

<b>I. <u>IT GOVERNANCE &amp; AMELIORATION DE LA RECHERCHE</u></b>	<b>3</b>
<b>A. PRINCIPALES ETAPES</b>	<b>4</b>
<b>B. FOCUS : PROCESSUS METIERS CLEF</b>	<b>5</b>
<b>C. ORGANISATION</b>	<b>6</b>
<b>II. <u>LA RECHERCHE SOUS MOSS</u></b>	<b>7</b>
<b>A. PARAMETRES DES SERVICES DE RECHERCHE</b>	<b>7</b>
<b>B. PARAMETRES DU SHARED SERVICES PROVIDER</b>	<b>7</b>
<b>C. PARAMETRES DE LA RECHERCHE POUR UNE COLLECTION DE SITE</b>	<b>8</b>
<b>III. <u>PARAMETRAGE DES SERVICES DE RECHERCHE – CENTRAL ADMIN</u></b>	<b>9</b>
<b>A. PARAMETRAGE DE LA RECHERCHE AU NIVEAU DE LA FERME</b>	<b>9</b>
1. PARAMETRE DU SERVEUR PROXY	9
2. REGLES D'IMPACT	10
<b>B. SERVEURS DE REQUETE ET D'INDEX</b>	<b>11</b>
1. REQUETE ET INDEXATION	11
2. COMPTE DE L'INDEXEUR	12
3. FRONTAL WEB ET ANALYSE	12
<b>IV. <u>PARAMETRAGE DU SHARED SERVICE PROVIDER</u></b>	<b>14</b>
<b>A. PARAMETRES DE L'INDEXATION</b>	<b>14</b>
1. COMPTE D'INDEXATION	14
2. SOURCE DE DONNEES	14
3. LIER LES METADATA AUX PROPRIETES	17
4. TYPE DE FICHIERS	19
5. PAGE A SUPPRIMER DE LA RECHERCHE	21
<b>B. SCOPES PARTAGÉS</b>	<b>21</b>
1. REGLES	21
2. PROGRAMMATION	22
<b>C. AUTHORITATIVE PAGES - PAGE FAISANT AUTORITE</b>	<b>22</b>
1. BONNES PRATIQUES	23
2. PROPOSITION DE CLASSEMENT	23
<b>V. <u>PARAMETRAGES AU NIVEAU SITE COLLECTION</u></b>	<b>25</b>

<b>A. SITE COLLECTION : PARAMETRES DE LA RECHERCHE</b>	<b>25</b>
<b>B. SCOPES SPECIFIQUES</b>	<b>25</b>
<b>C. MOTS CLEF</b>	<b>26</b>
<b>VI. <u>POUR ALLER PLUS LOIN</u></b>	<b>28</b>
<b>A. LES MOTS PARASITES &amp; THESAURUS</b>	<b>28</b>
<b>B. OUTILS DE REPORTING</b>	<b>28</b>
<b>C. CREATION DE PAGE DE RECHERCHE SPECIFIQUE</b>	<b>29</b>
<b>D. INTEGRATION DE LA RECHERCHE DANS INTERNET EXPLORER ET OFFICE</b>	<b>30</b>
<b>E. ENREGISTREMENT D'UNE RECHERCHE ET ALERTE</b>	<b>30</b>
<b>F. LIENS UTILES &amp; COMPLETS</b>	<b>30</b>

**Avant propos:**

Ce document s'adresse aux responsables fonctionnels, coordinateurs et administrateurs de site SharePoint. Il tente de présenter simplement les différents aspects à mettre en place pour que la recherche MOSS soit pertinente aux regards des spécificités de l'entreprise.

La première partie, plutôt générale, permet de sensibiliser le lecteur sur 2 axes à discerner par la suite :

- Les paramétrages techniques nécessaires (obligatoire – relatif à l'architecture et aux règles internes).
- Les paramétrages « Métiers » (pour l'optimisation du système)

## **I. IT Governance<sup>1</sup> & Amélioration de la recherche**

La gouvernance des systèmes d'information est à la mode. Pour être simple, c'est changer la vision « Centre de coût » du système d'information en « Apport de valeur €€ ». Au niveau du moteur de recherche, cela représente la capacité à présenter des résultats pertinents par rapport au métier de l'utilisateur. L'objectif visé étant le gain de temps, la mise à disposition de l'information profitable et l'adoption du produit par les utilisateurs.

Remarque : Aujourd'hui, en ce qui concerne la recherche, le problème est réel. Nombreuses sont les personnes qui ont le réflexe de rechercher de l'information sur Google ou autre moteur de recherche web, avant de penser aux possibilités offertes par leur organisation. La mise en place d'un bon système de recherche est donc primordiale.

Il est également important de comprendre l'aspect vivant d'un portail SharePoint. Il grandit et évolue avec l'organisation. Les utilisateurs, de par leurs besoins grandissant, l'alimenteront et le feront évoluer. Si lors des phases de mise en place de SharePoint, on pense généralement à l'autonomie des utilisateurs vis-à-vis de leur site, il est plus rare de voir des questions jaillir vis-à-vis de l'organisation autour du moteur de recherche. Comment s'organiser pour que le moteur de recherche soit, et reste **efficace** et **pertinent** ?

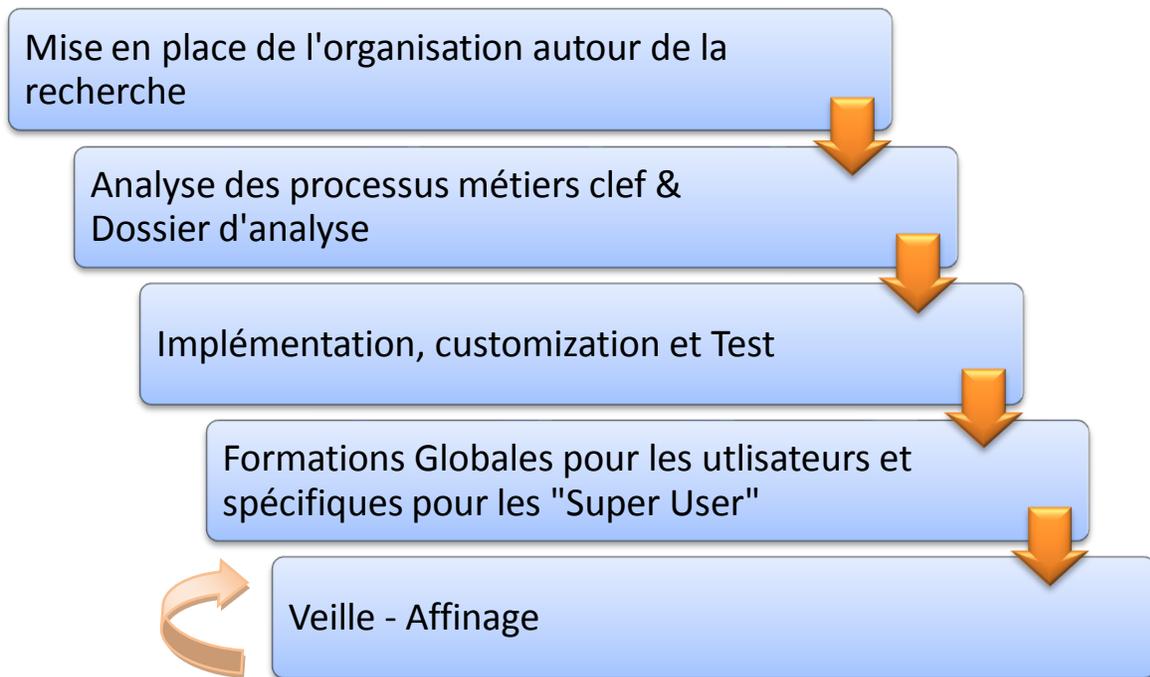
La mise en place d'un moteur de recherche fédérateur peut être vue comme un projet à part entière. D'une part, parce que c'est un outil, d'autre part à cause de sa complexité différente de celle de la mise en place d'un portail. Il est d'ailleurs fréquent de constater que des entreprises choisissent un produit spécifique dédié à la recherche.

Au travers de ce document, nous verrons la structure à adopter pour aborder ce type de projet, mais également les paramétrages clef à effectuer pour rendre MOSS Search efficace.

---

<sup>1</sup> Gouvernance des Systèmes D'information

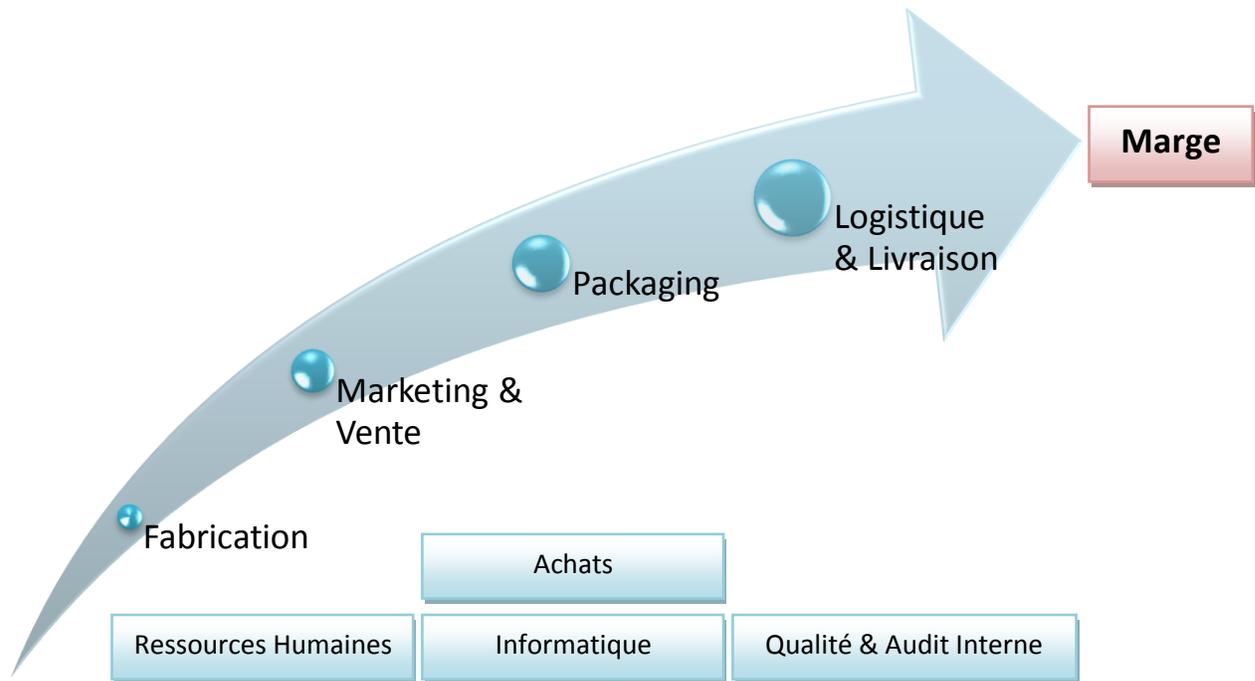
## A. Principales étapes



Trop souvent la mise en place du moteur de recherche est négligée ou mal abordée. Ces étapes, ci-dessus, bien qu'évidente, permettent de suivre un fil conducteur amenant à coup sur à une solution efficace et viable.

## B. Focus : Processus métiers clef

Les processus clef sont les activités de la chaîne de valeur de l'entreprise. La chaîne de la valeur<sup>2</sup>, c'est l'ensemble des activités de base qui permettent à l'entreprise dégager une marge. Les autres activités sont considérés comme secondaires ; elles ne contribuent pas directement à la marge de l'entreprise.



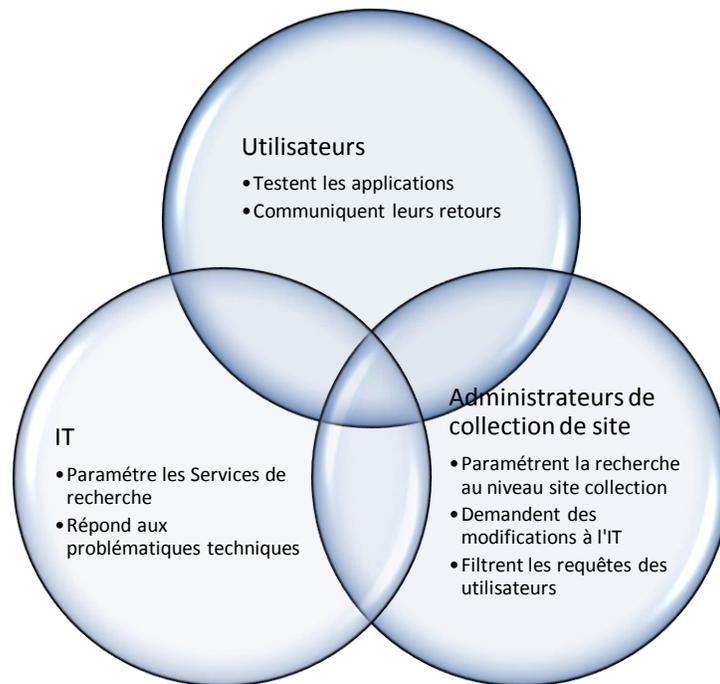
Une entreprise qui ne fait pas de marge est en difficulté. Il est donc vital pour une société de préserver l'ensemble des processus qui contribuent à sa survie ou pérennisation. Or mettre en place ou rénover un système d'information est un coût important. Il est donc normal de l'aligner sur la chaîne de valeur. Poussé à l'extrême, on parlera de refonte de système d'information : Révision des processus de l'entreprise, (modification et optimisation) afin que l'ensemble (Informatisation + automatisation + Processus Humain) forme un tout cohérent, pertinent, plus pertinent et optimisé (avec de nos jours une pointe de développement durable).

---

<sup>2</sup> Ref : Michael Porter : « L'avantage Concurrentiel » InterEditions 1986

## C. Organisation

Tout portail d'entreprise va s'étendre et grandir en même temps que la compagnie et les connaissances SharePoint des utilisateurs. Il est donc important de s'organiser en équipes et de répartir les rôles de surveillance de la cohérence, d'amélioration de la pertinence de la recherche, et ceci, tout au long de la vie du portail. Voici une proposition d'organisation simplifiée :



De cette organisation doit naître des interactions entre chaque groupe permettant l'amélioration du système.

## II. La Recherche sous MOSS

L'architecture physique étant un sujet à part entière et généralement traité avec la mise en place du portail complet, je n'en parlerai pas ici.

Les schémas suivant représentent le classement des paramètres du moteur de recherche. On les retrouve à plusieurs endroits :

- ✓ Central Admin,
- ✓ SSP (Shared Services Provider),
- ✓ Dans chaque Site Collection.

### A. Paramètres des services de recherche



### B. Paramètres du Shared Services Provider



***C. Paramètres de la recherche pour une collection de site***



### III. Paramétrage des services de recherche – Central Admin

Dans la central Admin : **Application Management** → **Search Service**

Central Administration > Application Management > Search Service

#### Manage Search Service

---

**Farm-Level Search Settings**

Proxy server used: None

Contact email address: [REDACTED]

Crawler impact rules: 0 defined

Farm-level search settings

Crawler impact rules

---

**Query and Index Servers**

Server name: W2K3

Search service: Office SharePoint Server Search Indexing and Query

SSPs hosted: None

All servers in this farm

---

**Shared Service Providers with Search Enabled**

SSP name: SharedServices2

Crawling Status: Search service offline

Items in index: 143

Propagation status: Search service offline

#### A. Paramétrage de la recherche au niveau de la ferme

##### 1. Paramètre du serveur Proxy

Si des serveurs sont à l'extérieur de la ferme SharePoint, Il se pourrait qu'ils soient accessibles uniquement en passant par un serveur Proxy. Constituer une liste des serveurs distants à indexer en précisant le proxy.

Exemple :

- ✓ **Proxy Name:** proxy
- ✓ **Port:** 3128
- ✓ **Bypass proxy server for local (intranet) addresses** → Yes

Server Name	URL
Ex: server1	http://server1*

Dans la central Admin : **Application Management** → **Search Service**

On the first section click: **“Farm-Level Search Settings”**

Central Administration > Application Management > Search Service > Farm-Level Search Settings

### Manage Farm-Level Search Settings

\* Indicates a required field

<p><b>Contact E-mail Address</b> Specify an e-mail address that external site administrators can contact if problems arise when their site is being crawled.</p>	<p>E-mail Address: *  <input type="text" value="admin@compagny.com"/>  <small>Example: someone@example.com</small></p>
<p><b>Proxy Server Settings</b> Configure proxy server settings to use when crawling other servers.</p>	<p> <input type="radio"/> Do not connect by using a proxy server  <input checked="" type="radio"/> Use the proxy server specified         </p> <p>Address: *  <input type="text" value="proxy"/></p> <p>Port:  <input type="text" value="3130"/></p> <p><input checked="" type="checkbox"/> Bypass proxy server for local (intranet) addresses.</p> <p>Do not use proxy server for addresses beginning with:  <input type="text" value="http://compagny"/>  <small>Use semicolons (;) to separate entries.</small></p>
<p><b>Timeout Settings</b> Type the amount of time that the search server will wait while connecting to other services.</p>	<p>Connection time (in seconds): *  <input type="text" value="60"/></p> <p>Request acknowledgement time (in seconds): *  <input type="text" value="60"/></p>
<p><b>SSL Certificate Warning Configuration</b> Select this checkbox only if you trust that sites are legitimate even if their certificate name does not exactly match.</p>	<p><input type="checkbox"/> Ignore SSL certificate name warnings</p>

## 2. Règles D’impact

L’indexation est consommatrice de ressource. Indexer un serveur peut ralentir ce dernier et donc nuire aux performances. Pour augmenter la disponibilité du serveur indexé, il est possible de paramétrer des règles d’indexation.

Server Name	Rules
Ex: Server1	Simultaneous requests: 8 OR Time to wait between request: 1 seconds

Dans la central Admin : **Application Management** → **Search Service**

Click on: **“Crawler Impact Rules”** then click **“Add a rule”**

Central Administration > Application Management > Search Service > Crawler Impact Rules > Add Crawler Impact Rule

## Add Crawler Impact Rule

\* Indicates a required field

<p><b>Site:</b> Type the name of a site. Do not include the protocol (for example 'http://').</p>	<p>Site: *</p> <input type="text" value="moss:2007/sites/Projects"/>
<p><b>Request Frequency</b> Indicate how the crawler will request documents from this site. For minimal impact, request fewer documents simultaneously or set a delay between requests.</p>	<p><input checked="" type="radio"/> Request up to the specified number of documents at a time and do not wait between requests Simultaneous requests: <input type="text" value="8"/></p> <p><input type="radio"/> Request one document at a time and wait the specified time between requests Time to wait (in seconds) <input type="text"/></p>

### B. Serveurs de requête et d'index

Cette section permet de paramétrer les rôles des serveurs et des informations de bases pour l'indexation comme le compte.

#### 1. Requête et indexation

Pour de meilleures performances et suivant la taille de la ferme mise en place, on découpe souvent les rôles des serveurs. Ainsi il est possible d'avoir un serveur dédié aux requêtes et à l'indexation, ou de multiples serveurs (par exemple : un pour l'indexation et un pour les requêtes). Grâce à cet écran, on peut sélectionner le rôle de chaque serveur vis-à-vis de la recherche.

Deux chemins pour y accéder :

- ✓ Central Admin: Application Management → Search → Manage Search Service - On the first section click: "Request server and Index" → "Index service and Request ..."
- ✓ Central Admin: Opération → services on server (Cet écran, plus complet donne accès à la configuration des services disponibles pour l'ensemble des serveurs)

Server Name	Request	Index
Application Server	No	Yes
Frontend 1	Yes	No
Frontend 2	Yes	No

## 2. Compte de l'indexeur

C'est un point important. Le serveur indexera les données des différents serveurs grâce à ce compte. Il doit pour cela avoir des droits administrateurs afin de pouvoir ouvrir tous les fichiers et indexer leur contenu.

Server name	Permission
Ex: w2k7	Domain Admin

**Note:** Ce Compte doit être différent de celui du Pool d'application. Sinon, le résultat de la recherche retournera toutes les versions d'un document au lieu d'un seul (pour les bibliothèques dont le versionning mineur est activé.)

## 3. Frontal Web et analyse

Permet de sélectionner quels serveurs effectueront l'analyse. Dans le cas où le serveur est dédié à la recherche et n'exécute pas d'autres services partagés, utiliser le comme web frontal pour l'analyse. Si vous avez un web frontal dédié, ou que ce serveur utilise des services partagé comme Excel services, choisir un autre web frontal.

Server	Index (Yes/No)
Ex: w2k7	Yes

<b>Query and Indexing</b> Use this option to specify if you want to use this server for search queries or indexing or both.	<input checked="" type="checkbox"/> Use this server for indexing content <input checked="" type="checkbox"/> Use this server for serving search queries
<b>Contact E-mail Address</b> Specify an e-mail address that external site administrators can contact if problems arise when their site is being crawled. This setting applies to all servers in the farm.	E-mail Address: <input type="text"/> Example: someone@example.com
<b>Farm Search Service Account</b> The search service will run using this account. Setting or changing this account affects all index and query servers in the server farm.  The farm search service account must not be a built-in account for security reasons and for it to access the database and content index. Examples of built-in accounts are Local Service and Network Service.	Select an account for this component: <input type="radio"/> Predefined <input type="text" value="Network Service"/> <input checked="" type="radio"/> Configurable User name: <input type="text" value="w2k3\administrator"/> Password: <input type="password"/>
<b>Index Server Default File Location</b> The search index will be located at this path by default on this server. For index servers, you can specify a different path when you create a Shared Services Provider. To change this index file location for an existing Shared Services Provider, use the command stsadm.exe -o editssp.	Default index file location: <input type="text" value="C:\Program Files\Microsoft Office Serv"/>
<b>Indexer Performance</b> Indexing information can place a large load on the local SQL Server database and might slow down the responsiveness of the local SharePoint sites. However, reducing the maximum allowed indexing activity will slow down the speed at which items are indexed, and therefore might cause search results to be outdated. Use information about the local server load to select the appropriate indexer performance level.	<input checked="" type="radio"/> Reduced <input type="radio"/> Partly reduced <input type="radio"/> Maximum
<b>Web Front End And Crawling</b> Use this option to specify a dedicated web front end for crawling. Crawling through a dedicated web front end will reduce the impact of crawling on the other web front ends in the farm.  If your index server is not running other shared services, it is recommended to enable the web front end role on this computer and use it as the dedicated web front end for crawling.  If your index server is also running the Excel Calculation service or other shared services, select no dedicated web front end for crawling. Otherwise, these services may not work as expected.	<input type="radio"/> Use all web front end computers for crawling <input checked="" type="radio"/> Use a dedicated web front end computer for crawling <input type="text" value="W2K3"/>

## IV. Paramétrage du Shared Service Provider

### Crawl Settings



Indexing status:	Computing ranking
Items in index:	143
Errors in log:	1
Content sources:	1 defined (Local Office SharePoint Server sites)
Crawl rules:	0 defined
Default content access account:	w2k3\administrator
Managed properties:	127 defined
Search alerts status:	Active
Propagation status:	Propagation not required

- Content sources and crawl schedules
- Crawl rules
- File types
- Crawl logs
- Default content access account
- Metadata property mappings
- Server name mappings
- Search-based alerts
- Search result removal
- Reset all crawled content

### Scopes



Scopes:	2 defined (People, All Sites)
Update status:	Idle
Update schedule:	Automatically scheduled
Next scheduled update:	1 second
Scopes needing update:	0

- View scopes
- Start update now

### Authoritative Pages



Authoritative pages:	1 defined (http://geveroffice/)
Non-authoritative pages:	0 defined

- Specify authoritative pages

### A. Paramètres de l'indexation

#### 1. Compte d'indexation

Ce compte doit être différent de celui du Pool d'application. Voir le paragraphe : **Error! Reference source not found.**

## 2. Source de données

Afin d'améliorer la recherche il est important d'identifier et de créer différentes sources de données.

### a) Travail à effectuer

- Lister les contenus qui doivent être analysés et indexés. Le résultat doit apparaître dans la recherche globale.
- Lister le contenu qui ne doit pas être analysé ni indexé.
- Lister les contenus qui doivent apparaître dans des recherches spécifiques.

### b) Quels sont les types de source de données ?

Type de Source de contenu	Type de contenu
Sites SharePoint	<ul style="list-style-type: none"> <li>• Sites SharePoint</li> <li>• Sites du Site Directory</li> </ul>
Sites Web	<ul style="list-style-type: none"> <li>• Site web de l'entreprise</li> <li>• Site Web d'internet</li> </ul>
Fichiers Partagés	<ul style="list-style-type: none"> <li>• Fichier avec un contenu important partagé au sein de l'entreprise</li> </ul>
Données métiers	<ul style="list-style-type: none"> <li>• Information Structurée</li> </ul>
Dossier Public Exchange	<ul style="list-style-type: none"> <li>• Contenu Exchange</li> </ul>

### c) Options des sources de contenu?

Type de source de contenu	Options
Sites SharePoint	<ul style="list-style-type: none"> <li>• Tout contenu en dessous du nom d'Hôte</li> <li>• Seulement les sites SharePoint contenant le même début d'adresse</li> </ul>
Sites Web	<ul style="list-style-type: none"> <li>• Seulement sur les serveurs de chaque adresse de chaque page principale.</li> <li>• Seulement la première page de chaque adresse de départ</li> <li>• Spécifique — spécification de la profondeur des pages à analyser. Par défaut illimité</li> </ul>
Fichiers Partagés	<ul style="list-style-type: none"> <li>• Le répertoire et le sous-répertoire de chaque adresse de départ</li> <li>• Seulement le répertoire de chaque adresse de départ</li> </ul>
Dossier Public Exchange	<ul style="list-style-type: none"> <li>• Le répertoire et le sous-répertoire de chaque adresse de départ</li> <li>• Seulement le répertoire de chaque adresse de départ</li> </ul>

## d) Automatisation

Les critères les plus importants à considérer lors de la planification de l'indexation sont la pertinence de l'information et l'impact sur la performance.

L'ajout de multiples sources de contenu et l'augmentation de la fréquence de l'indexation a un impact direct sur la qualité des résultats, plus riche, plus nombreux, mais aussi sur les performances de la ferme. En effet, l'indexation est le processus consommant le plus de ressource.

C'est donc avec une attention toute particulière qu'il faut paramétrer le moteur de recherche.

- **Indexation Full** : Importante, elle indexe et parcourt l'ensemble de la source de données. Ce processus est gourmand et long, mais impératif.
- **Indexation Incrémentale** : Met à jour l'index uniquement avec les éléments nouveaux ou modifiés.

On programme généralement l'indexation de la manière suivante : Full scan 1 fois par semaine et incrémentale chaque jour. **ATTENTION** : Cet exemple n'est pas valable pour les grandes quantités de données. Une étude sérieuse et des tests doivent être effectués au préalable.

**NOTE** : La durée de l'indexation est liée à beaucoup de paramètres (rapidité du/des serveurs, réseaux, ..). L'un des principaux est le type de documents indexés. L'indexation de 100'000 documents word sera plus longue que l'indexation de 100'000 images. Pourquoi ? Parce que le moteur de recherche, pour classer et indexer des documents spécifiques, ouvrent ces derniers avec le IFILTER<sup>3</sup> correspondant. L'IFILTER est une sorte de traducteur qui lui permet de comprendre le contenu du fichier. L'ouverture, plus la traduction, est donc plus lente que la simple indexation du nom et des metadata. (Exemple : 2 jours minimum pour 500 Gb )

## e) Tableau à remplir

Il est important d'abord d'avoir une vision globale de se que l'on souhaite indexer et à quelle fréquence:

Content Source Name	Type of content	Start Addresses	Type of crawling	Full Crawl Schedule	Incremental Schedule
Local Office SharePoint Sites	SharePoint Sites	All Sites collections	Recursively	1 per week	1 per day
Files Share	Files	\\Fileserver	Recursively	1 per month	1 per day
...					

En fonction des besoins, il faut trouver l'équilibre entre la quantité d'informations à indexer et la fréquence (↗ données ↗ fréquences = ↘ Performance).

<sup>3</sup> Voir IV.A.4

Use this page to add a content source.

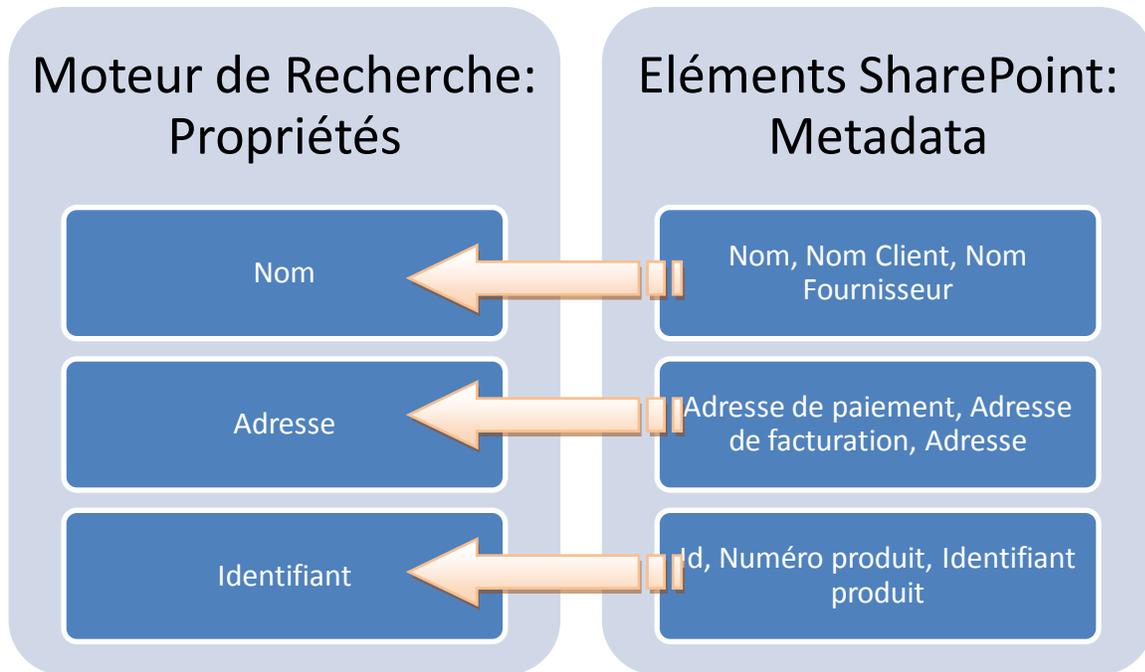
\*Indicates a required field

<p><b>Name</b></p> <p>Type a name to describe this content source.</p>	<p>Name: *</p> <input type="text"/>
<p><b>Content Source Type</b></p> <p>Select what type of content will be crawled.</p> <p>Note: This cannot be changed after this content source is created since other settings depend on it.</p>	<p>Select the type of content to be crawled:</p> <p><input checked="" type="radio"/> SharePoint Sites</p> <p><input type="radio"/> Web Sites</p> <p><input type="radio"/> File Shares</p> <p><input type="radio"/> Exchange Public Folders</p> <p><input type="radio"/> Business Data</p>
<p><b>Start Addresses</b></p> <p>Type the URLs from which the search system should start crawling.</p> <p>This includes all Office SharePoint Server sites and Windows SharePoint Services sites.</p>	<p>Type start addresses below (one per line): *</p> <div style="border: 1px solid gray; height: 60px; width: 100%;"></div> <p>Example: http://intranetsite</p>
<p><b>Crawl Settings</b></p> <p>Specify the behavior for crawling this type of content.</p> <p>Selecting to crawl everything under the hostname will also crawl all the SharePoint Sites in the server.</p>	<p>Select crawling behavior for all start addresses in this content source:</p> <p><input checked="" type="radio"/> Crawl everything under the hostname for each start address</p> <p><input type="radio"/> Crawl only the SharePoint Site of each start address</p>
<p><b>Crawl Schedules</b></p> <p>Select the crawl schedules for this content source.</p>	<p>Select the schedule that this should be a part of:</p> <p>Full Crawl  <input type="text" value="None"/> <a href="#">Create schedule</a></p> <p>Incremental Crawl  <input type="text" value="None"/> <a href="#">Create schedule</a></p>
<p><b>Start Full Crawl</b></p> <p>Select "Start full crawl of this content source" and click "OK" to start a full crawl of this content source.</p>	<p><input type="checkbox"/> Start full crawl of this content source.</p>

### 3. Lier les metadata aux propriétés

Les metadata liées aux éléments (fichier, liste) de SharePoint constituent une source d'information importante pour les utilisateurs. Elles sont le reflet de leur travail quotidien. Il est donc primordial et naturel que l'indexation des éléments prenne en compte ces metadata.

Le moteur de recherche indexe le contenu des sites selon des propriétés. Chacune d'entre elles sont liées à un ou plusieurs metadata. Le but étant de regrouper les metadata ayant une portée sémantique équivalente ou au contraire créer une nouvelle propriété.



Properties	Metadata
Ex : LastName	People:LastName(Text), ows_LastName(Text), ows_Last_x0020_Name(Text)

Shared Services Administration: SharedServices2 > Search Settings > Managed Properties  
**Metadata Property Mappings**

Crawled properties are automatically extracted from crawled content. Users can perform queries over managed properties. Use this page to create and modify managed properties and map crawled properties to managed properties. Changes to properties will take effect after the next full crawl.

**Managed Properties View**

Property Name	Type	May be deleted	Use in scopes	Mappings
AboutMe	Text	Yes	No	People>AboutMe(Text), ows_Notes(Text)
Account	Text	Yes	No	ows_Name(Text)
AccountName	Text	Yes	No	People:AccountName(Text)
AssignedTo	Text	Yes	No	ows_AssignedTo(Text), ows_Assigned_x0020_To(Text)
Assistant	Text	Yes	No	People:Assistant(Text)
Author	Text	No	Yes	Mail:6(Text), Office:4(Text)
Authority	Decimal	Yes	No	
BestBetKeywords	Text	Yes	No	SharePoint:BestBetKeywords(Text), SharePoint:HomeBestBetKeywords(Text)
Birthday	Date and Time	Yes	No	People:SPS-Birthday(Date and Time)
CachedPath	Text	Yes	No	Basic:26(Text)
CategoryNavigationUrl	Text	Yes	No	Basic:CategoryUrlNavigation(Text)
CollapsingStatus	Integer	No	No	
Company	Text	Yes	No	ows_Company(Text)
contentclass	Text	Yes	Yes	DAV:contentclass(Text), DAV:contentclass(Text), DAV:contentclass(Text), ...
ContentSource	Text	No	Yes	

#### 4. Type de fichiers

Type de fichier parcouru par défaut dans SharePoint:

File type extension	Default IFilter support	Default file type inclusions
<b>ASCX</b>	Yes	<b>Yes</b>
<b>ASM</b>	Yes	<b>No</b>
<b>ASP</b>	Yes	<b>Yes</b>
<b>ASPX</b>	Yes	<b>Yes</b>
<b>BAT</b>	Yes	<b>No</b>
<b>C</b>	Yes	<b>No</b>
<b>CMD</b>	Yes	<b>No</b>
<b>CPP</b>	Yes	<b>No</b>
<b>CSS</b>	Yes	<b>No</b>
<b>CXX</b>	Yes	<b>No</b>
<b>DEF</b>	Yes	<b>No</b>
<b>DIC</b>	Yes	<b>No</b>
<b>DOC</b>	Yes	<b>Yes</b>
<b>DOCM</b>	Yes	<b>Yes</b>
<b>DOCX</b>	Yes	<b>Yes</b>
<b>DOT</b>	Yes	<b>Yes</b>
<b>EML</b>	Yes	<b>Yes</b>
<b>EXCH</b>	No	<b>Yes</b>
<b>H</b>	Yes	<b>No</b>
<b>HHC</b>	Yes	<b>No</b>

<b>HHT</b>	Yes	<b>No</b>
<b>HPP</b>	Yes	<b>No</b>
<b>HTA</b>	Yes	<b>No</b>
<b>HTM</b>	Yes	<b>Yes</b>
<b>HTML</b>	Yes	<b>Yes</b>
<b>HTW</b>	Yes	<b>No</b>
<b>HTX</b>	Yes	<b>No</b>
<b>JHTML</b>	No	<b>Yes</b>
<b>JSP</b>	No	<b>Yes</b>
<b>LNK</b>	Yes	<b>No</b>
<b>MHT</b>	Yes	<b>Yes</b>
<b>MHTML</b>	Yes	<b>Yes</b>
<b>MPX</b>	Yes	<b>No</b>
<b>MSG</b>	Yes	<b>Yes</b>
<b>MSPX</b>	No	<b>Yes</b>
<b>NSF</b>	No	<b>Yes</b>
<b>ODC</b>	Yes	<b>Yes</b>
<b>ONE</b>	No	<b>No</b>
<b>PHP</b>	No	<b>Yes</b>
<b>POT</b>	Yes	<b>No</b>
<b>PPS</b>	Yes	<b>No</b>
<b>PPT</b>	Yes	<b>Yes</b>
<b>PPTM</b>	Yes	<b>Yes</b>
<b>PPTX</b>	Yes	<b>Yes</b>
<b>PUB</b>	Yes	<b>Yes</b>
<b>STM</b>	Yes	<b>No</b>
<b>TIF</b>	Yes	<b>Yes</b>
<b>TIFF</b>	No	<b>Yes</b>
<b>TRF</b>	Yes	<b>No</b>
<b>TXT</b>	Yes	<b>Yes</b>
<b>URL</b>	No	<b>Yes</b>
<b>VDX</b>	No	<b>Yes</b>
<b>VSD</b>	Yes	<b>Yes</b>
<b>VSS</b>	No	<b>Yes</b>
<b>VST</b>	No	<b>Yes</b>
<b>VSX</b>	No	<b>Yes</b>
<b>VTX</b>	No	<b>Yes</b>
<b>XLB</b>	Yes	<b>No</b>
<b>XLC</b>	Yes	<b>No</b>
<b>XLS</b>	Yes	<b>Yes</b>
<b>XLSM</b>	Yes	<b>Yes</b>
<b>XLSX</b>	Yes	<b>Yes</b>
<b>XLT</b>	Yes	<b>No</b>
<b>XML</b>	<b>Yes</b>	<b>Yes</b>

Par défaut, les fichiers pdf ne sont pas supportés. Il faut donc installer le IFilter pdf pour que la recherche fonctionne.

<b>Vendor</b>	<b>Application</b>	<b>Extension</b>
<b>Adobe</b>	Acrobat reader	*.pdf

## 5. Page à supprimer de la recherche

On peut parfois se rendre compte qu'une page apparaît dans la recherche alors qu'elle ne le devrait pas. Il suffit d'entrer son URL grâce à « **Search Result Removal** », et cette dernière sera immédiatement retirée des résultats. De plus, une règle d'analyse sera automatiquement créée afin que l'adresse indiquée ne soit plus indexée.

Shared Services Administration: SharedServices2 > Search Settings > Search Result Removal

### Remove URLs From Search Results

\* Indicates a required field

**URLs To Remove**

Type the URLs you want to remove from search results. These URLs will be removed from search results immediately.

Crawl rules will be created to exclude the specified URLs from future crawls.

To undo this action, delete the crawl rule and start a full crawl of the content source with which this URL was crawled.

URLs to remove: \*

### B. *Scopes Partagés*

Les Scopes créés à partir de la SSP sont partagés avec toutes les applications, contrairement à ceux créés dans le site collection.

Les scopes permettent de réduire le champ de recherche et donc de rendre cette dernière plus pertinente.

Scope Name	Description
People	Search for People
All Sites	Search for everything available for searching.
HR	Search through HR department sites
IT	Search through IT department sites
...	

### 1. Règles

Les scopes sont définies par des règles.

4 types de règles peuvent être mélangés :

- ✓ Adresse Web (http://server/site)
- ✓ Requête sur une propriété (Auteur = Paul Dupond)
- ✓ Source de contenu
- ✓ Tout contenu

Pour chaque type, il faut définir :

- ✓ Pour les adresses web :
  - Répertoire: http://site/subsite/folder
  - Nom d'hôte: nomduserveur
  - Domaine ou sous-domaine : office.microsoft.com
- ✓ Pour les propriétés : Name = Dupond
- ✓ Source de contenu : sélectionner une source de contenu.

Ces définitions peuvent être coordonnées par une inclusion, une exclusion, une obligation

Scope Name	Rules
People	contentclass = urn:content-class:SPSPeople
All Sites	All Content contentclass = urn:content-class:SPSPeople

## 2. Programmation

Par défaut, les scopes sont automatiquement programmés. Mais on peut également lancer le rafraîchissement des scopes manuellement ou spécifier un horaire spécifique.

Scope Name	Schedule
People	Automatic
All Sites	Automatic

### C. *Authoritative pages - Page faisant autorité*

C'est l'un des points les plus importants. Cela permet de classer les pages et sites par importance et pertinence. Le résultat de la recherche affichera en premier les éléments correspondant au classement le plus important.

Site Name	Site URL	Most IMPT	Second Level	Third Level	To demote
Ex : Product	http://MOSS:2007/Products	√			

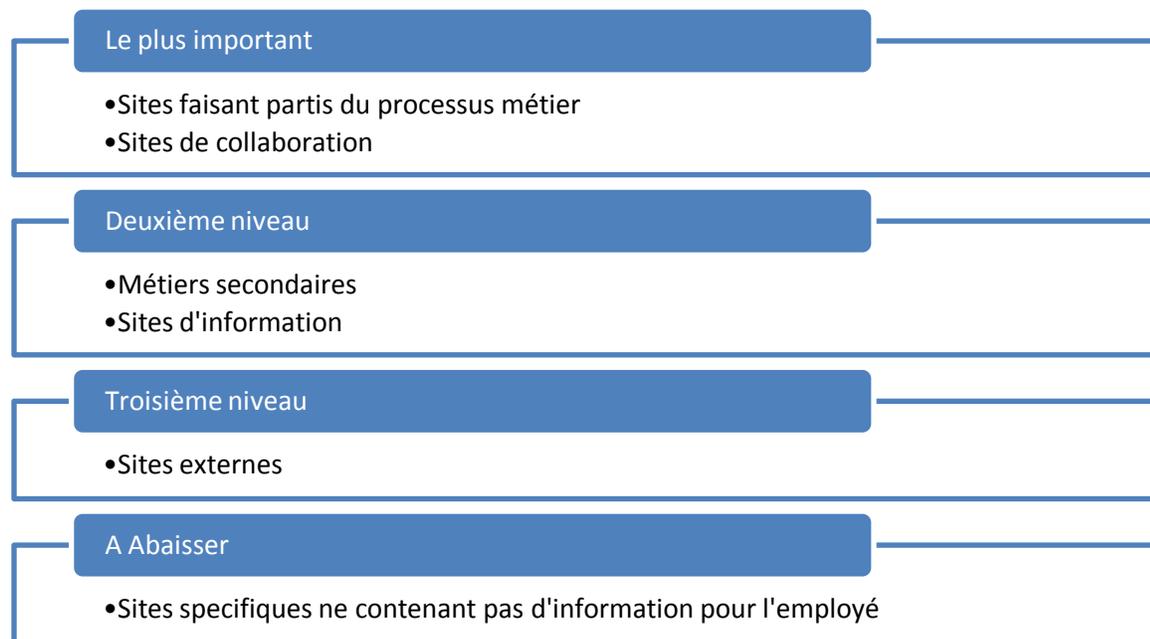
## 1. Bonnes pratiques

Bonne pratique pour classer les pages par importances :

- ✓ Les sites SharePoint correspondant à la chaîne de valeur du métier de l'entreprise seront les plus importantes
- ✓ Les sites encourageant la collaboration, les actions devraient être classées dans une catégorie moins importante.
- ✓ Les sites correspondant au métier secondaire devraient être en deuxième ou troisième catégories.
- ✓ Les sites externes dans la catégorie la plus faible, car l'entreprise ne peut contrôler leur contenu.
- ✓ Il n'est pas obligatoire de classer l'ensemble des sites.

La bonne approche est de sélectionner la pertinence d'un petit nombre de site que l'on juge très important ou très peu important, et par la suite d'ajuster le classement de ces derniers grâce au retour des utilisateurs et des informations dans le log des requêtes ou de l'indexation.

## 2. Proposition de classement



## Specify Authoritative Pages

Use this page to specify authoritative Web pages. Search uses these lists to enhance the overall ranking of results.

### Authoritative Web Pages

Authoritative pages are those that link to the most relevant information. Search uses this list to calculate the rank of every page in the index. There might be many authoritative Web pages in your environment. Type the most valuable in the topmost text box. You also have the option of specifying second and third-level authorities in the lower text boxes.

Enter one full URL per line. Example: `http://intranet/site-directory.html`.

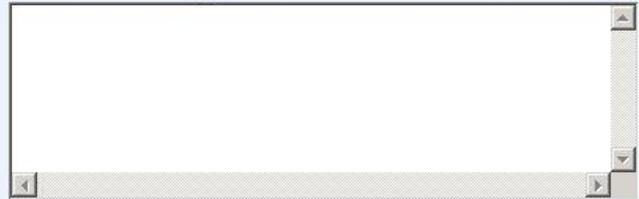
#### Most authoritative pages: \*

A large text area with a scroll bar on the right and a horizontal scroll bar at the bottom. It contains a single line of text that has been redacted with a black box.

#### Second-level authoritative pages:

A large empty text area with a scroll bar on the right and a horizontal scroll bar at the bottom.

#### Third-level authoritative pages:

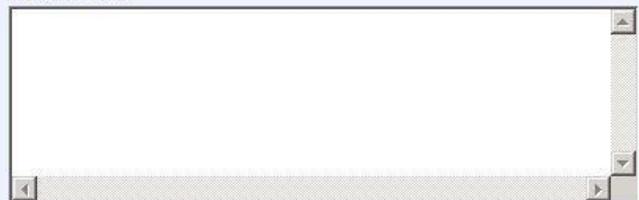
A large empty text area with a scroll bar on the right and a horizontal scroll bar at the bottom.

### Non-authoritative Sites

You can specify that all content from certain sites be ranked lower than that of all other sites. Type the URLs of those sites here, one per line.

URLs typed in this section are treated as prefix matches. Example: entering `http://archive/` will demote the rank of all URLs that begin with `http://archive/`.

#### Sites to demote:

A large empty text area with a scroll bar on the right and a horizontal scroll bar at the bottom.

### Refresh Now?

Ranking re-calculation will start as soon as you click OK. If unchecked, ranking calculations will occur later according to a predetermined schedule.

Refresh now

## V. Paramétrages au niveau Site Collection

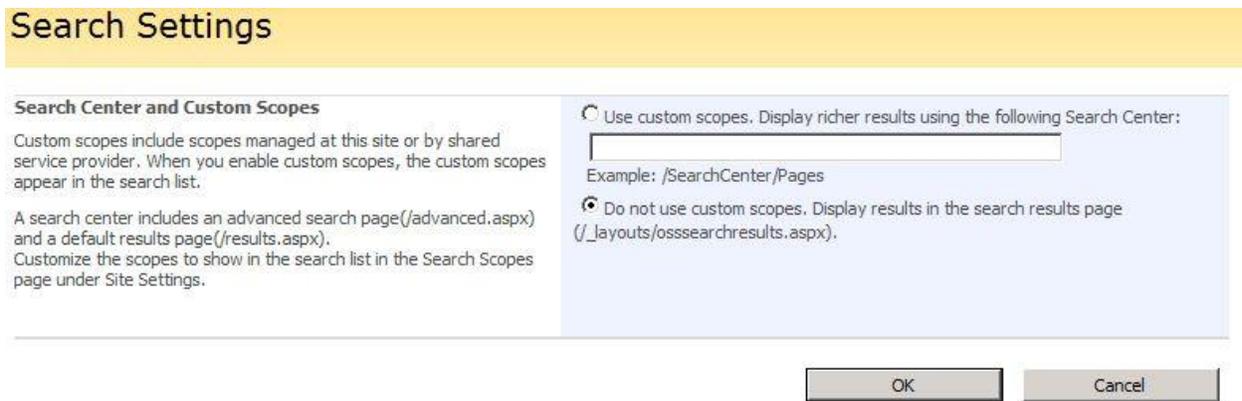
**Note** : Le paramétrage au niveau du Site Collection n'est disponible que dans MOSS (Pas dans WSS).

Dans cette partie, l'administrateur du site collection est responsable des paramétrages qui vont suivre. Ici, le retour des utilisateurs est très important.



### A. Site collection : paramètres de la recherche

Très utile par exemple pour rattacher un site collection isolé, à la recherche globale ou inversement...



### B. Scopes spécifiques

Les scopes créés dans une collection de site ne sont pas partagés. Ils sont dédiés aux utilisateurs des sites d'où l'importance d'avoir une bonne compréhension du besoin quotidien en recherche des utilisateurs.

Scope Name	Description
People	Search for People
All Sites	Search for everything available for searching.

Scope Name	Rules
People	contentclass = urn:content-class:SPSPeople
All Sites	All Content contentclass = urn:content-class:SPSPeople

New Scope                     New Display Group                     Refresh   <span style="float: right;">1 - 3</span>			
Title	Update Status	Shared	Items
<b>Display Group: Search Dropdown (2)</b>			
All Sites	Ready	<input checked="" type="checkbox"/>	89
People	Ready	<input checked="" type="checkbox"/>	0
<b>Display Group: Advanced Search (1)</b>			
All Sites	Ready	<input checked="" type="checkbox"/>	89
<b>Unused Scopes (0)</b>			

### C. Mots clef

Pour chaque mot clef, il faut définir les informations suivantes :

- ✓ Définition
- ✓ Un (ou plusieurs) synonyme
- ✓ Un (ou plusieurs) Best Bets: ce sont les url des sites considérés comme les plus significatifs par les utilisateurs au regard du mot clef.

Keyword	Definition	Synonyms	Best Bets
Ex : MOSS	Microsoft Office SharePoint Services	SharePoint, WSS	<a href="http://MOSS:2007/about_MOSS">Http://MOSS:2007/about_MOSS</a>



## VI. Pour aller plus loin

### A. *Les mots parasites & Thesaurus*

Les mots parasites, listés dans des fichiers spécifiques par langue, sont éliminés de l'indexation. Comme une sorte de dictionnaire, chaque mot sera considéré comme négligeable et ne sera pas pris en compte dans la recherche.

Le thesaurus (dictionnaire hiérarchisé) sert, sous MOSS, à traduire un mot par un autre. Par exemple, un fauteuil peut être converti en siège. Parce que conceptuellement c'est la même chose, même utilité. Le thesaurus est donc utile pour rassembler des éléments avec la même portée sémantique. Il est également utilisé pour corriger certaines fautes de frappe dans la recherche. A travailler, si beaucoup de résultat de recherche ne ramène aucun résultat.

Vous trouverez ces fichiers dans : \Program Files\Microsoft Office Servers\12.0\Data\Office Server\Applications\ « GUID »\Config.

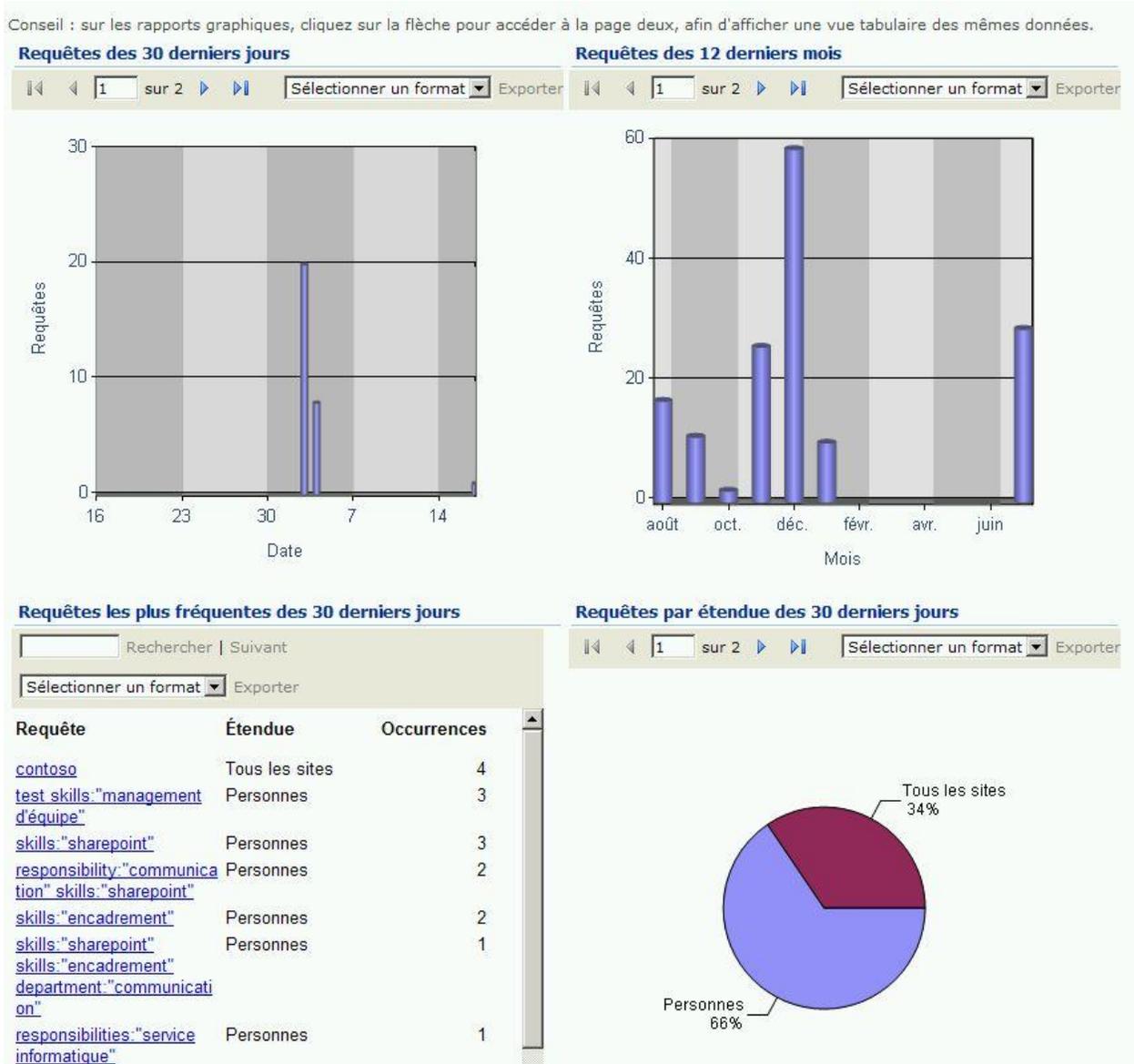
- ✓ Noiseeng.txt & Tseng.xml
- ✓ Noiseenu.txt & Tsenu.xml
- ✓ Noisefra.txt & Tsfra.xml

Pour le fichier de bruit il suffit d'éditer des d'ajouter les mots que l'on ne souhaite pas indexer. Le thesaurus est plus complexe, puisque c'est un fichier xml, exemple :

```
<thesaurus xmlns="x-schema:tsSchema.xml">
  <diacritics_sensitive>0</diacritics_sensitive>
  <expansion>
    <sub>maison</sub>
    <sub>appartement</sub>
    <sub>mas</sub>
    <sub>chalet</sub>
  </expansion>
  <replacement>
    <pat>miss</pat>
    <pat>mos</pat>
    <pat>moss</pat>
    <sub>Microsoft Office SharePoint Server</sub>
  </replacement>
</thesaurus>
```

### B. *Outils de reporting*

Un outil de reporting est présent au niveau de la SSP et au niveau du site collection. Il permet aux différents responsables de surveiller l'utilisation du moteur de recherche, les mots recherchés, et donc de pouvoir réagir et améliorer le système.



### C. Création de page de recherche spécifique

L'amélioration de la recherche passe également par la création ou modification des pages de résultats. Elles doivent être adaptées à l'utilisateur et à leur métier.

### ***D. Intégration de la recherche dans Internet Explorer et Office***

Toujours dans l'objectif de centraliser l'information et la recherche, il est possible d'intégrer la recherche MOSS dans Internet Explorer mais aussi dans la suite Office.

Ainsi, la plateforme de recherche devient incontournable. Quoi de plus naturel que de rechercher sur le portail d'entreprise des documents de référence, directement depuis Word, alors que l'on rédige un nouveau document ?

Voir blog : <http://blogs.codes-sources.com/pioute/archive/2007/10/11/moss-search-ie7-et-office-2007.aspx>

### ***E. Enregistrement d'une recherche et alerte***

Une autre fonctionnalité intéressante est l'alerte. Lorsque l'on effectue une recherche, l'alerte permet de prévenir l'utilisateur que le résultat vient de se modifier (Nouvel entrée dans les résultats, ou modification). Cela peut être intéressant pour un responsable produit ; il doit suivre la diffusion d'information de toute sorte, autour de son produit X, par exemple...

Ce genre d'astuce contribue fortement à l'adoption du moteur de recherche, pour autant que les résultats soient pertinents.

### ***F. Liens utiles & complets***

Fabrice Romelard : **Nouveauté !!** → <http://www.technos-sources.com/tutorial-nouveautes-dans-interfaces-recherche-moss-apres-mise-jour-93.aspx>

Microsoft :

- Capacity for search : <http://technet.microsoft.com/en-us/library/cc262574.aspx>
- Des infos utiles : <http://technet.microsoft.com/en-us/library/cc262043.aspx>
- Planning architecture : [http://technet.microsoft.com/en-us/library/cc261834\(TechNet.10\).aspx](http://technet.microsoft.com/en-us/library/cc261834(TechNet.10).aspx)

Un Ami très productif ... Il prépare une bible SharePoint© Fabrice Romelard :

- <http://blogs.developpeur.org/fabrice69/archive/2008/07/16/moss-search-quelle-fr-quence-et-type-d-indexation-pour-les-sources-de-contenu-sharepoint.aspx>
- <http://www.asp-php.net/tutorial/asp.net/sharepoint-set-search-center.php>
- <http://www.asp-php.net/tutorial/asp.net/sharepoint-search-moss.php>
- <http://www.asp-php.net/tutorial/asp.net/sharepoint-opensearch.php>

- <http://www.technos-sources.com/tutorial-installer-microsoft-office-sharepoint-server-moss-partie-70.aspx>

Un Grand Sage sous MOSS: Christian Hougardy → <http://www.asp-php.net/tutorial/asp.net/sharepoint-gouvernance-partie-1.php>

Précisions & Détails: <http://stephaneey.developpez.com/tutoriel/sharepoint/recherchemoss/>