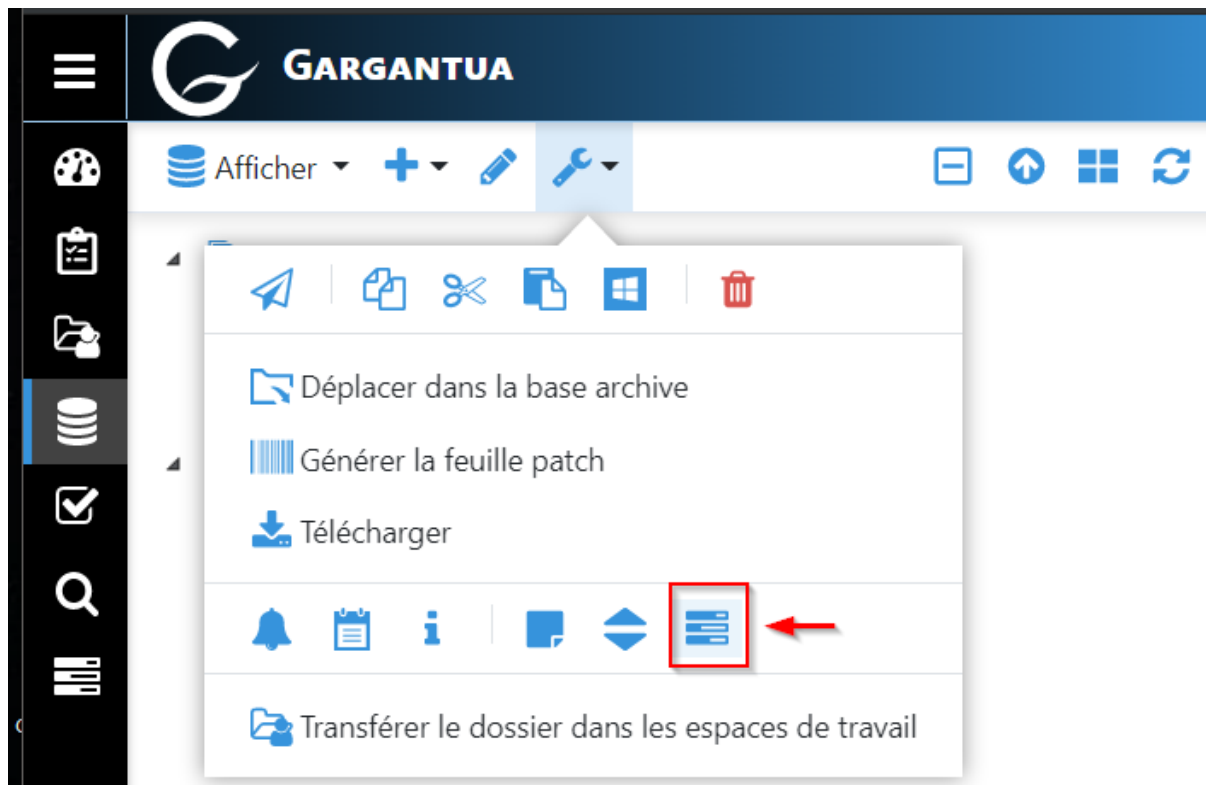
Pour les formats de fichiers acceptés, deux options sont possibles : tous les fichiers de type image ou une sélection de formats. En cas de sélection de formats, listez les de la manière suivante : .pdf; .jpg; .jpeg; .tif



4. Pour lancer une tâche asynchrone OCR, depuis l'interface utilisateur, créez ou sélectionnez un dossier dans la base documentaire, dossier dans lequel les résultats iront s'insérer.



5. Puis sur ce dossier sélectionnez l'exécution de tâches asynchrones.



6. L'écran liste les types de tâches asynchrones disponibles. Sélectionnez le type 'Plugin extension'.



7. Les options d'exécution de la tâche asynchrone s'affichent. La liste d'extensions disponibles s'affiche également. Sélectionnez l'extension 'Tâche asynchrone OCR' puis lancez la tâche avec une des commandes de démarrage ci-dessous.

Démarrer une opération différée

Nom de l'opération

Plugin extension 'ocr' 03/06/2019 17:00:05

☐ Notifications par e-mail

Démarrage

☒ Démarrer tout de suite

☐ Démarrer après [] minutes

☐ Démarrer le []

Horaires d'exécution

☒ Utiliser le planificateur des opérations différées

☐ Envoyer des notifications concernant les horaires d'exécution

Plugin extension options

☒ Récursif (appliquer aussi aux sous-dossiers)

Extensions

☒ Tâche asynchrone OCR

Retour Démarrer Démarrer et afficher l'état Annuler