



AFS

Mode d'emploi du Back-office version 2.5

Manuel back-office AFS 2.5

Version BO :	2.5
Version de documentation :	2.5-05
Date :	03/12/2007
Référence :	AFS/DOC/BO-USERSGUIDE

Table des matières

1.Introduction.....	4
1.1.Connexion au Back-Office.....	4
1.2.Applications et Menus.....	4
1.2.1.Interfaces	5
1.2.2.Services et sites.....	6
1.2.3.Utilisateurs.....	7
2.Home : Page d'accueil.....	8
3.Statistiques.....	9
4.Hit-parade.....	10
4.1.Onglet Hit-Parades	10
4.1.1.Les mots les plus recherchés.....	11
4.1.2.Les documents les plus cliqués.....	11
4.1.3.Les requêtes sans réponse.....	13
4.1.4.Les requêtes lentes.....	14
4.1.5.Les clics par agent.....	14
4.2.Les Mots.....	15
4.2.1.Les mots vides.....	15
4.2.2.Les synonymes.....	16
4.2.3.Les Mots Composés.....	16
4.3.Les dernières requêtes.....	17
4.4.Préférences.....	18
5.Kword.....	19
5.1.Introduction.....	19
5.2.Le mode de programmation des KWord.....	19
5.2.1.Éditer une campagne.....	19
5.2.2.Saisie des caractéristiques de la campagne.....	20
5.2.3.Suivi des campagnes.....	21
6.SEMA.....	22
6.1.Ajout d'un document Sema.....	22
6.2.Définition de la source.....	23
6.3.Déclaration des expressions ou mots à exclure.....	24
6.4.Publication.....	25
6.4.1.Options.....	25
6.4.2.Publication manuelle.....	26
6.4.3.Publication automatique.....	26
7.Configuration.....	27
7.1.Référentiel projet.....	27
7.2.Paramétrage Global.....	27
7.2.1.Paramètres communs.....	27
7.2.2.Définition de catégories de tri.....	28
7.3.Sources de données.....	29
7.3.1.Présentation.....	30
7.3.2.Processus de configuration d'un périmètre.....	30
7.3.3.Créer/modifier la configuration d'un périmètre	31
7.3.4.Ajouter/modifier un site du périmètre	34
7.3.5.Soumettre un périmètre	40

1. Introduction

1.1. Connexion au Back-Office

Le back-office est accessible depuis l'adresse : <https://admin.antidot.net> pour les services hébergés.

Si vous disposez d'AFS en mode licence installé sur vos serveurs, l'URL de connexion doit vous être communiquée par la personne qui a réalisée l'installation d'AFS.

Prérequis : Afin de pouvoir utiliser le back-office (BO), il est impératif que votre navigateur Internet supporte le Javascript et accepte les cookies. Le BO Antidot supporte les navigateurs Firefox à partir de la version 2.0 et Internet Explorer à partir de la version 6.0.

antidot

back office

LE SITE CLIENTS ET PARTENAIRES D'ANTIDOT VOUS OFFRE L'ACCÈS AUX FONCTIONS D'INSTALLATION ET DE CONFIGURATION D'AFS SUR VOTRE SITE.

VOUS ETES WEB AGENCY OU SSII, VOUS DESIREZ PROPOSER NOS SOLUTIONS A VOS CLIENTS... INSCRIVEZ-VOUS !

VOUS ETES WEBMESTRE, VOUS SOUHAITEZ EXPLOITER AFS@WEB BRONZE SUR VOTRE SITE... INSCRIVEZ-VOUS !

Accès clients et partenaires

Utilisateur :

Mot de passe :

Mot de passe oublié ? [cliquez-ici.](#)

v2.5 © Tous droits réservés Antidot 2005-2007

Pour se connecter, il suffit de saisir son nom d'utilisateur (identifiant fournit par Antidot ou par l'administrateur BO ; il s'agit habituellement de votre adresse de courrier électronique) et son mot de passe.

Attention : les identifiants utilisateur et les mots de passe sont sensibles à la casse (majuscules / minuscules)

1.2. Applications et Menus

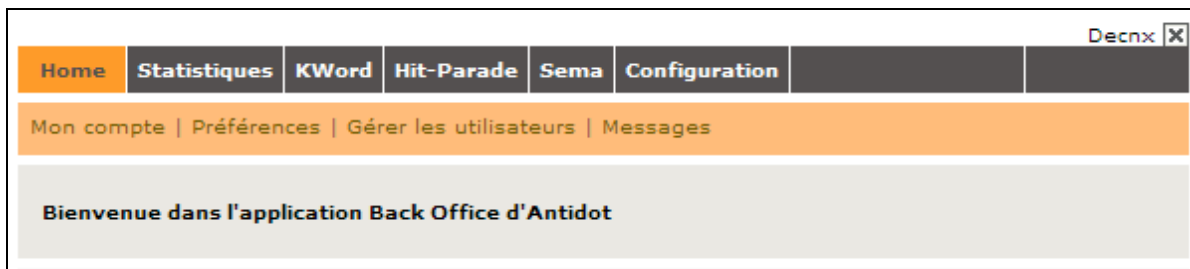
Une fois l'utilisateur authentifié, la page d'accueil permet d'accéder aux différentes applications, suivant ses droits affectés à l'utilisateur, les onglets du menu principal sont activés ou désactivés.

Le back office (BO) propose différentes applications qui permettent :

- **Home** : de gérer les informations administratives liées à votre compte sur le BO ;
- **Statistiques** : de suivre l'évolutions de différents indicateurs quantitatifs ;
- **Hit-parade** : d'analyser les recherches faites par les utilisateurs (analyse qualitative) ;
- **Kword** : de gérer les campagnes de communication ;
- **Sema** : de générer des index de recherche ;
- **Configuration** : de définir certains paramètres de configuration de votre service de recherche, comme par exemples les sites web à indexer dans le cas d'indexation HTML le périmètre de crawl (menu).

Ces applications sont accessibles à travers le bandeau noir situé en haut de chaque page.

Chaque application dispose d'un sous-menu qui lui est propre et qui se situe dans le bandeau orange.






1.2.1. Interfaces




Les différentes applications du BO disposent d'une ergonomie homogène avec le menu de l'application dans le bandeau orange et un coeur de page contenant des informations souvent affichées sous forme de tableau. On y retrouve :

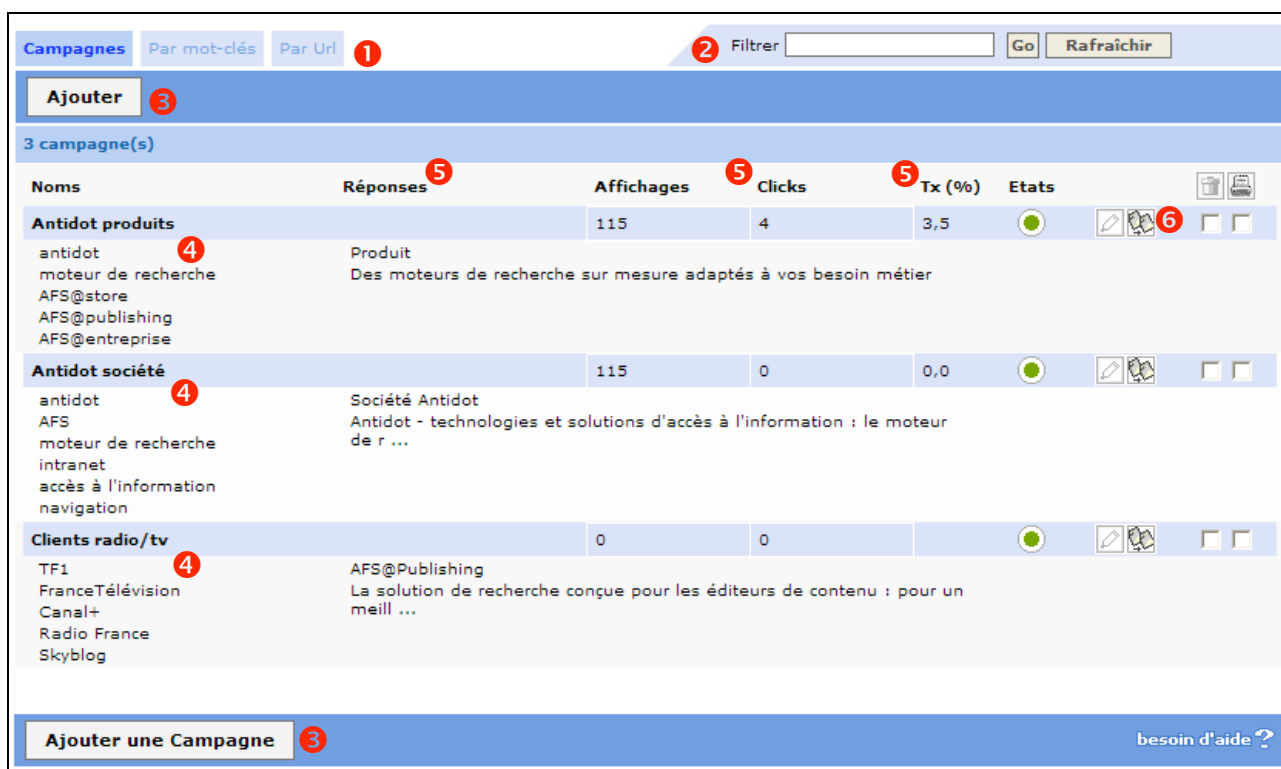
1. des onglets permettant d'accéder à différentes vues ou parties de l'application
2. une zone de filtrage permettant de sélectionner et retrouver plus facilement l'information recherchée. Attention : n'oubliez pas d'effacer le filtre et de cliquer sur «go» pour retrouver l'ensemble des données.
3. deux bandeaux (un en haut et un en bas) contenant des boutons pour les différentes actions possibles (créer des éléments, ...)
4. les différentes « lignes » du tableau
5. le haut de chaque colonne comporte un libellé sur lequel il peut être possible de cliquer pour modifier l'ordre d'affichage du tableau selon les données de cette colonne. Dans le cas où la tableau contient trop de lignes pour tenir sur une seule page, le bandeau haut du tableau affiche le nombre de pages totales, la page courante, et vous pouvez vous déplacer de page en page '<' (page précédente) et '>' (page suivante). Vous pouvez également aller directement à la première page '<<' ou à la dernière page '>>'.

Position	Mot	Nb occurrences	Public	Requête
1	NORME	2942	Indéfini	Voir les requêtes




6. des boutons qui permettent de réaliser pour une ligne donnée les actions suivantes :

Bouton	Action
	Editer
	Ouvrir
	Supprimer

Bouton	Action
	Imprimer
	Copier
	Enregistrer



The screenshot shows the 'Campagnes' section of the AFS back-office. At the top, there are tabs for 'Par mot-clés' and 'Par Url', a search filter with a 'Go' button, and a 'Rafraîchir' button. Below this is an 'Ajouter' button. The main area displays a table with 3 campaigns. The table has columns for 'Noms', 'Réponses', 'Affichages', 'Clicks', 'Tx (%)', and 'Etats'. Each row represents a campaign and includes a list of associated services/sites, a status indicator (green dot), and a set of action buttons (edit, open, delete, print, save). Red circles with numbers 1 through 6 highlight specific UI elements: 1 (filter tabs), 2 (filter input), 3 (Ajouter button), 4 (campaign names), 5 (statistic columns), and 6 (action buttons).

Noms	Réponses	Affichages	Clicks	Tx (%)	Etats
Antidot produits antidot moteur de recherche AFS@store AFS@publishing AFS@entreprise	Produit Des moteurs de recherche sur mesure adaptés à vos besoin métier	115	4	3,5	
Antidot société antidot AFS moteur de recherche intranet accès à l'information navigation	Société Antidot Antidot - technologies et solutions d'accès à l'information : le moteur de r ...	115	0	0,0	
Clients radio/tv TF1 FranceTélévision Canal+ Radio France Skyblog	AFS@Publishing La solution de recherche conçue pour les éditeurs de contenu : pour un meill ...	0	0		

At the bottom, there is an 'Ajouter une Campagne' button and a 'besoin d'aide ?' link.

1.2.2. Services et sites

Les applications fournies sont conçues pour fonctionner à plusieurs niveaux :

- Au niveau du **service**. Un service de recherche est caractérisé par un ensemble de sources de données, les réglages d'indexation propre à chacune de ces sources, et le paramétrage de la recherche sur ces sources.
- Au niveau d'un **site**. Un site est une vue d'un service. Ainsi, un même service de recherche peut être exploité différemment (par plusieurs applications ou plusieurs sites web) en faisant varier des critères de filtrage et les feuilles de style pour l'affichage.
- Sur un **groupe de sites**. Un groupe de sites permet de regrouper plusieurs sites dans un même ensemble afin de les gérer de façon uniforme (cumul de statistiques ...).

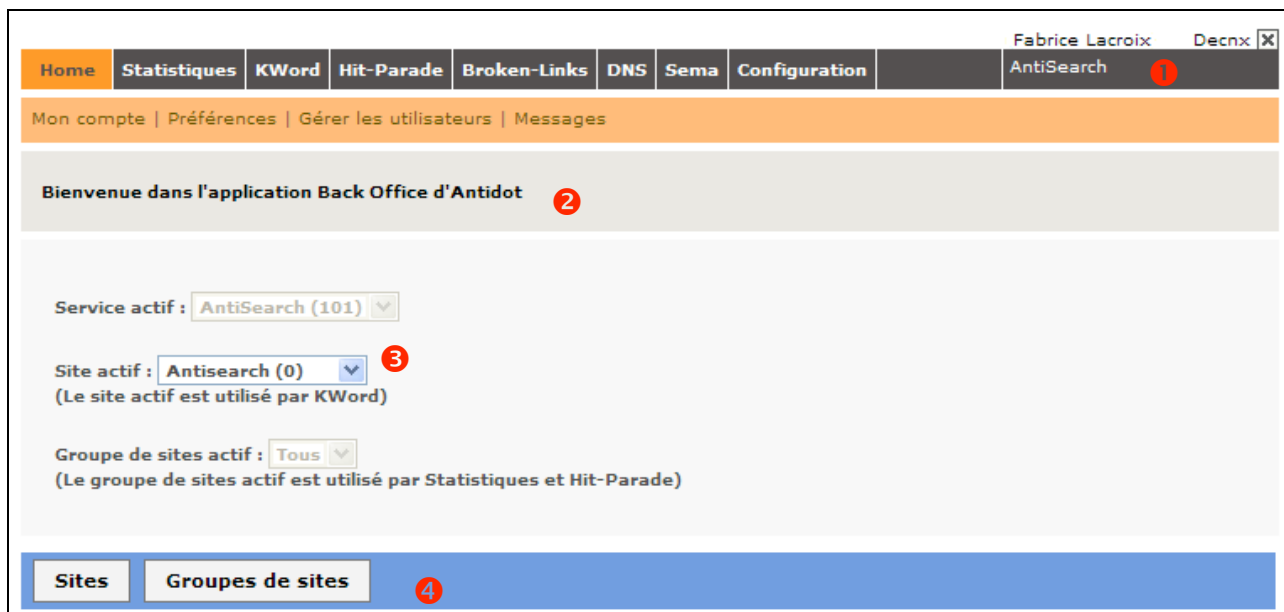
1.2.3. Utilisateurs

Il existe plusieurs types d'utilisateurs, selon son type un utilisateur est associé à un profil de droits.

- Administrateur : Peut créer des sites et groupes de sites, créer des utilisateurs et utiliser les applications.
- Utilisateur et utilisateur avancé : peuvent utiliser les applications avec des niveaux de droit différents (lecture, écriture, ...).

En plus des droits par défaut défini par son type, un utilisateur pourra se voir associer des droits spécifiques par applications sur les services, sites et groupes de sites.

2. Home : Page d'accueil



La page d'accueil du back-office contient les éléments suivants :

1. Le nom d'utilisateur sous lequel vous êtes connecté ainsi qu'une petite croix sur laquelle vous pouvez cliquer pour clore votre session de travail.
2. Une zone dans laquelle sont affichés d'éventuels messages de services.
3. La zone centrale vous permet de sélectionner le service de recherche sur lequel vous désirez travailler, ainsi que pour un service de recherche donné, éventuellement le site ou groupe de site qui vous intéresse.
4. Les boutons « Sites » et « Groupes de sites » ne sont disponibles que si vous êtes connecté avec un profil « Administrateur ». Ils vous permettent de créer un site ou un groupe de sites.

Le sous-menu (bandeau orange) de cette page permet d'accéder aux informations suivantes :

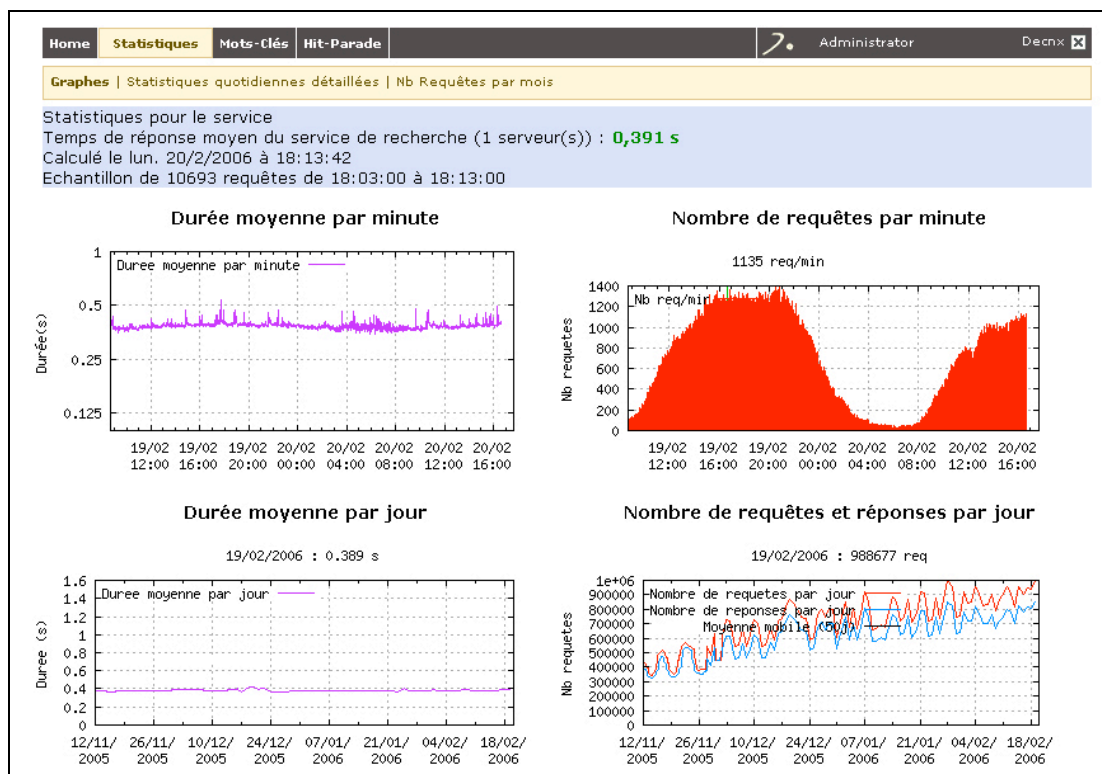
- Le sous menu **Mon compte** permet de déclarer/modifier vos nom, prénom et adresse et principalement de modifier les caractéristiques de votre connexion au BO : adresse électronique et mot de passe.
- Le sous menu **Préférences** permet de définir les paramètres commun à toute l'application back-office qui vous conviennent. En particulier, vous pouvez choisir le nombre de ligne à afficher par page du back-office (en particulier dans les tableaux des outils comme le hit-parade). La valeur par défaut est 10. Pour un plus grand confort d'utilisation, si votre écran le permet, nous vous conseillons d'afficher au moins 25 lignes par page.
- Le sous menu **Gérer les utilisateurs** (disponible seulement à l'Administrateur) permet d'ajouter ou supprimer un utilisateur et de gérer ses droits d'accès sur le Back Office.
- Le sous menu **Messages** permet d'accéder à l'historique des messages de service publiés par l'administrateur de la solution (opérations de maintenance, informations diverses).

3. Statistiques

Cette application regroupe les indicateurs statistiques tels que :

- le temps de réponse moyen du moteur dans les dernières minutes, heures et jours
- le nombre de requête par minute, ces données cumulées par jour ou par mois
- la proportion de requêtes sans réponses.

Ces statistiques sont calculées sur les groupes de sites uniquement.



4. Hit-parade

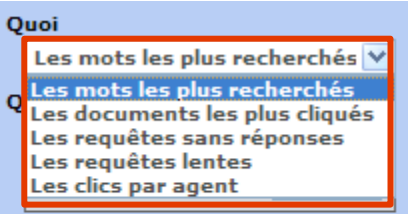

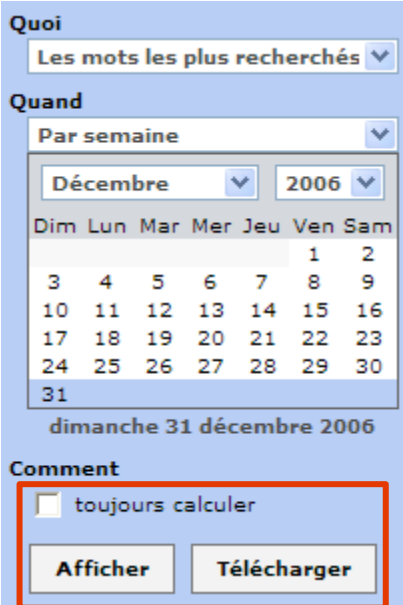
L'application hit-parade permet d'analyser les recherches des visiteurs.

Il permet de dresser le top de mots les plus recherchés, de connaître les requêtes sans réponses, les dernières requêtes ...

Il peut être calculé sur un groupe de sites ou un site.

4.1. Onglet Hit-Parades

Cet onglet vous permet d'accéder aux résultats de l'analyse, la colonne de gauche offrant le moyen de sélectionner :

<p>Quoi : ce que vous désirez visualiser :</p> <ul style="list-style-type: none"> - les mots les plus recherchés - les documents les plus cliqués - les requêtes sans réponses - les requêtes lentes - la répartition des clics par agent (et donc par source de données) 	<p>Quand : la période à analyser (par jour, semaine, mois ou année). Il suffit alors d'utiliser le calendrier qui s'affiche pour sélectionner la période qui vous intéresse.</p>	<p>Comment : cliquez sur «Afficher» pour que le résultat de l'analyse s'affiche à l'écran, et sur «Télécharger» pour l'exporter vers votre ordinateur au format XML ou CSV (Excel). La case à cocher 'toujours calculer' permet de forcer le calcul même si celui-ci a déjà été réalisé.</p>																																																	
 <p>Quoi Les mots les plus recherchés Les documents les plus cliqués Les requêtes sans réponses Les requêtes lentes Les clics par agent</p> <p>dimanche 31 décembre 2006</p> <p>Comment <input type="checkbox"/> toujours calculer</p> <p>Afficher Télécharger</p>	 <p>Quoi Les mots les plus recherchés</p> <p>Quand Par année Par jour Par semaine Par mois Par année</p> <p>Comment <input type="checkbox"/> toujours calculer</p> <p>Afficher Télécharger</p>	 <p>Quoi Les mots les plus recherchés</p> <p>Quand Par semaine Décembre 2006</p> <table border="1"> <thead> <tr> <th>Dim</th> <th>Lun</th> <th>Mar</th> <th>Mer</th> <th>Jeu</th> <th>Ven</th> <th>Sam</th> </tr> </thead> <tbody> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td>1</td> <td>2</td> </tr> <tr> <td>3</td> <td>4</td> <td>5</td> <td>6</td> <td>7</td> <td>8</td> <td>9</td> </tr> <tr> <td>10</td> <td>11</td> <td>12</td> <td>13</td> <td>14</td> <td>15</td> <td>16</td> </tr> <tr> <td>17</td> <td>18</td> <td>19</td> <td>20</td> <td>21</td> <td>22</td> <td>23</td> </tr> <tr> <td>24</td> <td>25</td> <td>26</td> <td>27</td> <td>28</td> <td>29</td> <td>30</td> </tr> <tr> <td>31</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p>dimanche 31 décembre 2006</p> <p>Comment <input checked="" type="checkbox"/> toujours calculer</p> <p>Afficher Télécharger</p>	Dim	Lun	Mar	Mer	Jeu	Ven	Sam						1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31						
Dim	Lun	Mar	Mer	Jeu	Ven	Sam																																													
					1	2																																													
3	4	5	6	7	8	9																																													
10	11	12	13	14	15	16																																													
17	18	19	20	21	22	23																																													
24	25	26	27	28	29	30																																													
31																																																			

4.1.1. Les mots les plus recherchés

Le Hit-Parade permet d'obtenir le top des mots clés les plus recherchés : par jour, par semaine, par mois ou par an.

Position	Mot	Nb occurrences	Public	Requête
1	NORME	2942	Indéfini	Voir les requêtes
2	CHAUSSURES	994	Indéfini	Voir les requêtes
3	EMPLOI	949	Indéfini	Voir les requêtes
4	CHAUSSURE	883	Indéfini	Voir les requêtes
5	OFFRE	709	Indéfini	Voir les requêtes
6	FEMME	644	Indéfini	Voir les requêtes
7	CERTIFICATION	594	Indéfini	Voir les requêtes
8	HOMME	472	Indéfini	Voir les requêtes
9	EAU	455	Indéfini	Voir les requêtes
10	FRANCE	386	Indéfini	Voir les requêtes

Pour obtenir le classement, il faut sélectionner la période souhaitée dans la liste ainsi que le mois et l'année souhaitée au niveau du calendrier (1), puis cliquer sur le bouton «**Afficher**».

Si l'analyse pour cette période n'a pas déjà été réalisée par AFS, ou si vous cochez la case « toujours calculer », le calcul s'exécute avant affichage.

La liste des mots clés apparaît dans l'ordre de fréquence décroissante. Le mot clé le plus recherché apparaît en premier (2).

Pour chaque mot clé, il est possible de connaître les requêtes exactes dans lesquelles il apparaît en cliquant sur le lien « **Voir les requêtes** »

NB : Le top des mots correspond au top des mots contenus dans les requêtes et non pas aux requêtes complètes.

4.1.2. Les documents les plus cliqués

Cette fonctionnalité permet de connaître la liste des documents sur lesquels les utilisateurs ont cliqué.

Remarque : pour être disponible, ce décompte doit être préalablement configuré lors de la mise en oeuvre du service (utilisation du script Redirect).

Quoi
Les documents les plus cliqués

Quand
Par semaine

Novembre 2007

				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	

mercredi 21 novembre 2007

Comment
 toujours calculer

Afficher **Télécharger**

Hit-Parade des sites

Filtrer **go**

70 entrée(s), page 5 sur 7

Position	Site	Nb occurrences	Public
41	outils.abondance.com/antidot.html	1	Indéfini
42	www.ensmp.fr/%7Escherer/adminet/cawa/cawa0202.html	1	Indéfini
43	www.copernic.com/fr/	1	Indéfini
44	www.glisse66.com/meteo.php	1	Indéfini
45	tv.edicom.ch/	1	Indéfini
46	www.snowtignestravel.com/ete/reservation.html	1	Indéfini
47	leswebs.com/liste_de_liens1.htm	1	Indéfini
48	dogstory.net/british_airways.htm	1	Indéfini
49	fnac-micro.calindex.com/micro-bigben-ecran-5.4-extra-plat-pour-playstation-bigben-interactive-1452763.html	1	Indéfini
50	www.telesuisse.ch/	1	Indéfini

La liste des documents apparaît par fréquence décroissante, il est possible de définir des règles pour le back office lors de l'analyse considère plusieurs URLs comme un même document. Pour cela, cliquer sur le lien du document afin d'accéder à la page qui permet d'éditer les paramètres de l'analyse pour une URL/un site donné. Cette fonctionnalité permet alors de considérer un site et tous ses sous-répertoires comme un même document.

Retour

Gestion d'un site

Host

Port

Path (Racine du site)

Nombre de sous-répertoire formant le site à partir de la racine

Description

Type du site Normal Interdit

Site réservé à un public Tout public Adulte

Ajouter le site

4.1.3. Les requêtes sans réponse

Les requêtes sans réponses permettent d'obtenir la liste des requêtes n'ayant remonté aucun résultat : par jour, par semaine, par mois ou par an.

Quoi
Les requêtes sans réponses

Quand
Par mois
Novembre 2007
mercredi 21 novembre 2007

Comment
 toujours calculer

Afficher Télécharger

Hit-Parade des requêtes sans réponses

Filtrer go

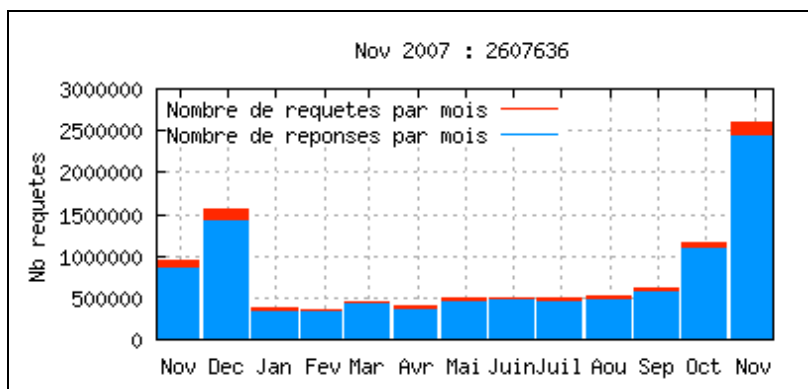
943 entrée(s), page 1 sur 95 << < > >>

Position	Requête	Nb occurrences
1	haute-vienne	44
2	entreprise haute-vienne	40
3	maire haute-vienne	36
4	emploi haute-vienne	34
5	departement haute-vienne	34
6	musee haute-vienne	33
7	decouvrir haute-vienne	32
8	tourisme haute-vienne	29
9	hybridhaus	26
10	sport haute-vienne	22

La liste des requêtes n'ayant pas ramené de réponses apparaît dans l'ordre de fréquence décroissante. La requête sans réponse la plus fréquemment jouée apparaît en premier.

L'analyse de ces requêtes sans réponses permet de mesurer l'écart entre les informations disponibles, et les attentes des utilisateurs. L'analyse de ces recherches permet également la cas échéant d'enrichir dictionnaires des synonymes ou thésaurus utilisés lors de l'indexation.

La proportion globale de requêtes sans réponse est visible à travers l'outil « Statistiques ». Il s'agit de la partie rouge de chaque barre dans le graphe ci-dessous qui matérialise pour chaque mois ces requêtes.



Cette analyse est particulièrement importante dans des domaines comme les sites eCommerce car elle donne des indications sur les produits recherchés par les consommateurs.

4.1.4. Les requêtes lentes

Affiche les requêtes lentes permet de cibler les requêtes qui ont nécessité un temps de traitement important. Cela permet de suivre la qualité perçue du service et une éventuelle surcharge des serveurs.

Le délai minimum à partir duquel une requête lente est considérée comme lente se paramètre dans le sous menu « Préférences » de l'application 'Hit Parades' (cf §4.4).

Quoi

Les requêtes lentes

Quand

Août 2007

Dim	Lun	Mar	Mer	Jeu	Ven	Sam
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

mardi 21 août 2007

Requêtes lentes

9 entrée(s)

Heure	Requête	Durée
10:51:54	charles de gaulle	5.159
21:06:18	cours	4.936
21:06:23	cours	4.879
05:25:12	france	4.774
10:52:01	service de l'eau	4.586
20:25:36	dire d'expert	3.629
21:04:13	droit	3.465
20:18:20	sport	3.185
21:04:09	droit	3.174

4.1.5. Les clics par agent

Si un service de indexe et permet de rechercher sur plusieurs sources de données, alors il est constitué de plusieurs agents répondeurs. L'analyse des « clics par agent » permet de connaître pour chaque agent le nombre de clics reçus, c'est à dire les sources de données fournissant les informations les plus adéquates aux questions des utilisateurs.

Quoi

Les clics par agent

Quand

Novembre 2007

Dim	Lun	Mar	Mer	Jeu	Ven	Sam
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	

mardi 6 novembre 2007

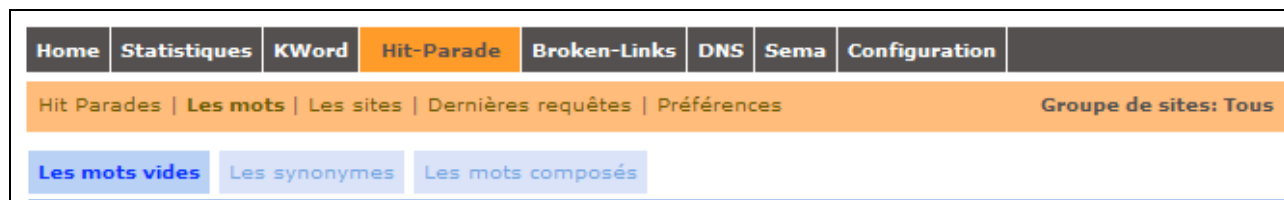
Répartition par agents

4 entrée(s)

Agent	Nombre de clics
DNS	2
DIRECTORY	10
ENGINE1	44
ENGINE2	77

4.2. Les Mots

Ce sous-menu permet de paramétrer le calcul du Hit-Parade afin d'en faciliter la lecture et de l'adapter à l'environnement métier en déclarant : des mots vides, des mots composés ou encore des synonymes. Ainsi, lors du prochain calcul du hit parade ces paramètres seront pris en compte.



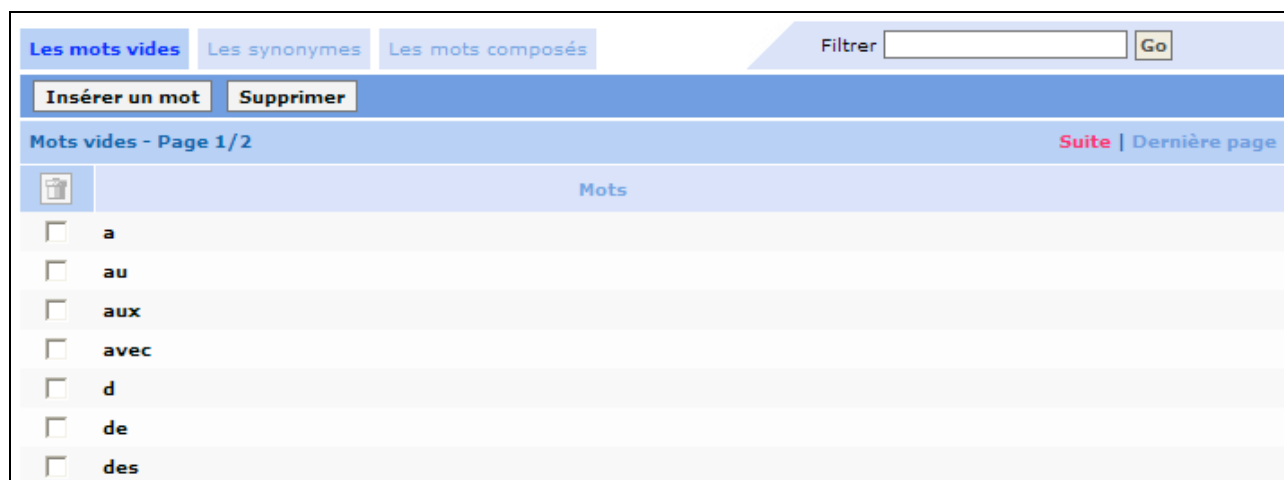
Attention : Les réglages enregistrés ici n'interviennent que dans le calcul du Hit Parade, et n'ont aucun effet sur le moteur de recherche. De plus, ces modifications ne seront prises en compte que lors des prochains calculs. Si un calcul a déjà eu lieu, il faudra cocher la case " toujours calculer " pour que les modifications soient prises en compte.

4.2.1. Les mots vides

Afin de faciliter la lecture du hit-parade, il est nécessaire de traiter la liste des mots et d'identifier les mots dits "**vides**" (c'est-à-dire pour "vides de sens") et n'apportant pas d'information significative dans les requêtes.

Les déterminants (le, la les, de, du, ...) et autres pronoms sont extrêmement fréquents dans les requêtes et pour autant sont inintéressants dans l'analyse du hit-parade, c'est pourquoi tous les pronoms et déterminants sont d'emblée inscrits dans la liste des mots vides.

Pour un métier particulier, certains mots pourtant significatifs peuvent être considérés comme vides, il faudra alors les identifier dans le hit-parade et les insérer dans la liste.



4.2.2. Les synonymes

La définition de synonymes permet de considérer une forme et ses formes alternatives ou ses formes fléchies comme une forme unique permettant ainsi de regrouper les occurrences de ces formes lors de l'analyse des requêtes.

Les mots vides	Les synonymes	Les mots composés	Filter <input type="text"/>	Go
<input type="button" value="Insérer un mot"/> <input type="button" value="Supprimer"/>				
Aucun enregistrement				
	Mots	Synonymes		
<input type="checkbox"/>	1ere compagnie	compagnie		
<input type="checkbox"/>		première compagnie		
<input type="checkbox"/>		1er compagnie		
<input type="checkbox"/>	journal televise	13h		
<input type="checkbox"/>		13 heures		
<input type="checkbox"/>		journal tv		
<input type="checkbox"/>		20h		
<input type="checkbox"/>		journal du soir		

4.2.3. Les Mots Composés

De même pour rendre la liste de mots plus lisible, il est possible de recomposer les groupes nominaux les plus fréquents.

Ces groupes nominaux sont identifiables lorsque l'on clique sur le lien "**voir les requêtes**" et peuvent être reconstitués par la fonction "**les mots composés**".

Les mots vides	Les synonymes	Les mots composés	Filter <input type="text"/>	Go
<input type="button" value="Insérer un mot"/> <input type="button" value="Supprimer"/>				
Mots composés				
	Mots	Action	est composé de	
<input type="checkbox"/>	accessoires de tennis	Nouvelle composition	accessoires de tennis	
<input type="checkbox"/>	chassures enfants	Nouvelle composition	chaussures enfants	
<input type="checkbox"/>	chassures enfants	Nouvelle composition	chaussures enfant	
<input type="checkbox"/>	Mot cle ou ref	Nouvelle composition	mot cle ou ref	
<input type="checkbox"/>	pantacourts de tennis	Nouvelle composition	pantacourts de tennis	
<input type="checkbox"/>	sac a dos	Nouvelle composition	sac a dos	
<input type="checkbox"/>	sandales de bodyboard	Nouvelle composition	sandales de bodyboard	

L'outil de calcul du Hit Parade comptabilise chaque requête selon le groupe de mots le plus long. Ainsi pour les groupes de mots : "chaussure enfant" et "chaussure enfant fille", la requête "chaussure enfant fille" sera comptabilisée uniquement dans le groupe de mots "chaussure enfant fille" et non pas dans "chaussure enfant" ni dans les mots "chaussure", "enfant" et "fille".

4.3. Les dernières requêtes

Cette fonction permet de visualiser les dernières requêtes des utilisateurs. Il est ainsi possible de connaître en temps réel les recherches réalisées des utilisateurs. La dernière colonne permet de connaître le numéro de la première réponse renvoyée et ainsi de déterminer sur quelle page est l'utilisateur (1 correspondant à une réponse de la première page, 11 de la deuxième page etc).

Hit Parades Les mots Les sites Dernières requêtes Préférences					Groupe de sites: Tous
Dernières requêtes					
Date	Heure	Temps de réponse	Requête	N° 1ère rép.	
2007-12-06	11:55:09	0.357	rondes	1	
2007-12-06	11:55:09	0.277	accessoires beaute	1	
2007-12-06	11:55:07	0.256	jupe	1	
2007-12-06	11:55:06	0.27	sexualite	1	
2007-12-06	11:55:05	0.253	geoportail	1	
2007-12-06	11:55:05	0.332	les feux de l amour	1	
2007-12-06	11:55:03	0.342	jean	1	
2007-12-06	11:55:03	0.28	ingrid betancourt	1	
2007-12-06	11:55:02	0.275	matilda	1	
2007-12-06	11:55:01	0.255	pomme	1	

4.4. Préférences

Cet onglet permet de fixer les paramètres spécifique à l'affichage de l'application Hit-Parade dans le back-office.

Pour le calcul des requêtes lentes, il est possible de définir le temps de réponse à partir duquel une réponse sera considérée comme lente. Pour l'ensemble des Hit-Parades, il est possible de ne travailler qu'en mode synthétique, c'est à dire de prendre en compte dans le calcul seules les requêtes des premières pages.



The screenshot shows a web interface with a navigation menu at the top containing 'Home', 'Statistiques', 'KWord', 'Hit-Parade', 'Sema', and 'Configuration'. Below the menu is a breadcrumb trail: 'Hit Parades | Les mots | Les sites | Dernières requêtes | Préférences'. The main content area is titled 'Préférences' and contains two settings: 'Requêtes lentes au delà de (en secondes):' with a text input field containing '3.0', and 'Ne travailler que sur les premières pages (Mode synthétique):' with radio buttons for 'Oui' and 'Non', where 'Non' is selected. At the bottom, there are two buttons: 'Valider' and 'Annuler'.

5. Kword

5.1. Introduction

L'outil KWord permet de gérer la promotion de certains articles via le moteur de recherche. Les articles définis dans une promotion se caractérisent alors par une présentation différente qui les distinguent des réponses classiques.

Le parallèle est fait entre ces promotions et la notion de campagnes publicitaires, ces promotions pouvant avoir une durée de vie, et un objectif d'affichage et de mesure d'impact sur les utilisateurs.

5.2. Le mode de programmation des KWord

5.2.1. Éditer une campagne

Une campagne est définie par un nom obligatoire, saisi dans la zone de texte «Libellé de la campagne»

Et potentiellement :

- Une durée de vie caractérisée par une date de début et/ou une date de fin
- Un commentaire libre
- Une zone d'affichage (en marge des autres réponses ou au sein des autres réponses)
- Un ordre d'affichage
- Un nombre d'affichage à atteindre
- Un nombre de clics à atteindre

Une fois saisie la définition de la campagne, cliquer sur **Enregistrer**

Remarque : Pour que les campagnes soient visibles dans le flux de résultats, il faut que le service de recherche ait été configuré en conséquence au préalable.

Edition de la campagne		
Libellé de la campagne * :	<input type="text" value="Test"/>	Du : <input type="text" value="09/08/2007"/> Au : <input type="text" value="10/08/2007"/>
Commentaire :	<input type="text"/>	
Répartition des réponses :	<input type="text" value="à côté (publicité)"/>	
Ordre des réponses :	<input type="text" value="par ordre de priorité"/>	
Affichages :	<input type="text"/>	
Clics :	<input type="text"/>	
Fin de la campagne: la fin de la campagne est atteinte lorsque la fin de la période est atteinte ou le seuil d'affichages ou le seuil de clics.		

5.2.2. Saisie des caractéristiques de la campagne

Une campagne consiste donc en la mise en avant d'un document (accès via une url) dès que l'internaute saisit certains mots clés ❶.

Cette réponse sera alors affichée en marge des autres réponses et contiendra les caractéristiques définis dans le Back Office ❷ .

Mot-Clés

1 mot-clé(s)

Mots-Clés Du / Au

test_toto

❶ Mot-Clé * :

Pendant toute la campagne seulement Du : Au :

Ajouter

Réponses

1 ur(s)

Réponses Du / Au

Test de la campagne

Ceci est un test de campagne publicitaire

http://www.koodza.fr/Rockrider-5-1-h_56_14722.htm

❷ Titre * :

Position :

Description * :

URL * :







Image :

Pendant toute la campagne seulement Du : Au :

Ajouter

5.2.3. Suivi des campagnes




Une fois les campagnes saisies et en ligne, il est intéressant d'en suivre l'efficacité et éventuellement de décider de leur poursuite ou d'une meilleure définition des mots clés ou des réponses.

1 campagne(s)					
Noms	Réponses	Affichages	Clicks	Tx (%)	Etats
Test	du 09/08/2007 au 10/08/2007	0	0		     
test_toto	Test de la campagne Ceci est un test de campagne publicitaire				

[Ajouter une Campagne](#) [besoin d'aide ?](#)

L'accueil de l'application KWord, donne un aperçu de l'ensemble des campagnes saisies dans l'interface.

Les indicateurs sont :

- Noms = nom de la campagne tel que saisi dans la zone de définition de la campagne.
- Réponses = colonne qui va d'une part donner les éventuelles dates de parution de la campagne puis le titre de la campagne tel qu'il va s'afficher sur le site, ainsi qu'un résumé de la description.
- Affichages = affiche le décompte du nombre de fois où la campagne a été affichée.
- Clicks = affiche le décompte du nombre de clics sur la campagne
- Tx(%) = taux de clic de la campagne.
- Etats = traduit l'état de la campagne :
 -  : Campagne valide
 -  : Campagne terminée
 -  : Campagne programmée

Les sous menu "**Par url**" et "**Par mot-clé**" permettent l'affichage des campagnes existantes selon ces critères.

6. SEMA

Le module SEMA est un module optionnel qui est conçu pour l'aide au référencement des contenus dans les moteurs de recherche grand public. Cette application n'apparaît dans le menu du back-office que si vous avez souscrit à l'option Sema.

Sema permet à partir de différentes sources (entrée manuelle, Related Topic Expressions ou Hit parade des requêtes) de générer un index de recherche.

Cet index peut avoir deux objectifs :

- proposer une navigation orientée contenu et basée sur le moteur de recherche et les utilisateurs
- améliorer le référencement du site sur les moteurs de recherche en améliorant la visibilité des pages importantes du site.

6.1. Ajout d'un document Sema

Documents SEMA				
			Nom <input type="text"/>	Filter
2 document(s), page 1 sur 1				
Nom	Créé	Publié	Autopub.	
test	18 juin 2007 21:20:06	Jamais	<input type="checkbox"/>	
TopRecherche	30 avr. 2007 16:22:59	Jamais	<input type="checkbox"/>	

Un document sema représente une publication. Il définit les expressions à publier, leur source, leur format de sortie mais également la façon dont on obtient ces expressions et la façon de les publier.

L'ajout d'un document sema est possible à partir du bouton 'Ajouter', les boutons de droite permettent l'ouverture d'un document, la modification de son nom ou sa suppression.

6.2. Définition de la source

① **Mes sources** | Mes exclusions | Ma liste | Ma publication Fermer le document (TopRecherche)

Hit-Parade (HPR)

liste figée (Manuelle)

② **Ajouter** **Rafraîchir**

Expression Etat Score minimum - 0 + **Go**

500 document(s), page 1 sur 10

③ Expression	Etat	Score	Occurrences	④
sac de sport		100%	4 202	
velo		98%	3 944	
tente		91%	2 887	
vtt		81%	1 843	
tente famille		79%	1 646	
velos route		71%	1 114	
short		65%	819	
sac a dos		65%	818	
sac		63%	770	
maillot de bain		61%	687	
chaussures		60%	648	

Onglet : ①

Il est possible d'ajouter une source pour la génération de la liste ②

La liste des mots clés ③ apparaît par défaut dans l'ordre de fréquence décroissante. Cet ordre de tri peut être modifié par score ou alphabétique en cliquant sur la colonne désirée.

Le score est calculé en fonction de la fréquence de la requête par rapport à la requête la plus populaire.

Les boutons de droite ④ permettent de valider ou d'exclure une expression ou un mot.

Les différents états possibles sont les suivants :

	Requête exclue
	Mot de la requête exclu
	Requête validée
	Nouvelle requête

6.3. Déclaration des expressions ou mots à exclure

Mes sources | **Mes exclusions** | Ma liste | Ma publication Fermer le document (TopRecherche)

Mots exclus	Gestion des mots exclus		
Expressions exclues	Mot <input type="text"/> <input type="button" value="Go"/>		
11 document(s), page 1 sur 1 << < > >>			
<u>Mot exclu</u>	<u>Créé</u>	<u>Modifié</u>	
32290	30 juil. 2007 14:22:01	30 juil. 2007 14:22:01	
57007	30 juil. 2007 14:21:53	30 juil. 2007 14:21:53	
61778	30 juil. 2007 14:21:44	30 juil. 2007 14:21:44	
66114	30 juil. 2007 14:21:30	30 juil. 2007 14:21:30	
adidas	19 juin 2007 09:56:21	19 juin 2007 09:56:21	
mixte	19 juin 2007 09:51:05	19 juin 2007 09:51:05	
nike	19 juin 2007 09:56:08	19 juin 2007 09:56:08	
polar	19 juin 2007 09:56:13	19 juin 2007 09:56:13	
puma	19 juin 2007 09:56:38	19 juin 2007 09:56:38	
salomon	19 juin 2007 09:56:02	19 juin 2007 09:56:02	
tankini	19 juin 2007 10:01:24	19 juin 2007 10:01:24	
<input type="button" value="Ajouter"/>			

Cet onglet **1** contient la liste des expressions ou mots exclus. Concernant les mots, il est possible de les entrer manuellement, ce qui n'est pas possible pour les expressions où seule l'annulation de leur exclusion est possible.

6.4. Publication

6.4.1. Options

Mes sources | Mes exclusions | Ma liste | **Ma publication**

Options de publication

v Options de filtrage

Score minimum *

Nombre maximum d'expressions *

Nombre minimum de résultat *

v Options des résultats

Activer *

Site *

Nombre de résultats par page *

Nombre de pages *

Nom du fichier

v Options des index

Activer *

Nombre d'expression par page *

Type d'index *

Clé de tri *

Nom du fichier

Il est possible de paramétrer la publication selon les options suivantes :









- **filtrage** : score minimum, nombre de requêtes maximum, nombre de résultats minimum par requête
- **résultats** : permet de générer des pages de résultats statiques
- **index** : génération de la liste des requêtes, types (répertoire : résultats groupés par initiale, page : une page par initiale, global : toutes les expressions à la suite), tri des expressions
- **sortie xml** : format de sortie avec définition de l'extension

- **sortie html : définition** de la feuille de style à appliquer

6.4.2. Publication manuelle

Publication manuelle

Lorsque votre liste d'expression est prête, vous pouvez lancer la publication des pages en cliquant sur le bouton "Publier". Elle peut nécessiter de quelques minutes à près de 45 minutes. A la fin de la publication vous trouverez une archive avec tous les fichiers dans la liste ci-dessous.

Date de publication	Fichier Zip
1 août 2007 16:14:00	 
1 août 2007 15:33:00	 
2 juil. 2007 10:54:00	 
22 juin 2007 00:03:00	 

Publier **Rafraîchir**

Le bouton 'Rafraîchir' permet de lancer une publication immédiate.

La liste des publications est affichée par date descendante.

6.4.3. Publication automatique

Options de publication automatique

Activer *

Dernière publication Jamais

Fréquence de publication *

Destination *

Enregistrer **Annuler**

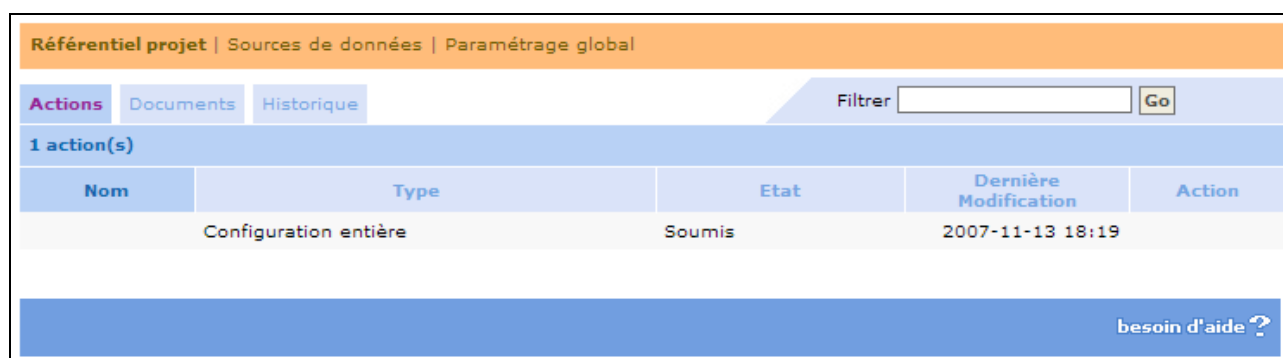
Permet d'activer la publication automatique en définissant la fréquence de publication, la destination du document (FTP ou SFTP).

7. Configuration

Cette application n'est accessible qu'aux utilisateurs ayant la profil d'administrateur car elle permet de configurer le service de recherche : sources de données, traitement linguistique, ...

7.1. Référentiel projet

En plus d'interfaces de configuration, cette application offre un un espace « Référentiel projet » qui permet de suivre la vie de son service de recherche et de permettre un échange de données entre le client et Antidot.



The screenshot shows a web interface for 'Référentiel projet'. At the top, there are navigation links: 'Sources de données' and 'Paramétrage global'. Below this, there are three tabs: 'Actions' (selected), 'Documents', and 'Historique'. To the right of the tabs is a search bar labeled 'Filtrer' with a 'Go' button. Below the tabs, it says '1 action(s)'. A table follows with the following data:

Nom	Type	Etat	Dernière Modification	Action
Configuration entière		Soumis	2007-11-13 18:19	

At the bottom right of the interface, there is a link 'besoin d'aide ?'.

Les différentes fonctions offertes sont disponibles à partir des onglets suivants :

- L'onglet **Actions** permet de suivre les différentes **actions en cours** (indexation, déploiement des agents, ...).
- L'onglet Documents permet l'échange de fichiers (spécifications techniques, fichiers de configuration etc) entre Antidot et le client.
- L'onglet Historique permet d'accéder à la liste les actions déjà réalisées et terminée.

7.2. Paramétrage Global

Le paramétrage global permet de définir :

1. des paramètres communs au service de recherche pour l'indexation : traitement linguistique, ...
2. des catégories permettant de trier les données indexées et d'offrir à l'utilisateur des fonctions de tri et de sélection lors de ses recherches. Certaines valeurs de tri peuvent être extraites automatiquement des données lors de l'indexation alors que d'autres peuvent être définies dans la back-office par cette interface.

7.2.1. Paramètres communs

Seulement quelques uns des paramètres d'indexation sont accessibles par le back-office :

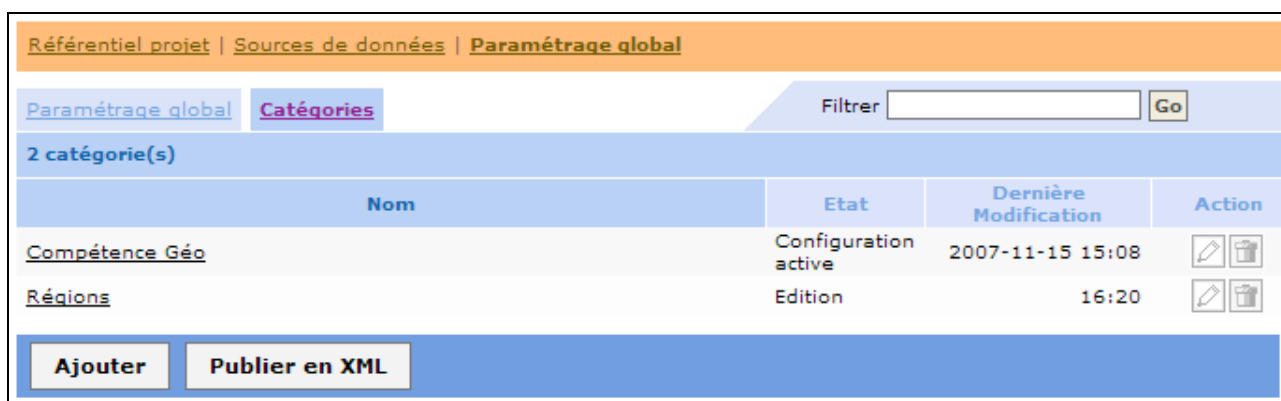
- le traitement principal à appliquer lors de l'indexation

- l'autorisation des requêtes vides à la recherche, ainsi que de l'extension automatique.



7.2.2. Définition de catégories de tri

Il est bien entendu possible de définir plusieurs catégories et de les gérer à travers cette application. Certaines catégories, comme le type d'un document (HTML, PDF, Word, ...) ou sa langue (Français, Anglais, ...) sont fournies de façon automatique par AFS et n'ont pas à être créées. De même les catégories créées à partir de données indexées (champs d'une base de données, méta données de fichiers, ...) n'ont pas à être créées ici.



La copie d'écran ci-dessus fait apparaître deux catégories, l'une appelée « Compétence Géo », et l'autre « Régions ».

Chaque catégorie est définie par un libellé, un identifiant, une cardinalité :

- **Libellé** : le nom utilisé pour l'affichage.

- **Identifiant** de la catégorie : il est utilisé dans la configuration du périmètre pour référencer cette catégorie ou lors de l'appel au moteur.
- **Cardinalité** : indique si un document peut prendre une seule ou plusieurs valeurs parmi la liste des valeurs possibles pour cette catégorie. La cardinalité peut donc être simple ou multiple. Elle est simple lorsqu'un document indexé ne peut être associé qu'à une seule valeur d'une catégorie, et multiple lorsqu'un document peut avoir plusieurs valeurs dans une même catégorie.

Référentiel projet | Sources de données | Paramétrage global

Paramétrage global **Catégories**

Editer la catégorie Type de site

Libellé de la catégorie * :

Identifiant de la catégorie :

Cardinalité : ▾

Valeurs

Libellés	Identifiants	
<input type="text" value="Constructeur"/>	<input type="text" value="CONS"/>	<input type="button" value="🗑"/>
<input type="text" value="News et info"/>	<input type="text" value="NEWS"/>	<input type="button" value="🗑"/>
<input type="text" value="Forums et Blogs"/>	<input type="text" value="FRM"/>	<input type="button" value="🗑"/>
<input type="text" value="Vendeur"/>	<input type="text" value="STORE"/>	<input type="button" value="🗑"/>

La copie d'écran ci-dessus montre que la catégorie appelée «Type de site » a pour identifiant « TYP », que sa cardinalité est multiple, et que les différentes valeurs possibles sont au nombre de 4.

Il est très facile d'ajouter une valeur, de modifier ou de supprimer une valeur existante.

La section suivante (Sources de données) illustre comment ces catégories peuvent être utilisées par exemple lors de la définition d'un périmètre de sites afin de catégoriser les différents sites et leurs pages.

7.3. Sources de données

L'application de configuration présente dans le back-office AFS permet enfin de configurer les sources de données à indexer. Seules les sources de données de type « périmètre de sites web » sont configurable à travers la version en mode hébergé. Les sources de données autres (bases de données, flux XML, ...) doivent être déclarées et configurées par le moyen de fichiers de configuration détaillés dans le manuel du développeur.

La suite de cette section détaille la façon dont un ou plusieurs périmètres de sites peuvent être déclarés et gérés.

7.3.1. Présentation

L'indexation de sites Web nécessite de définir au préalable un **périmètre**. Un périmètre est un ensemble de **pages Web** à parcourir et à indexer. Ces pages peuvent appartenir à un ou plusieurs sites.

Ainsi, la **définition d'un périmètre** consiste à définir, d'une part, l'ensemble des URLs à indexer et, d'autre part, les paramètres de cette indexation, y compris l'affectation des sites dans d'éventuelles catégories précédemment créées (cf section 7.2.2).

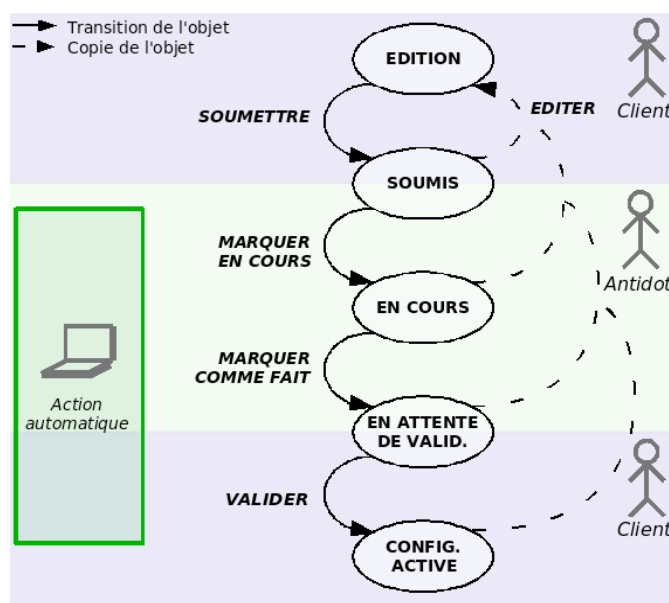
7.3.2. Processus de configuration d'un périmètre

Le processus de configuration d'un périmètre repose sur l'interaction entre l'administrateur, qui définit le périmètre, et AFS qui utilise ce périmètre comme paramètre de l'indexation.

Dans la mesure où l'indexation d'un périmètre est un processus qui peut prendre de quelques minutes à plusieurs semaines selon sa taille (nombre de sites déclarés, nombres de pages à indexer, ...), un mécanisme de soumission et de suivi d'une configuration est incluse.

Ce processus d'évolution dans le temps d'un périmètre à indexer comporte un ensemble d'actions illustrées dans la figure ci-dessous. Chaque action effectuée implique un changement d'état du périmètre. Le périmètre peut être dans l'un des cinq états suivants :

- « Edition » : l'état de base, lorsque vous êtes en train de modifier le périmètre en ajoutant des sites ou en modifiant des sites existants.
- « Soumis » : une fois que la définition du périmètre est terminée, le client peut soumettre (cf. §7.4.3) le périmètre à AFS afin qu'il soit pris en compte lors de la prochaine indexation. Le périmètre passe alors dans l'état « Soumis ».
- « En cours de réalisation » : ensuite, AFS effectue l'indexation du périmètre selon la configuration soumise. Le périmètre est alors « En cours de réalisation ».
- « En attente de validation » : lorsque l'indexation est terminée, le client doit valider le périmètre (une phase de recette en quelque sorte).
- « En Configuration active » : une fois validé, le périmètre est mis en production. C'est celui qui est utilisé par les agents pour répondre. Il est dit en « Configuration active ».



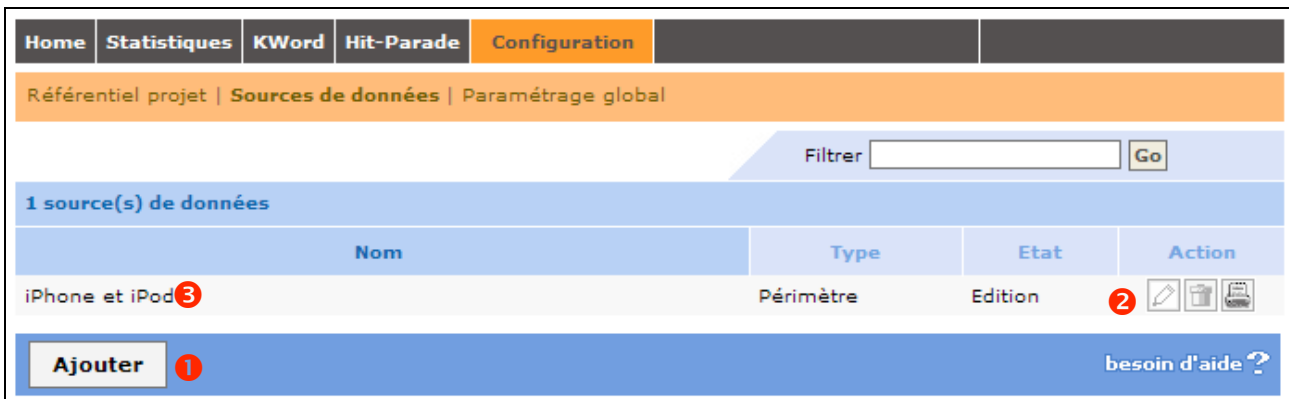
Les différentes actions possibles (demandes de chagement d'état) sont résumées dans le tableau ci-dessous :

Action effectuée	Auteur de l'action	Etat du périmètre une fois l'action effectuée
Enregistrer (cf. § 7.4.1)	Client	Edition
Soumettre (cf. § 7.4.3)	Client	Soumis
Marquer en cours	AFS	En cours de réalisation
Marquer comme fait	AFS	En attente de validation
Valider	Client	Configuration active

Remarque : Les actions « Enregistrer », « Soumettre », et « Valider » sont accessibles uniquement par le client et les actions « Marquer en cours » et « Marquer comme fait » ne peuvent être effectuées que par Antidot.

7.3.3. Créer/modifier la configuration d'un périmètre

1. Pour créer un périmètre, allez dans Configuration -> Sources de Données, puis cliquez sur le bouton "Ajouter".
2. Pour modifier la configuration un périmètre, cliquez sur le crayon
3. Pour entrer dans un périmètre et modifier la liste des sites qui constituent ce périmètre, cliquer sur son nom.



Créer un périmètre ou modifier sa configuration vous permet d'accéder à l'écran suivant :

Référentiel projet | Sources de données | Paramétrage global

Edition du périmètre iPhone et iPod

Configuration du périmètre

Nom du périmètre * :

Nombre max de sites :

Nombre max de pages : 1

Nombre max de pages par site :

Profondeur maximum : 2

Fréquence d'indexation : 3

Langues conservées :

Français Anglais Allemand 4

Espagnol Italien Portugais

Japonais Chinois Coréen

Formats indexés :

5 HTML Flash PDF

Texte, Doc, RTF Excel Powerpoint

Méthode d'indexation avancée :

6 Aucune

Lemmatisation

Phonétique

Floue

Paramètres

Réordonner les paramètres

Nombre de paramètres pris en compte : 7

Variables de session : 7

Divers

Détecter les miroirs 8

Accepter les soumissions externes

Dans ce formulaire, les champs suivis du caractère * sont obligatoires. Ainsi, seul le champs « Nom du périmètre » est obligatoire. Les autres paramètres peuvent être laissés vides, auquel cas les paramètres par défaut d'AFS s'appliqueront.

Les options définies à partir de ce formulaire seront appliquées sur l'ensemble des sites du périmètre.

Ces différents paramètres sont :

1. en ensemble de zones grisées et non accessibles qui sont définies par Antidot et qui dépendent de l'offre commerciale souscrite ou d'un réglage accessible seulement à l'administrateur de la solution.
 - **Nombre max de sites** : le nombre maximum de sites qui peuvent être définis dans un périmètre
 - **Nombre max de pages** : le nombre maximum de pages cumulées qui seront indexées dans ce périmètre.
 - **Nombre max de pages par site** : le nombre maximum de pages qui seront indexées pour un site. Une fois ce nombre atteint, les liens ne sont plus suivis.Si ces champs sont laissés vides, alors aucune limite ne s'applique.

2. **Profondeur maximum** : permet de définir la profondeur maximale des pages à indexer. Les pages définies comme point d'entrée dans le site (l'URL du site et les URLs ajoutées dans la zone Seed) ont par définition une profondeur de 0 (zéro). Les pages pointées par ces pages d'entrée ont une profondeur de 1, ... les pages pointées par une page de profondeur N ont une profondeur N+1. Lorsqu'une page est pointée par plusieurs pages, le première valeur de profondeur est conservée, c'est-à-dire la valeur la plus petite qui caractérise le parcours (nombre de clics) le plus court pour aller d'une page d'entrée à cette page.
Lorsque la profondeur maximum est définie, seule les pages Web ayant une profondeur inférieure ou égale à cette valeur sont conservées.

3. **Fréquence d'indexation** : précise la fréquence pour une indexation automatique du périmètre. La liste des valeurs possibles est définie dans le tableau ci-dessous :

Fréquence	Description
everyday	Tous les jours (monday, tuesday, wednesday thursday, friday, saturday, sunday)
weekday	Tous les jours de la semaine (monday, tuesday, wednesday thursday, friday)
weeknight	Tous les soirs de la semaine (sunday, monday, tuesday, wednesday thursday)
yearly	Une fois par an (exemple : yearly@january@01@22h50 qui peut être abrégé par january@01@22h50)
monthly	Une fois par mois (exemple : monthly@01@22h50 qui peut être abrégé par monday@22h50)
daily	Une fois par jour (exemple : daily@22h25)
hourly	Une fois par heure (exemple : hourly@15)

Remarques : Chaque élément, séparé par @ dans la fréquence d'indexation, peut contenir une liste de valeurs séparées par des virgules.

Exemple : (monday,sunday)@22h50

(= le lundi ou le dimanche à 22h50, le ou étant inclusif)

NB : Il faut noter que cette fréquence d'indexation n'est pas prise en compte actuellement. Elle sera exploitée dans la prochaine version d'AFS.

4. **Langues conservées** : permet de définir les langues des documents à conserver, et celles à ignorer.
5. **Formats indexés** : définition des formats des fichiers à prendre en compte lors de l'indexation, et les formats à ignorer.
6. **Méthode d'indexation avancée** : permet de définir le type de traitement linguistique voulu. Les différentes méthodes d'indexation (lemmatisation, floue, et phonétique) sont définies dans le manuel de configuration AFS (§ 4.1.1).
7. Les options définies, dans la section « Paramètres » du formulaire, permettent de limiter les risques de doublons ou d'aider AFS dans la détection des paramètres de session. Cet ensemble de paramètres est destiné aux utilisateurs avancés et ils ne devraient pas être renseignés tant que des problèmes de crawl précis n'ont pas été relevés.
 - **Réordonner les paramètres** : si cette option est cochée l'ordre des paramètres d'une URL ne sera pas conservé et les paramètres seront réordonnés lors de l'indexation afin d'obtenir une forme normalisée des URLs.
 - **Nombre de paramètres** : permet de définir le nombre de paramètres HTTP à conserver lors de l'analyse des URLs
 - **Variables de session** : déclaration des variables de session possibles pour ce périmètre (la plupart des variables sessions habituelles sont détectées et gérées par AFS, il y a donc peu de chance d'avoir à déclarer de nouveaux paramètres).
8. Les options définies dans la partie « Divers » du formulaire sont les suivantes :
 - **Détecter les miroirs** : permet d'activer dans AFS le mécanisme de détection automatique des pages miroirs. Ce sont des pages au contenu strictement identique mais présentes sur des sites différents. AFS ne conservera alors dans l'index qu'une seule version de chaque page.
 - **Accepter les soumissions externes** : pour autoriser le service qui permet l'ajout de sites ou de page à indexer sans passer par l'interface de back-office.

Une fois l'ensemble des paramètres renseignés ou motivés, il est essentiel de cliquer sur le bouton "**Enregistrer**" situé en bas du formulaire, sans quoi les modifications seront perdues si vous changez de page.

7.3.4. Ajouter/modifier un site du périmètre

Une fois un périmètre créé et configuré, vous pouvez déclarer les sites constituant ce périmètre à indexer.

Pour cela, il suffit de cliquer sur le nom du périmètre lui-même comme indiqué en début de section 7.3.3.



Deux onglets sont disponibles :

- L'onglet **Sites** qui permet de déclarer et gérer les sites du périmètre (voir plus loin).
- L'onglet **Alias de Sites** permet de définir comme un même site des URLs différentes. Par exemple, le site Apple France a pour URL www.apple.fr. Mais cette URL renvoie vers www.apple.com/fr et certaines parties du site comme l'espace boutique se situent sous l'URL store.apple.com. Si vous entrez comme URL uniquement apple.fr, toutes les pages se trouvant sous les autres URLs seront ignorées dans le cas où vous n'indexez des pages que dans la mesure où leur nom de domaine est dans le périmètre. Un moyen simple d'éviter cela est de regrouper ces différentes URL dans un même groupe appelé Alias.

Pour rassembler un groupe d'URL sous un unique site, cliquez sur l'onglet « **Alias de Sites** », puis sur « Ajouter ». Cette page propose un formulaire qui permet de regrouper sous un même libellé un ensemble d'URL qui seront considérées comme équivalentes par le moteur. Une fois les différentes formes d'URL ajoutées, cliquez sur « Enregistrer ».



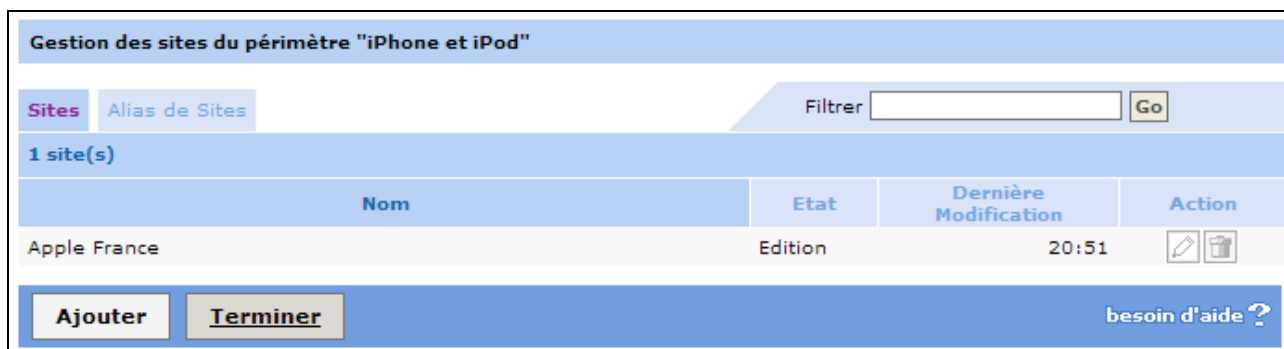
Il est possible de modifier un alias en ajoutant ou supprimant des URL à un groupe en cliquant sur son libellé dans la liste des Alias déclarés.

La boîte de recherche **Filter** permet de limiter l'affichage des différents sites du périmètre, en

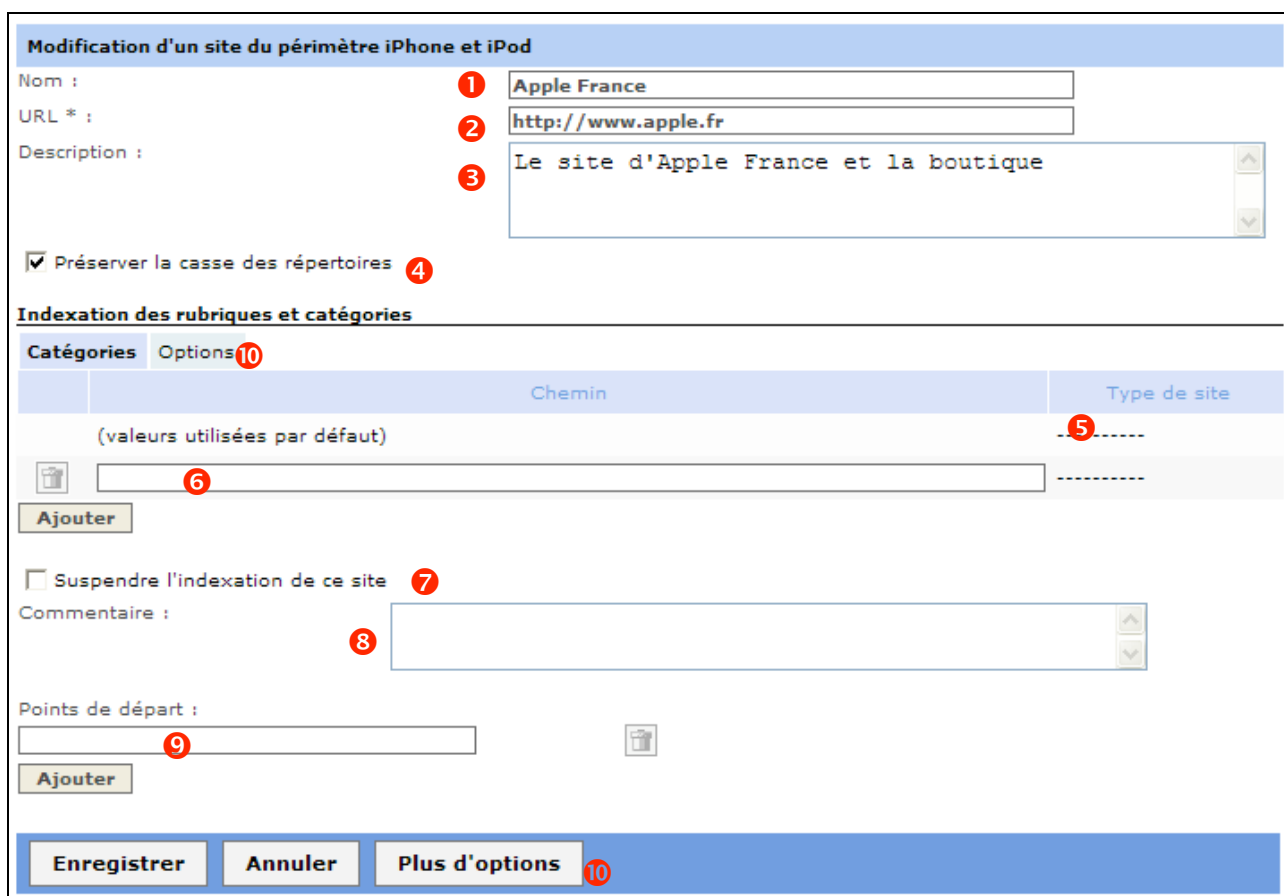
effectuant une recherche sur les champs de définition d'un site. Il prend en compte la troncature à droite et à gauche. Par exemple, si les sites www.apple.fr, store.apple.com et www.apple.com/fr sont déclarés comme différents sites dans le périmètre, la recherche « apple » renverra les trois sites cités précédemment. Cette boîte de recherche permet en particulier de vérifier si un site a déjà été indexé.

Déclaration d'un site

Pour ajouter un site, il suffit de cliquer sur le bouton Ajouter qui se trouve en bas de la liste des sites déjà existants du périmètre, ou bien de cliquer sur le titre d'un site existant ou sur l'icône représentant un crayon pour modifier sa déclaration.



Dans les deux cas, vous arrivez sur le formulaire de déclaration d'un site.



Les deux premiers champs sont obligatoires :

1. le **Nom** que vous donnez au site. C'est ce nom qui apparaît dans la liste du périmètre.
2. l'**URL** racine du site (exemple : <http://www.apple.fr>). Attention, il n'est pas possible dans cette case de saisir un chemin. Seule l'URL racine est acceptée, elle doit toujours commencer par <http://>. Il ne faut pas mettre de slash à la fin de l'URL.
3. Un champ **Description** vous permet de saisir un texte libre de description du site (son contenu, son intérêt dans le périmètre, ...). Ce champ est rendu disponible dans AFS et peut être exploité dans une application.
4. La case à cocher **Conserver la casse** permet le cas échéant de préciser si les URLs du site sont insensibles à la casse (majuscule/minuscule). En effet dans certains sites Web développées avec les technologies de Microsoft, les URLs comme <http://site.com/Répertoire1/Fichier1.html> et <http://site.com/répertoire1/fichier1.html> sont en fait les mêmes pages. Il est donc important pour AFS de considérer ces deux URLs comme identiques, auquel cas, il faut décocher cette case. Par défaut, laissez la cochée.
5. Il est possible pour ce site de lui attribuer des valeurs dans les différentes **catégories** créées. Dans cet exemple, nous avons précédemment créé une catégorie « Type de site » (voir la section 7.2.2 page 28). Nous retrouvons cette catégorie en tant que colonne du tableau. La première ligne de ce tableau correspond à la valeur par défaut pour tous le site dans la catégorie.

Cliquez sur la ligne (« ---- ») afin de faire apparaître la fenêtre de saisie des valeurs :

Modification d'un site du périmètre iPhone et iPod

Nom :

URL * :

Description :

Préserver la casse des répertoires

Indexation des rubriques et catégories

Catégories	Options
(valeurs utilisées par défaut)	

Suspender l'indexation de ce site

Editer la catégorie : Type de site

Cardinalité : Multiple

Valeurs

Constructeur

Forums et Blogs

News et info

Vendeur

[Enregistrer](#) [Annuler](#)

Cochez les valeurs que vous voulez affecter à ce site pour cette catégorie, puis cliquez sur Enregistrer pour mémoriser votre choix et fermer cette fenêtre de saisie.

6. Il est également possible de raffiner cette catégorisation sur des sous parties du site. Pour cela, vous pouvez ajouter déclarer des sous-répertoires (en commençant toujours par / et en n'oubliant pas de terminer par /* pour que la règle s'applique à tous les contenus qui se trouvent sous ce répertoire). Puis pour les sous-répertoires ainsi déclarés, vous pouvez cliquer dans la colonne d'une catégorie pour en déclarer la valeur.

Indexation des rubriques et catégories	
Catégories	Options
Chemin	Type de site
(valeurs utilisées par défaut)	Constructeur
<input type="text" value="/support/*"/>	Constructeur, News et info
<input type="button" value="Ajouter"/>	

- La case à cocher **Suspendre l'indexation** permet d'indiquer à AFS de suspendre l'indexation du site, sans pour autant le supprimer du périmètre. Il sera ainsi possible de reprendre l'indexation du site, en décochant la case.
- La zone de saisie **Commentaire** est destinée à la saisie d'un commentaire interne au back-office sur la configuration ou l'indexation du site (par exemple la raison pour laquelle l'indexation du site a été suspendue ou pour laquelle un site a été ajouté au périmètre ou noter qu'il faut une surveillance particulière sur ce site...).
- La zone de saisie **Point de départ** permet de donner au moteur des URLs différentes de l'URL racine du site pour commencer son indexation. En effet, dans certains cas, l'URL racine du site ne contient pas de liens à suivre et le moteur restera bloqué sur cette première page sans pouvoir aller plus loin. Cela arrive par exemple avec les sites dont la page d'accueil utilise la technologie Flash ou pour les sites dont tous les liens sont écrits en javascript. Par ailleurs, ajouter des points de départ permet d'optimiser le crawl du site et de tendre plus facilement vers l'exhaustivité.

Vous pouvez alors ajouter manuellement des URLs vers des pages d'entrées plus adaptées (comme le plan du site par exemple) en commençant la saisie de l'URL après l'URL racine et toujours précédé d'un slash.

Points de départ :	<input type="text" value="/fr/sitemap/"/>	<input type="button" value="Ajouter"/>
	<input type="text"/>	<input type="button" value="Ajouter"/>

- Enfin, vous pouvez accéder à plus d'options de configuration soit de façon globale pour le site en cliquant sur le **bouton « Plus d'options »** qui se situe en bas de la page, soit en cliquant sur l'**onglet « Options »** pour les options propres aux sous-répertoires.

L'onglet Options permet :

- « **Crawl** » : de préciser si un sous-répertoire ou des pages dont l'URL contient une suite de caractères dont la règle sera précisée seront crawlés. Par exemple, si le site propose les pages dans une version imprimable, un paramètre du type format=print peut être utilisé dans l'URL. Afin d'éviter un doublon dans la page, il est possible d'interdire le crawl sur ces pages. Pour cela, il suffit de déclarer la chaîne de caractère précédée et suivie du caractère *. Dans le cas de notre exemple, cela donnera *format=print*. Pour les sous-répertoires donc un exemple est donné ci-dessous, nous indiquons au moteur que le contenu du sous-répertoire /fr/downloads (téléchargements) ne doit pas être indexé.
- « **Langue** » : de forcer la langue pour un sous-répertoire (et tous les sous contenus), car même si le moteur détecte automatiquement la langue d'une page, celle-ci peut

contenu plusieurs langues et on peut ainsi forcer le moteur.

- « **Rafrâichissement** » : définir une fréquence de rafraîchissement spécifique pour le site ou une sous partie du site (la syntaxe est la même que celle présentée au 7.3.3 pour le périmètre). Non pris en compte par la version 6.5 d'AFS.

- « **Priorité** » : de modifier la priorité des sous parties du site (en sous ou surpondérant le contenu). Non pris en compte par la version 6.5 d'AFS.

Indexation des rubriques et catégories					
Catégories		Options			
	Chemin	Crawl	Langue	Rafrâichissement	Priorité
(valeurs utilisées par défaut)					
	<input type="text" value="/support/*"/>	<input checked="" type="checkbox"/>	(automatique) ▾	<input type="text"/>	<input type="checkbox"/>
	<input type="text" value="/fr/*"/>	<input checked="" type="checkbox"/>	Français ▾	<input type="text"/>	<input type="checkbox"/>
	<input type="text" value="/fr/downloads/*"/>	<input type="checkbox"/>	(automatique) ▾	<input type="text"/>	<input type="checkbox"/>
<input type="button" value="Ajouter"/>					

Cliquer sur « Plus d'options » fait apparaître les éléments suivants :

Options	
Priorité	<input type="checkbox"/> 1
Profondeur maximum	<input type="text"/> 2
Vitesse de crawl	
<input checked="" type="checkbox"/> Utiliser la valeur définie pour le périmètre	3
<input type="checkbox"/> Ce site nécessite une authentification	4
Nom d'utilisateur :	<input type="text"/>
Mot de passe :	<input type="text"/>
Domaine :	<input type="text"/>
Nombre max de pages pour ce site :	<input type="text"/> 7
Paramètres	
<input checked="" type="checkbox"/> Utiliser la configuration du périmètre	
Variables de session :	5
<input type="text"/>	
<input type="button" value="Ajouter"/>	
<input type="button" value="Enregistrer"/> <input type="button" value="Annuler"/> <input type="button" value="Moins d'options"/> 6	

1. **Priorité** : permet de sous/sur pondérer ce site au sein du périmètre (non utilisé par AFS en v6.5. Prévu pour la version 7).

2. **Profondeur maximum** : permet de limiter la profondeur d'indexation de ce site en particulier. Ce champ permet de redéfinir pour ce site la valeur qui a éventuellement été fixée dans la configuration du périmètre. Dans le cas où il ne faut crawler que la page d'accueil du site, cette option devra être fixée à 1.
3. **Vitesse de crawl** : permet de définir pour ce site une vitesse de crawl différente, c'est à dire le temps d'attente (en secondes) entre deux requêtes. Cela peut-être utile pour des sites web ayant peu de bande passante afin de ne pas les surcharger lors de l'indexation. La zone de saisie apparaît si vous décocher la case.
4. **Authentification** : si le site nécessite une authentification, saisir ici un nom d'utilisateur, un mot de passe et éventuellement un domaine (dans le cas d'une authentification de type NTLM).
5. **Paramètres** : permet de redéfinir les paramètres qui ont été fixés dans la configuration du périmètre.
6. **Moins d'options** : cliquer sur ce bouton permet de masquer ces options supplémentaires.
7. **Nombre de pages maximum par sites** : cette option permet de déterminer le nombre de pages maximum à crawler pour le site configuré. Cette option ne doit être renseignée que dans le cas où il est connu que le site contient plus de pages que la configuration par défaut indiquée pour le périmètre.

La validation du site et/ou des modifications se fait en cliquant sur Enregistrer situé en bas de liste.

7.3.5. Soumettre un périmètre

Une fois le périmètre enregistré, n'oubliez pas qu'il faut le Soumettre afin que les modifications apportées soient prises en compte par AFS pour de la prochaine indexation. Dans le cas contraire cette nouvelle configuration restera dans le statut « En édition » et c'est la configuration précédente qui sera utilisée.

Pour soumettre le périmètre, allez dans la liste des sources de données (Configuration -> Sources de données) puis cliquez sur Edition (le petit crayon) correspondant au périmètre. Le bouton Soumettre se trouve en bas de la page.