

OVIDENTIA

Connecteur Swish-e / OVIDENTIA





Version Documentation	Version OVIDENTIA / Module	Date
1.0	6.0.0	15/11/2006
1.1	6.0.5	04/12/2006
1.2	6.0.5	31/01/2008

Table des matières

1 - Introduction.....	4
2 - Installation du moteur d'indexation swish-e.....	4
3 - Configuration du connecteur Swish-e dans OVIDENTIA.....	4
3.1 - Exemple dans un environnement Windows.....	5
3.2 - Exemple dans un environnement Linux.....	5
4 - Exécution et configuration du processus d'indexation.....	6
5 - Planification du processus d'indexation.....	7
5.1 - Exemple dans un environnement Linux avec wget.....	7
5.2 - Exemple dans un environnement Windows avec wget.....	7
6 - Recommandations.....	8

1 - Introduction

Le connecteur Swish-e est disponible à partir de la version 6.0.0 d'**OVIDENTIA**.

A ce jour, le connecteur Swish-e/Ovidentia n'a été validé qu'avec la version 2.4.3 de Swish-e.

L'indexation de fichiers concerne les fonctions du noyau suivantes :

- Les fichiers du gestionnaire de fichiers
- Les fichiers liés aux articles
- Les fichiers liés aux contributions de forum

OVIDENTIA gère les demandes d'indexation et la recherche plein-texte sur les types de fichiers suivants (les « Modules » Swish-e doivent être présents) :

doc, xls, xml, html, txt, csv, pdf, sxw.

La version Windows de Swish-e ne permet l'indexation que des fichiers doc, xml, html, txt, csv et pdf. Les modules xls2csv et unzip n'étant pas proposés dans les librairies lors de son installation.

Important :

OVIDENTIA délègue totalement l'indexation et la recherche plein texte à Swish-e et ses modules d'indexation (catdoc, pdftotext, unzip ...). **OVIDENTIA** n'est donc pas directement à l'origine de la pertinence de l'indexation ou des résultats remontés.

2 - Installation du moteur d'indexation swish-e

Le moteur d'indexation « Swish-e » doit être installé sur le serveur hébergeant **OVIDENTIA**..

Swish-e peut être téléchargé depuis le site : <http://swish-e.org>

Nous vous invitons à utiliser la documentation d'installation et de configuration présent sur le site de Swish-e pour son intégration sur votre serveur. <http://swish-e.org/docs/index.html>

3 - Configuration du connecteur Swish-e dans **OVIDENTIA**.

Cette étape ne doit être réalisée qu'après l'installation de Swish-e sur votre environnement.

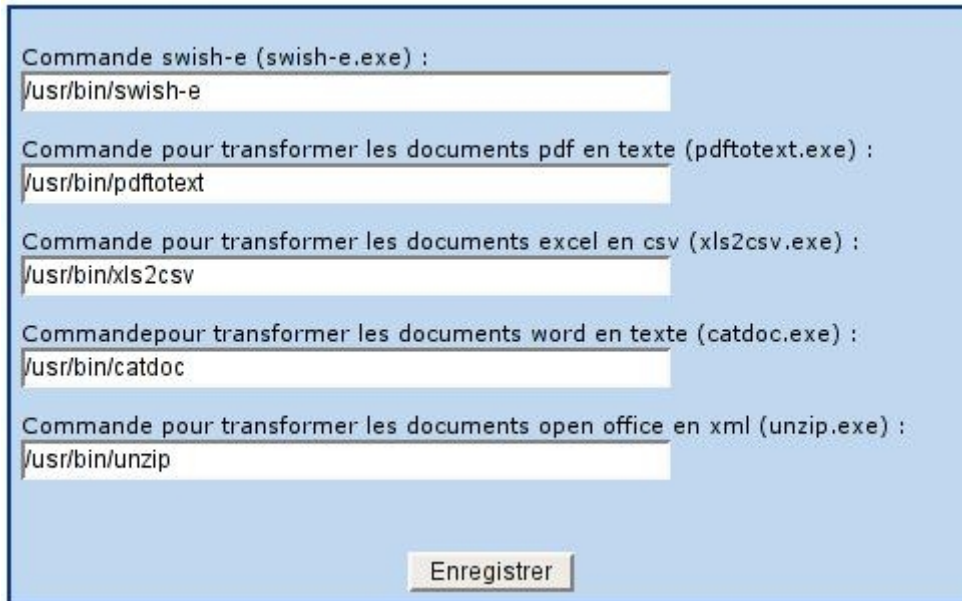
Pour activer la configuration du moteur d'indexation dans **OVIDENTIA**., il faut ajouter la ligne suivante dans le fichier config.php qui se trouve à la racine d'**OVIDENTIA**. (à placer avant la balise « ?> ») :

```
$babSearchEngine = 'swish';
```

L'accès à la configuration du connecteur Swish-e dans **OVIDENTIA** s'effectue en tant qu'administrateur du portail.

Depuis la section « Administration », cliquer sur « sites » puis sur le nom de votre site.

En sélectionnant « Configuration du moteur de recherche », **OVIDENTIA** vous présentera l'interface suivante :



Commande swish-e (swish-e.exe) :

/usr/bin/swish-e

Commande pour transformer les documents pdf en texte (pdftotext.exe) :

/usr/bin/pdftotext

Commande pour transformer les documents excel en csv (xls2csv.exe) :

/usr/bin/xls2csv

Commande pour transformer les documents word en texte (catdoc.exe) :

/usr/bin/catdoc

Commande pour transformer les documents open office en xml (unzip.exe) :

/usr/bin/unzip

Enregistrer

Illustration 1: Configuration du connecteur Swish-e / Ovidentia

Il est nécessaire de renseigner le chemin complet d'accès aux exécutable.

3.1 - Exemple dans un environnement Windows

Commande swish-e (swish-e.exe) :

C:\Progra~1\SWISH-E\swish-e.exe

Commande pour transformer les documents pdf en texte (pdftotext.exe) :

C:\Progra~1\SWISH-E\lib\swish-e\pdftotext.exe

Commande pour transformer les documents word en texte (catdoc.exe) :

C:\Progra~1\SWISH-E\lib\swish-e\catdoc.exe

....

Attention : Les noms longs ne sont pas supportés, dans notre exemple C:\Program Files devient C:\Progra~1.

3.2 - Exemple dans un environnement Linux

Commande swish-e (swish-e.exe) :

/usr/bin/swish-e

Commande pour transformer les documents pdf en texte (pdftotext.exe) :

/usr/bin/pdftotext

Commande pour transformer les documents excel en csv (xls2csv.exe) :

/usr/bin/xls2csv

Commande pour transformer les documents word en texte (catdoc.exe) :

/usr/bin/catdoc

Commande pour transformer les documents open office en xml (unzip.exe) :

/usr/bin/unzip

A noté :

Pour que Ovidentia « communique » avec Swish-e, il est nécessaire que l'utilisateur qui exécute le serveur web (Apache, IUSR_NomDuServeur ...) puissent demander l'exécution de Swish-e.

Dans un environnement IIS 6/Windows 2003 SRV, il est nécessaire de mettre le droit de lecture/exécution au fichier "cmd.exe" qui se trouve dans "windows/system32" pour l'utilisateur IUSR_NomDuServeur.

4 - Exécution et configuration du processus d'indexation

En tant qu'administrateur, depuis la section « Administration », cliquer sur l'entrée « Indexation »

Indexation

Nom	Indexer tous les fichiers	Indexer les fichiers en attente	Indexer au moment du chargement	Désactivé
Gest. de fichiers			<input checked="" type="checkbox"/>	<input type="checkbox"/>
Fichiers liés aux articles			<input checked="" type="checkbox"/>	<input type="checkbox"/>
Fichiers liés aux contributions de forum			<input checked="" type="checkbox"/>	<input type="checkbox"/>
Adresse IP autorisée :	<input type="text" value="127.0.0.1"/>			

Mettre à jour

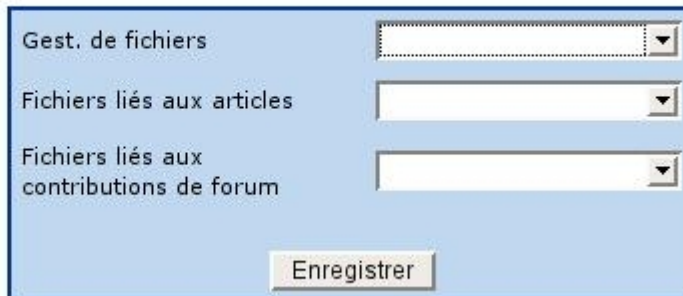
[Indexer tous les fichiers](#) - [Indexer tous les fichiers en attente](#)

Cette interface vous permet :

- De gérer l'activation de l'indexation par fonction.
- De configurer l'indexation des fichiers lors de leur chargement dans le portail.
- D'exécuter l'indexation de tous les fichiers ou uniquement des fichiers en attente d'indexation, fonction par fonction ou globalement. (Pour des raisons de sécurité, l'exécution de l'indexation nécessite au préalable de renseigner dans le champ « Adresse IP autorisée » l'adresse IP du client qui demande cette action. La planification de l'indexation (voir le paragraphe traitant ce point) nécessite de renseigner l'adresse IP du poste à l'origine de la requête planifiée (en général 127.0.0.1, la planification étant exécutée par le même serveur que celui qui héberge **OVIDENTIA.**)

En cliquant sur l'onglet « État des fichiers »

Modifier l'état d'indexation de tout les fichiers



Cela permet d'initialiser l'ensemble des fichiers du gestionnaires de fichiers, des fichiers liés aux articles ou des fichiers liés aux contributions de forum aux états suivants :

- Pas d'indexation : Les fichiers ne seront pas indexés à la prochaine indexation.
- Indexé : Les fichiers sont notés comme étant indexés.
- En attente d'indexation : Les fichiers sont notés comme étant en attente d'indexation.

5 - Planification du processus d'indexation

La planification du processus d'indexation sous-entend une ré-indexation de tous les fichiers « indexables » du serveur.

Les étapes à planifier pour automatiser l'indexation sont les suivantes :

1. Générer le fichier de configuration de l'indexation.
2. Exécuter le fichier de configuration de l'indexation.

L'utilisation de wget (<http://www.gnu.org/software/wget/>) est nécessaire à l'exécution de ce processus. wget doit être dans le PATH de votre serveur.

Remarque : si pour une raison ou une autre, le programme wget ne s'exécute pas avec l'adresse IP indiquée dans votre configuration Ovidentia, il est parfois nécessaire d'ajouter le paramètre bind en indiquant l'adresse désirée.

Exemple : wget --bind-address=127.0.0.1

5.1 - Exemple dans un environnement Linux avec wget

1. wget -q -O /dev/null '<http://VotreServeur/index.php?tg=usrindex&idx=2&prepare=1>'
2. source /VotreRepertoireUploadOvidentia/SearchIndex/index.bat

5.2 - Exemple dans un environnement Windows avec wget

1. C:\RepertoireWget\wget.exe" -q "<http://VotreServeur/index.php?tg=usrindex&idx=2&prepare=1>
2. C:\VotreRepertoireUploadOvidentia\SearchIndex\index.bat

6 - Recommandations

Le processus d'indexation étant très gourmand en ressources systèmes, il est nécessaire d'utiliser **OVIDENTIA** et Swish-e sur un serveur dédié à cette application.

De plus, pour les mêmes raisons, il est préférable de planifier le processus d'indexation (la nuit par exemple) plutôt qu'une indexation à la volée (indexation au chargement du fichier).