

Fonctionnalités AFS

Référence : AFS/WP/FONCS v6.5c

Introduction : principe de fonctionnement

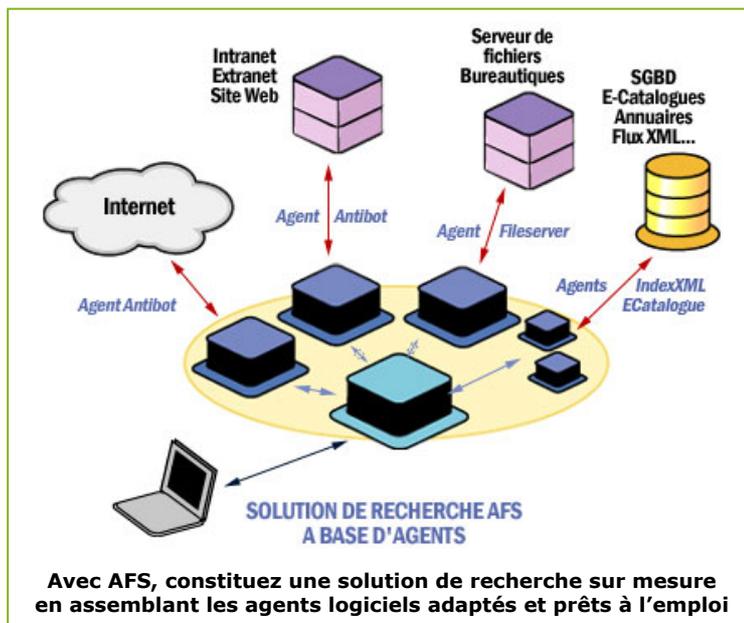
Comme décrit dans le document « Technologie & Innovation » (référence AFS/WP/TEKI), **Antidot Finder Suite** (AFS) est une boîte à outils d'**agents¹ logiciels** permettant de construire des solutions de recherche et d'accès à l'information.

Pour bâtir une **solution** parfaitement **adaptée**, il suffit de choisir, d'assembler et de configurer un ou plusieurs agents parmi tout un ensemble de **modules prêts à l'emploi**. Chacun des agents mis en œuvre aura la charge d'une source de données (sites web, serveurs de fichiers ou de mails, archives, bases de données, annuaires, flux XML, catalogues électroniques, ...).

Cette technologie à base d'agents permet de traiter au mieux en chaque source de données

en prenant en compte ses contraintes d'accès, de sécurité, de format, de structure et de contenu. Les agents les plus utilisés sont :

- **AntiBot** : indexation de sites et de serveurs de fichiers ;
- **IndexXML** : indexation de données structurées comme les SGBD, les flux XML, RSS, ... ;
- **eCatalogue** : catalogues électroniques ;
- **Directory** : annuaires ;
- **Keywords** : associations de mots clés à des contenus éditoriaux ;
- **Hit-Parade** : analyse des recherches des utilisateurs, des documents les plus accédés, ...



Lorsqu'un utilisateur pose une question, celle-ci est envoyée à l'élément central de la solution appelé **RMS** (Request Monitoring System) qui a la charge de transférer la requête aux différents agents qui composent la solution, puis de récupérer leurs réponses afin de les agréger pour produire la réponse finale qui sera fournie à l'utilisateur. C'est au niveau du RMS que sont appliquées les **règles de sécurité**, de **tri**, de **dédoublonnage**, de **pondération** et de **catégorisation** qui sont automatiquement configurées lors de l'installation de la solution et ajustées par la suite.

Techniquement parlant, AFS met en œuvre des mécanismes de **propagation / agrégation de flux XML** avec la possibilité de traiter à la volée ces flux XML en appliquant des **règles métiers** qui sont spécifiées de façon externe.

Cette **architecture unique** fait d'AFS le produit le plus à même d'aborder les nouveaux challenges de la recherche d'information en intégrant de nouveaux supports et de nouveaux formats tels que l'audio et la vidéo, mais aussi de s'interfacer avec des systèmes externes pour consolider leurs connaissances.

¹ un agent est un logiciel qui remplit de façon autonome une tâche particulière au sein d'un système plus vaste.

Grâce aux compétences et aux spécificités de ses différents agents logiciels, AFS offre une **couverture fonctionnelle extrêmement large** à même de répondre aux besoins les plus avancés. Ce document présente quelques unes de ces fonctionnalités en s'appuyant sur des exemples concrets issus de cas clients réels, la plupart étant accessibles et visibles sur internet.

Les fonctionnalités couvertes par AFS se regroupent dans les catégories suivantes :

1.	Accès aux données.....	2
2.	Formats supportés	2
3.	Indexation	3
4.	Modes de recherche	4
5.	Pertinence	7
6.	Aides à la recherche	7
7.	Présentation des résultats	11
8.	Sécurité.....	12
9.	Intégration.....	13
10.	Performance.....	12
11.	Back-office.....	15

1. Accès aux données

Que les sources de données soient internes ou externes à l'entreprise, publiques ou privées, en accès libre ou sécurisé, AFS est capable de mettre en œuvre des stratégies adaptées à chaque situation.

Crawl

Le crawl, ou « parcours », est une façon d'accéder aux données qui simule le comportement d'un utilisateur en utilisant les méthodes standard. AFS supporte nativement les protocoles **HTTP, HTTPS, FTP, SFTP, SCP, NFS** et **CIFS, POP3** et **IMAP4, LDAP**.

Pour l'indexation des contenus web sur internet comme en intranet, AFS peut indexer un **périmètre de sites** choisis en fonction de critères fixés par l'administrateur (liste de sites, parties de sites, ...). Lors du parcours de ces sites, le module d'indexation normalise les URLs dynamiques afin de ne pas indexer plusieurs fois le même document. Le module **détecte** automatiquement **les doublons** (par exemple : les pages standards versus les versions imprimables ou celles que l'on envoie par mail) et **les pages miroirs** et il peut ne retenir que les documents « maîtres » au moyen d'algorithmes basés sur des analyses hypertextes ou sur un principe de génération de signature des contenus.

Le système est également capable de reconnaître les **identifiants de session** et donc de parcourir les **sites dynamiques**. Il est même possible de prendre en compte d'autres paramètres, de **s'authentifier** et de **scripter le parcours d'un site** (par exemple pour en extraire une base de données).

Lors de l'indexation, AFS se base sur le contenu, la forme des URLs et les méta-données pour réaliser un **enrichissement taxonomique** qui sera utilisé lors de la présentation des réponses pour regrouper les informations ayant un lien (voir catégorisation).

La fréquence de rafraîchissement peut être définie spécifiquement pour chaque site ou partie de site afin d'optimiser la mise à jour des index.

Bases de données

L'indexation des bases de données est réalisée soit au travers d'**exports** (natifs ou **XML**), soit par un **accès direct** à la base au moyen d'un driver **ODBC** ou natif (LDAP dans le cas d'un annuaire par exemple, **jsr170** pour les systèmes documentaires comme Documentum).

La mise à jour du moteur peut se faire soit de façon différée par **indexation différentielle**, soit de façon **synchrone** par une prise en compte des modifications au fil de l'eau.

Dans tous les cas, la fréquence d'indexation et de mise à jour des index est réglable (de 1 seconde à 1 an) et peut être différente selon les éléments.

Extensibilité

Grâce à l'architecture ouverte à base de plug-in d'AFS, il est facile d'ajouter d'autres protocoles pour répondre à des **besoins spécifiques** afin d'indexer des sources de données internes ou externes particulières.

2. Formats supportés

AFS supporte les formats de document suivants :

- **formats web** : **HTML, PDF, Flash**, TXT, RFT (Rich Text Format), ...
- **formats bureautiques** : **Microsoft Office** (Word, Excel, PowerPoint) et **Open Office**.

D'une façon générale, AFS ne convertit pas les documents en texte avant de les indexer mais procède à une **analyse** dans leur **format natif** afin de conserver les informations de présentation associées : ainsi les mises en forme (gras, italique, souligné, taille de police, titre, paragraphe, positions des mots, ...) sont conservées et utilisées dans les algorithmes de calcul de **pertinence**.

Pour les formats web, et tout particulièrement pour le HTML et le PDF, AFS dispose de ses propres analyseurs ce qui lui permet d'assurer que **100% des documents** seront lus et **indexés** dans leur totalité. C'est là un aspect différenciant, particulièrement dans le cas du PDF où les autres solutions autres ont un taux d'échec de lecture des documents qui est souvent de plus de 20%.

AFS indexe également de façon optimale les **données structurées**, qu'il s'agisse de formats simples comme le **CSV** ou de formats plus complexes comme le **XML**. Dans le cas du XML par exemple, AFS conserve l'ensemble des informations de structure **XPath** afin de permettre des recherches multicritères fines.

3. Indexation

Analyse

AFS détecte et gère les différents jeux de caractères des documents, qu'ils soient sur un ou sur plusieurs octets comme l'Unicode. Elle gère nativement les accents, ainsi que les différences de casse (majuscule / minuscule) ou les abréviations ("CEE" et "C.E.E" sont reconnus comme identiques).

Pour indexer les documents, AFS utilise des technologies avancées d'**analyse lexicale**¹ et **syntaxique**² ce qui lui permet de réaliser diverses fonctions comme :

- la **gestion des flexions**, c'est-à-dire des différentes formes d'un même mot : singulier/pluriel, masculin/féminin ou les formes conjuguées pour un verbe.
- la **radicalisation** : technique qui permet de ramener un mot à sa racine afin de lier entre eux des mots ayant un même signifiant. Exemple : sifflet, siffler et siffleur sont ramenés à leur radical siffl.
- l'**extraction de groupes nominaux** et de formes singulières (noms propres, adresses).

Ces différentes technologies peuvent être exploitées directement ou alors servir de base à des fonctions plus complexes comme le calcul des expressions contextuelles (voir « 6. Aides à la recherche »).

Exemples :

- Par défaut, que la recherche soit `Téléphone`, `téléphone`, `telephone` ou `teléPhone`, les réponses seront les mêmes.
- Avec la prise en compte des flexions, les recherches `téléphone` et `téléphones` rendront les mêmes résultats.
- Avec la radicalisation activée, les requêtes `téléphone` et `téléphoneront` donnent des réponses identiques.

Pour réaliser les fonctions d'analyse linguistique avancées, AFS intègre la technologie XeLDA de Témis. Cette technologie, issue à l'origine des laboratoires de recherche de Xerox, est aujourd'hui propriété de Témis, un des partenaires technologiques d'Antidot.



Langues

AFS est une solution de **recherche multilingue** :

- elle **détecte** automatiquement **la langue des documents** ; elle choisit alors de les indexer ou de les ignorer selon leur langue et en fonction de la configuration (par exemple indexer les documents en français et anglais mais ignorer ceux en chinois). Ces choix

¹ L'analyse lexicale est la phase de découpage d'un texte en mots ce qui inclut la reconnaissance des cas particuliers comme les abréviations, ...

² L'analyse syntaxique consiste à retrouver les éléments grammaticaux d'une phrase (groupes nominaux, verbes, ...)

peuvent être définis par source de données, ou même selon les sous parties d'une même source.

- lors de leur indexation, **AFS catégorise** automatiquement les données **selon leur langue** et selon leur source si bien que lors d'une requête, elle remontera les documents en fonction de leur langue, de leur source et de la langue de l'utilisateur selon la configuration. Par exemple, lors d'une requête en anglais, il sera possible de ne répondre qu'avec des documents anglais (doc, pdf, ...) mais avec des entrées de base de données en anglais et en français.

Dictionnaires et thésaurus

Lors de l'indexation, AFS peut appliquer des **dictionnaires de synonymes** ou des **thésaurus**, multilingues ou cross-lingues, permettant ainsi de prendre en compte des spécificités métier ou tout simplement d'élargir les requêtes.

Une spécificité d'AFS repose sur sa capacité à exploiter pleinement les thésaurus et leur structure :

- lors de la recherche en faisant de l'**extension sémantique (autopostage)** de façon optimale puisque la pertinence d'un terme varie en fonction de son éloignement avec le terme synonyme recherché.
- AFS peut extraire des **taxonomies** (plans de classement) du thésaurus et les utiliser pour **catégoriser** l'information en rattachant automatiquement les documents aux catégories et en exploitant ces plans de classement comme outil d'aide au filtrage et à la navigation dans les réponses.

4. Modes de recherche

Recherche plein-texte

AFS propose par défaut aux utilisateurs un mode de recherche simple et en phase avec celui qu'ils utilisent au quotidien sur les moteurs de recherche Internet : la **recherche plein-texte par mots clés**.

AFS supporte les **expressions booléennes** (ET, OU, NON, recherche exacte), soit exprimées explicitement par l'utilisateur au moyen des opérateurs classiques, soit à travers une interface composée de plusieurs champs de saisie (« doit contenir les mots », « peut contenir les mots », « ne doit pas contenir les mots », « doit contenir exactement »).

Exemples :

- `location |bateau |voilier` pour louer un bateau ou un voilier
- `emploi -"mode d'emploi"` pour chercher les documents sur l'emploi, mais pas sur les "modes d'emploi".

Recherche phonétique

AFS propose également des fonctions de **recherche phonétique**, y compris l'indexation des contenus écrits "à la SMS". Cette fonctionnalité, activable pour source, partie de source ou type de document, ajoute de la tolérance et est à ce titre particulièrement adaptée à la recherche dans les espaces collaboratifs (forums, blogs, wiki) ou dans les mails du fait de la liberté d'écriture et des fautes récurrentes que l'on y trouve.

Recherche floue

AFS dispose d'algorithmes de **recherche floue** permettant d'étendre automatiquement la requête aux **orthographes approchantes**. Ces algorithmes peuvent être appliqués sélectivement à certaines sources, à des types de documents, ou même à un nombre limité de champs dans le cas d'une recherche dans des contenus structurés afin de répondre à des besoins spécifiques comme la recherche dans les annuaires ou dans les catalogues.

Exemples :

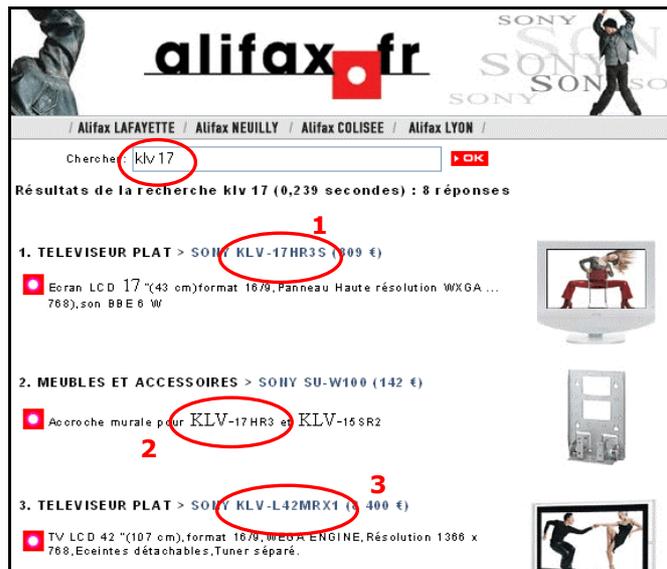
- L'activation de la recherche floue sur un annuaire de personnes permet pour la requête `dupont` de retourner en priorité les `dupont` mais aussi les `dupond`, les `duppon` ou les `dupind`.
- L'utilisation de la recherche floue au niveau d'un catalogue permet de retrouver facilement un produit par sa référence même si celle-ci est mal orthographiée ou incomplète.

Site : alifax.fr

Requête : klv17

La recherche floue appliquée à ce catalogue permet de remonter aussi bien :

1. la réponse la plus adaptée qui est le produit de référence klv-17hr3s.
2. le meuble associé à ce produit car une référence compatible a été trouvée dans la fiche descriptive.
3. un téléviseur différent (réf klv-l42mr1) mais appartenant à la même famille de produit et donc susceptible d'intéresser le client.



Extension automatique

AFS supporte la recherche par **extension automatique**, dite aussi par **troncature**. Ce mode permet de saisir le début d'un mot et d'obtenir des réponses avec tous les mots commençant de la même façon. Bien que ce ne soit pas le comportement par défaut d'AFS, ce mode peut être activé dans la configuration de deux façons différentes : soit par utilisation explicite (l'utilisateur doit ajouter le signe " * " à la fin du mot à étendre) soit par utilisation implicite (tous les mots recherchés seront automatiquement étendus).

Exemples :

- en configuration explicite, la requête "ecran pla*" permettra de trouver aussi bien les "écrans plats" que les "écrans plasma".
- en configuration implicite, la requête "voitur" retournera des réponses sur les voitures, mais aussi sur les `voiturettes` et sur les `voituriers`.

Recherche multicritères

Au-delà de la recherche plein-texte, AFS offre des possibilités avancées de **recherche multicritères** de façon similaire à ce qui est faisable avec une base de données mais avec beaucoup **plus de possibilités** puisque toutes les fonctionnalités décrites dans ce document sont également applicables à la recherche multicritères.

Les **critères utilisables** sont :

- pour les documents non structurés (web, bureautique) : les **méta-données existantes** (auteur, dates de création ou de modification, ...) ou **générées dynamiquement** lors de l'indexation à partir de règles applicatives définies lors de la configuration.
- pour les données structurées (bases de données, XML) : il est possible de rechercher dans n'importe quel **champ**, **tag** ou **attribut**, soit indépendamment soit en les regroupant.

Ces **critères** peuvent être **typés** (nombres, dates, ...) afin de ne pas être considérés comme du texte et donc d'offrir des mécanismes de filtrage spécifiques.

Site : lyon-business.org

AFS indexe un ensemble de données structurées (agenda d'événements, annuaires de personnes et d'entreprises, articles, news, FAQs, forums, médiathèque, espaces collaboratifs, ...) contenues dans des bases de données et dans un système de gestion de contenu.

Le portail propose une recherche globale mais aussi des fonctions de recherche avancée multicritères par type de contenu. Ici une recherche d'événements.

Cet exemple, illustre qu'il est possible de ne pas saisir de mot clé, afin de lister l'ensemble des données répondant à des critères donnés.

LYON-BUSINESS.ORG
PORTAIL ÉCONOMIQUE DE L'AGGLOMÉRATION LYONNAISE

ENTREPRENEUR | CRÉATEUR | INVESTISSEUR | PROFESSIONNEL TOURISME

AGENDA Recherchez les événements dans l'agglomération lyonnaise en sélectionnant un ou plusieurs critères

Mots-clés : Période : Le mois à venir

Type : Tous les types Date : du 18/02/2005 au 20/08/2005

Communes : Toutes les communes

RECHERCHER Tous les événements

ACCUEIL > OUTILS / AGENDA ÉCONOMIQUE > RÉSULTAT

522 réponses correspondent à votre recherche : ""

Retour

73 arts & culture
365 concerts & spectacles

// ARTS & CULTURE

Du 02/05/2005 au 04/05/2005
TUNING

Le 13/05/2005
MARYSE DELENTE : LES PETITES ET LA BELLE ... AU BOIS DORMANT

Du 01/04/2005 au 02/04/2005
JOSETTE BAIZ : TONIGHT

VERSION IMPRIMABLE
AJOUTER À MON ESPACE
ENVOYER CETTE PAGE

Recherche en langage naturel

AFS propose un module spécifique d'analyse des requêtes qui permet de transformer des recherches en langues naturelles en recherches structurées propres aux sources de données à interroger, permettant ainsi d'offrir une recherche unifiée sur des sources structurellement différentes.

Ce module, appelé « **Query Miner** », a par exemple été mis en œuvre par Pages Jaunes dans le cadre de son nouveau service de petites annonces (Annonces Jaunes).

Le **module d'analyse du langage naturel** spécifique permet aux utilisateurs de formuler leur recherche de façon libre, offrant ainsi un confort d'utilisation important.

Le vocabulaire à analyser peut être automatiquement récupérés depuis les thésaurus et les listes d'autorité.

QUOI ? Immo Auto Moto

appart avec cheminée parquet plus de 150m2
ex : appart à louer; maison à vendre; voiture année 2005

OÙ ?

Paris
ex : Toulouse; 59; Alsace; 75012

COMBIEN ?

entre 2 et 3 ME
ex : moins de 500 euros; 1000 euros; entre 400 et 500 €

TROUVEZ

L'ensemble des critères exprimés par l'utilisateur sont alors analysés et transformés en filtres compatibles avec les données exploitées et l'information disponible.

Les résultats sont affichés en liste de façon classique, alors que l'ensemble des paramètres permettant d'affiner ou d'élargir la recherche sont repris dans la colonne de gauche et peuvent être modifiés, ce qui a pour effet de relancer la recherche.

11 annonces

Vous êtes ici: Accueil > Immobilier

LISTE **CARTE** Voir ma sélection (0)

Ajouter à ma sélection Comparer mes annonces (0) Page(s) 1

	Date	m²	Pièces	Ville	Prix
<input type="checkbox"/>	24/11/06	195m²	8p.	Paris 07	2 100 000 €
<input type="checkbox"/>	07/01/07	204m²	7p.	Paris 07	2 180 000 €
<input type="checkbox"/>	17/11/06	265m²	7p.	Paris 07	2 180 000 €
<input type="checkbox"/>	07/01/07	204m²	6p.	Paris 07	2 180 000 €
<input type="checkbox"/>	17/11/06	284m²	7p.	Paris 16	2 236 500 €
<input type="checkbox"/>	29/12/06	262m²	6p.	Paris 07	2 240 000 €
<input type="checkbox"/>	18/11/06	240m²	8p.	Paris 06	2 321 000 €

PRECISER MA RECHERCHE

Vente / Location

Appartement

Nombre de pièces
Min Max OK

Surface (en m²)
Min 150 Max OK

Localité / Département
Paris OK

Budget (en €)
Min 2e +06 Max 3e +06 OK

Ancienneté

Etage

Oriantation

5. Pertinence

De très nombreux paramètres rentrent en ligne de compte dans le calcul de la pertinence. Ils proviennent des différentes phases de l'indexation :

- mode d'accès : pondération selon la source de données, selon les méta-données disponibles, ...
- format du document : pondération selon le type de document, pondération des différents champs dans le cas des données structurées, ...
- indexation : prise en compte de la mise en forme propre à chaque mot, de sa position dans le document (dans les titres, en début de paragraphe, ...), de sa densité, ...

Dans le cas où les différents **documents** sont **liés** entre eux (comme c'est le cas sur le web ou dans les intranets), AFS met en œuvre des **algorithmes** extrêmement **sophistiqués** permettant de calculer des scores associés à chaque document :

- le score d'autorité : il indique les sources ou documents de références pour un sujet donné, c'est-à-dire ceux qui sont les plus populaires (les plus référencés).
- le score de hub : il permet d'identifier les documents ou sites pivots, c'est-à-dire les documents "experts" susceptibles d'amener à la bonne information.

Ordre des mots

Dans tous les cas, AFS prend en compte non seulement la **position relative** mais également l'**écart entre les mots** si bien que l'utilisation d'opérateurs booléens de recherche comme « before » et « near » devient inutile. Avec AFS, les recherches « location de voiture » et « voiture de location » sont différentes et génèrent donc des résultats plus précis.

Analyse des liens entre les documents

Dans le cas où les différents **documents** sont **liés** entre eux (comme c'est le cas sur le web ou dans les intranets), AFS met en œuvre des **algorithmes** extrêmement **sophistiqués** permettant de calculer des scores associés à chaque document :

- le score d'autorité : il indique les sites ou documents de références pour un sujet donné, c'est-à-dire ceux qui sont les plus populaires (les plus pointés).
- le score de hub : il permet d'identifier les documents ou sites pivots, c'est-à-dire les "experts" susceptibles d'amener à la bonne information.

Tous ces critères sont utilisés conjointement afin de calculer la pertinence de chaque document pour une requête donnée, ce qui donne à AFS une **pertinence de tout premier plan**.

6. Aides à la recherche

Suggestion orthographique

Cette fonctionnalité permet de proposer à l'utilisateur des orthographes différentes de celle qu'il a utilisée pour la requête et qui donneraient plus de résultats.

Cette suggestion est forcément pertinente puisqu'elle ne propose que des mots présents dans les documents. Elle fonctionne par **apprentissage automatique** lors de l'indexation si bien qu'elle ne nécessite **aucun dictionnaire** ni **aucune intervention manuelle**.

Il est intéressant de noter qu'avec ce fonctionnement, si les documents contiennent un mot mal orthographié mais que la requête est juste, alors AFS suggère d'utiliser la mauvaise orthographe afin d'obtenir plus de résultats.

Requête : micorsoft

Cet exemple met en évidence que malgré l'existence de réponses à une requête erronée, AFS est capable de proposer une orthographe amenant plus de réponses.



Automatic Cross Content (ACC)

AFS propose une fonction ACC de repérage automatique des données ayant un lien sémantique qui permet de calculer, pour chaque document du corpus les **documents similaires**, c'est-à-dire les documents qui **traitent du même sujet** ou de sujets **connexes**. Une fois ce travail effectué, il est possible lors de l'affichage d'un document de proposer à l'utilisateur de poursuivre sa lecture en lui affichant la liste des informations similaires susceptibles de l'intéresser.

Cette fonction "Voir Aussi" permet donc de mettre en avant la richesse éditoriale tout en offrant à l'utilisateur un réel service à valeur ajoutée puisqu'on lui propose une aide à la navigation et un accès simplifié à l'information qui l'intéresse.

Analyse de corpus

Le moteur de catégorisation d'AFS peut également être utilisé sur un "tas" de documents afin qu'il **l'analyse automatiquement** et **découvre** seul **les catégories** et sous catégories (principe de segmentation de corpus).

Le résultat de cette analyse permet de déterminer les catégories les plus discriminantes pour classer l'information.

Ces catégories ne sont pas nécessairement les mêmes que celles qui auraient été choisies de manière empirique ou de façon à reprendre une stratégie de classement existante, mais elles peuvent constituer un complément ou une alternative efficace.

Related Topic Engine

Le module RTE d'extraction et de génération d'expressions contextuelles permet de lister les mots ou expressions les plus souvent associés à la recherche effectuée dans les données indexées. Ces expressions contextuelles peuvent contribuer à éclairer l'utilisateur sur un sujet donné afin de l'aider à mieux formuler sa requête.

Site : [Ecobase21](#)

Requête : climatisation

Cet exemple montre les expressions et mots clés les plus courants dans les documents répondant à la recherche.

The screenshot shows the Ecobase21 search interface. At the top, there is a logo with the letters 'E', 'C', and 'O' in colored boxes, followed by 'BASE 21'. The search term 'climatisation' is entered in a search bar, and the results show '51 documents trouvés'. Below the search bar, there is a section titled 'Filtrer les résultats' with a list of categories: Environnement (10), Social (8), Gouvernance (2), Economie (6), and Culture (1). A 'Valider' button is at the bottom of this list. To the right, there is a section titled 'Suggestions de recherche' with three suggestions: 'énergies renouvelables', 'matériaux de construction', and 'énergie géothermique'. Below this, there is a section titled 'Formation, Energie' with a brief description of the content.

Catégorisation

La problématique principale de l'utilisateur n'est pas tant de formuler sa requête de façon convenable que de pouvoir **naviguer rapidement** dans la liste des réponses afin d'y **retrouver** le plus **facilement** et le plus rapidement possible l'information qu'il cherche. A ce titre, il est évident que les longues listes de réponses telles que retournées par les moteurs de recherche Internet constituent un contre exemple d'**ergonomie** et d'**efficacité**.

Pour rompre avec cette difficulté AFS propose des fonctions de **catégorisation automatique** des données permettant de découper les réponses selon des catégories intelligibles par l'utilisateur et lui permettant ainsi de naviguer plus rapidement.

Ces **catégories** peuvent être :

- soit **définies à l'avance** afin de répondre à des exigences éditoriales. Par exemple en intranet, regrouper ensemble les documents techniques, commerciaux, marketing, juridiques, ...
- soit **définies à la volée** lors de la constitution de la liste de réponses en fonction du contenu des documents retournés ou du profil de l'utilisateur.

Site : credit-agricole.fr et les sites des caisses régionales.

La stratégie marketing du Crédit Agricole s'articule autour de 3 axes :

1. les types de clients (particuliers, entreprises, ...)
2. les univers de besoins (retraite, assurance, ...)
3. les produits bancaires.

AFS indexe les contenus des différentes caisses régionales ainsi que plusieurs sites nationaux et regroupe automatiquement les réponses aux requêtes selon les axes de communication marketing afin que le moteur soit à l'unisson



RECHERCHE
Accueil > Recherche

RESULTATS DE VOTRE RECHERCHE : 10 résultats pour deduction fiscale

→ PARTICULIERS

→ Assurances

1. Plan Vert Avenir
... de revenus pour votre retraite tout en bénéficiant d'économies d'impôts. Vos versements sont rémunérés chaque année et cette rémunération vous est définitivement acquise.

2. Plan Vert Vitalité
... performances des marchés financiers sur le long, voire le très long terme, et une sécurisation progressive de votre capital à mesure que la retraite approche.

3. Predissime 9
... de supports financiers proposés, en fonction de vos objectifs et de votre sensibilité au risque et décidez des évolutions de cette répartition dans la durée.

→ AGRICULTEURS

→ Résultats trouvés sur pleinchamp.com

1. Plan du site Pleinchamp.com
... ELEVAGE Remise en cause de l'exonération des GFA Irrévocabilité de l'option pour l'impôt sur les sociétés Comptes d'associés débiteurs et **déduction** des frais financiers Droit de vote et parts sociales démembrées Cession de bail entre associés et contrôle des structures Signatures des cautions et sociétés Obligation ...

→ PROFESSIONNELS

→ Résultats trouvés sur ce site

1. Extension Retraite Supplémentaire
... au moment de leur départ en retraite. Tous les salariés peuvent en bénéficier, ainsi que le chef

Search & Navigation Exploration System (SNES)

Les fonctions de catégorisation décrites ci avant permettent de découper les listes de réponses afin de regrouper les résultats appartenant à une même catégorie. Lorsque ces catégories deviennent trop nombreuses ou qu'elles ne sont plus disjointes (c'est-à-dire qu'un document peut appartenir en même temps à plusieurs voir à toutes les catégories), il devient difficile voire impossible d'effectuer la présentation.

Il existe une alternative qui consiste à **construire dynamiquement des catégories** candidates puis à lister pour chacune de ces catégories les **filtres** possibles. Ces catégories et leurs paramètres peuvent alors être affichés en marge des réponses.

L'exemple ci-dessous permet d'illustrer cette fonctionnalité dite de "recherche par paramètres".

Site : [intranet d'Atisreal Auguste-Thouard](http://intranet.d'Atisreal-Auguste-Thouard).

AFS indexe les bases de tiers (prospects, clients et offreurs) ainsi que les bases contenant les différents biens immobiliers à louer ou à vendre. Le moteur est intégré dans l'application métier et sert à rechercher un bien ou un tiers pour la constitution d'offres commerciales.

Dans l'exemple ci-dessous, la recherche se fait sur la base des tiers et la requête est "ville=limonest". On distingue dans la page de résultat les éléments suivants :

1. la zone de recherche multicritères
2. une liste de réponses formatée dans un tableau
3. un ensemble de **catégories construites dynamiquement** ("Type de tiers", "Profil du tiers", "Dossier en cours", ...) et pour chacune d'elle les paramètres possibles et le nombre de réponses correspondantes.
4. certains champs de la recherche multicritères (code postal ou code NAF, ...) ont été renseignés automatiquement avec les valeurs existantes dans la liste des réponses.

Grâce à ces informations contextuelles, l'utilisateur dispose d'une **vue d'ensemble** et d'**indicateurs précis** lui permettant d'**analyser** en un coup d'œil les différents types de réponses. Il peut alors facilement **affiner sa requête** en quelques clics qui lui permettent de ne conserver que les réponses correspondant à sa recherche.

1 RECHERCHE D'UN TIERS

Raison sociale : Code postal : Siret :
 Enseigne : Ville : N° tiers :
 Nom du contact : Code NAF :

3

Type de tiers :
 Clients (24)
 Prospects (86)

Profil du tiers :
 Confrères (2)
 Autres (108)

Direction Grands Comptes :
 Oui (1)
 Non (109)

Dossier en cours :
 Oui (20)
 Non (90)

2

110 réponses à limonest

N°	Raison sociale	Adresse	Code postal	Enseigne	Type	Nature	NAF	Effectif	Dossier	DGC
46068	HNE MEDICAL	451 CHE de Champivost BP 20 - 69579 LIMONEST cédex	69760	EST	PR				Au	0
58589	IDSYS	CHEMIN DU PAISY	69760	LIMONEST	CL	UTI			Au	0
58770	A.P.I	43 ALL DE LA CROISEE	69760	LIMONEST	PR	UTI			Au	0
58834	SOCIETE GITEM SCAME	540 ALL des hêtres	69760	LIMONEST	PR	UTI			Au	0
58927	SOCOMIE	1 R PARC D'ACTIVITES DE LIMONEST 540 ALLEE DES HETRES	69760	LIMONEST	CL	ADM			Au	0
58953	EUROTHERM CHESSELL	513 R DE SANS SOUCI	69760	LIMONEST	PR	UTI			Au	0
58997	SOCIETE MAJORETTE	IMMEUBLE L'ESSENTIEL 4 ALLEE DES HETRES	69760	LIMONEST	PR	UTI			Au	0
59011	SOCIETE ECS MAINTENANCE	1 R des Vergers PARC D'ACTIVITES	69760	LIMONEST	CL	UTI			Au	0
59054	SCI TOPEL	Parc du Vieux Logis	69760	LIMONEST	PR	UTI			Au	0
59093	MONSIEUR MONTVENEUR	165 ALL du MATHIAS	69760	LIMONEST	PR	UTI			Au	0

<< Précédent - 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 - Suivant >>

Fermer

Ce principe de navigation dans l'information grâce à des critères de choix proposés en marge du flux de réponse est appelé selon son mode de mise en œuvre « **recherche paramétrique** » quand les critères sont issus de paramètres ou « **recherche par facette** » dans le cas d'une navigation multidimensionnelle où les axes d'exploration sont disjoints et calculés à partir de taxonomies et de métadonnées.

Les exemples ci-après illustrent les avantages de la recherche paramétrique et de la recherche par facette appliquées à des corpus "textuels".

Site : recherches.nievre-tourisme.com

Ce site met en œuvre la recherche par facette. La colonne de gauche propose les différentes dimensions d'exploration de l'information. Un clic sur une catégorie permet de filtrer les objets présentés et de proposer des sous catégories permettant d'affiner la recherche (1).

L'ensemble des critères de recherches sélectionnés sont résumés en haut de colonne et peuvent être supprimés afin de ré-ouvrir la recherche (2).

Enfin, l'utilisateur peut, de façon classique, saisir un ou plusieurs mots clés lui permettant de préciser sa recherche (3). Les mots clés saisis agissent comme un filtre supplémentaire.

1

2

3

1

Nos propositions

Camping de la Base Sport et Nature de Chaumeçon
MARTIN-DU-PUY
 Situé au bord du Lac de Chaumeçon le centre caractérise par la qualité de ses activités sportives comprenant 7 emplacements avec sanitaires et bouée nautique de Chaumeçon 58140 SAINT-MARTIN-DU-PUY
 2 Suggestions sont associées à cette offre

Gîte de la Base Sport et Nature de Chaumeçon
PUY
 Au bord du lac de Chaumeçon, gîte de 29 coins de couchage entre lacs, rivières et forêts au sein du Parc National du Morvan, la base Sport et Nature de Chaumeçon vous permettra de procurer sensations et émotions dans un cadre naturel exceptionnel. Base nautique de Chaumeçon 58140 SAINT-MARTIN-DU-PUY
 4 Suggestions sont associées à cette offre

Gîte 639 - CERVON
 Grande location indépendante mitoyenne à Chaumeçon sans vis à vis dans un ancien corps de ferme avec auvent jardin d'agrément de 300 m². Au rdc : cuisine équipée canapé wc. Chauffage électrique. A l'étage : salle de bains avec wc. A l'Orée du Morvan possibilité de pêche nature randonnées. Nombreux sites de randonnée.

Gîte 482 - CHAUMARD
 Chalet tout bois confortable et douillet directement au bord du lac de Chaumeçon avec coin cuisine convertible 2 pers. 1 ch (1 pers) 1 pers salle de bains wc chauffage électrique terrasse. Possibilité ménage fin de séjour: 40 euros. Chalet typique bourg morvandiau de Blaisy. Dans cet écrin de verdure vous pourrez profiter de la vue sur le lac de Chaumeçon.

Site :
www.securite-sociale.fr

En cliquant sur les catégories proposées en colonne de gauche, l'utilisateur peut dynamiquement affiner sa recherche.

7. Présentation des résultats

La solution AFS offre un **large éventail de possibilités** quant à la présentation des résultats. Au delà du classique "titre, résumé, URL" habituellement utilisé, il est possible de construire des **réponses plus riches** incluant des **images** ou des **informations complémentaires** (date, auteur, ...) issues des métadonnées ou extraites des documents eux-mêmes.

Dans le cas de la recherche dans des données structurées (bases, XML), il est possible d'exploiter n'importe quel champ, tag ou attribut dans la présentation des réponses.

Mise en évidence des mots

AFS présente pour chaque réponse un **extrait du document** ayant justifié la sélection de ce document dans la réponse ainsi que le contexte exact d'apparition des mots de la requête sous la forme dite KWIC (KeyWords In Context), c'est à dire avec **mise en couleur ou en gras des mots** concernés.

Exemple : réponse avec mise en gras pour la requête « Antidot » :

1. La Poste.net choisit Antidot pour son moteur de recherche

JDNet... janvier 2003 La Poste.net choisit **Antidot** pour son moteur de recherche En ... La Poste La Poste.net choisit **Antidot** ... http://solutions.journaldunet.com/ilsontchoisi/030114_ilsontchoisi.shtml

Personnalisation

AFS fournit par défaut les **résultats de recherche au format XML**.

Ceux-ci peuvent être récupérés tels quel afin d'être **intégrés dans une application**, ou alors les réponses peuvent être **formatées** en appliquant la **charte graphique** désirée via un processus de transformation XSLT, nourrir une application **Flash** ou une interface Web **AJAX**.

Il est possible d'utiliser différentes **feuilles de style XSL** selon les objectifs à atteindre. La solution AFS est ainsi nativement **compatible** tant avec le Web classique (production des réponses en **HTML** compatible **XHTML 1.0**) qu'avec les autres types de terminaux (**SMS**, **PDA**, **I-Mode**,...). Il est également aisé de mettre en place une feuille de style spécifique permettant de générer des pages réponses compatibles avec les besoins des **déficients visuels (WAI)**. L'utilisation de feuilles de style **CSS** permet par ailleurs de faire varier les aspects graphiques très facilement.

Géolocalisation

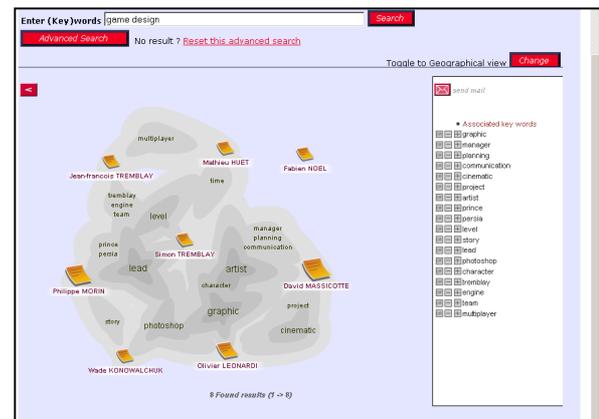
AFS permet de générer des flux de réponse spécifiques destinés à la représentation de l'information sur des systèmes de géolocalisation afin de positionner l'information.

Dans le cas d'Annonces Jaunes par exemple, les résultats sont ventilés sur la carte selon un principe de grille et l'utilisateur à la possibilité de zoomer en cliquant dans la case qui l'intéresse afin de restreindre sa recherche à cette zone et d'affiner la consultation.



Visualisation cartographique

Grâce à son partenariat technologique avec Kartoo, Antidot est à même de proposer des outils de visualisation d'information et de cartographie sémantique.



8. Sécurité

Conçue pour être intégrée dans des **environnements exigeants en terme de sécurité**, AFS intègre un ensemble de mécanisme à même de répondre à des contraintes avancées de **protection des données**, de **sécurité applicative** et de **filtrage des réponses** en fonction des **profils utilisateurs**.

Lors de l'indexation, tous les accès aux données peuvent se faire en mode **authentifié** et être **cryptés**.

Chaque document, chaque enregistrement de base de données, élément de flux XML, champ, tag ou attribut peut être étiqueté séparément lors de l'indexation et ces étiquettes sont ensuite utilisées dynamiquement lors de requêtes afin de valider les **droits d'accès de chaque utilisateur**. Ces droits peuvent par exemple être récupérés dans un annuaire **LDAP**.

Avec AFS, il est possible de faire dépendre les agents qui participent à la recherche, les résultats retournés et les éléments de réponse en fonction de l'application qui requête, en fonction du profil utilisateur et des droits dont il dispose.

9. Performance

L'architecture à base d'**agents logiciels** mise en œuvre par AFS offre de très nombreux avantages. Elle permet en effet nativement :

- de **maximiser la pertinence** car chaque agent va exploiter au mieux les spécificités de la source de données qu'il indexe ;
- d'intégrer facilement de **nouvelles sources** sans compromettre l'existant puisque les agents en place ne sont pas modifiés ;
- de garantir la **pérennité** du système grâce à son **évolutivité** et sa **scalabilité** par simple ajout d'agents ou de ressources matérielles ;

- de garantir la **performance** et la **disponibilité** : l'insertion d'un nouvel agent au sein d'un service est possible à chaud, de même que la modification ou la suppression, permettant ainsi l'ajout d'une nouvelle fonctionnalité ou tout autre opération de maintenance à tout moment sans interruption du service.

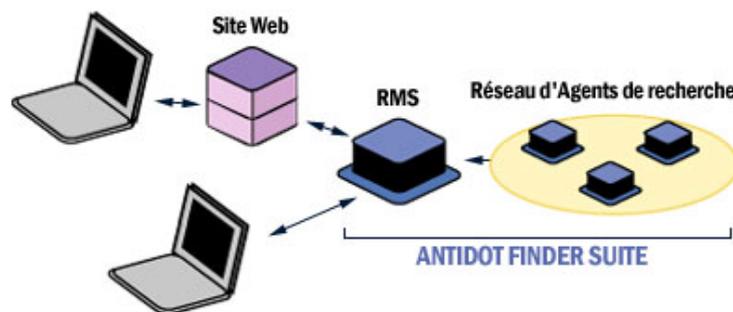
AFS est, de par sa conception, une solution conçue pour la haute disponibilité avec des temps de réponses prédictifs, c'est-à-dire qui ne dépendent ni du volume de données indexées, ni du nombre de requêtes à traiter. Ainsi, quelque soit le nombre de documents et le nombre de requêtes, le temps de réponse moyen du moteur est inférieur à 500 ms.

AFS est une solution totalement extensible puisqu'il suffit d'ajouter des ressources pour que ses capacités augmentent de façon linéaire. Il est possible d'indexer plusieurs centaines de millions de documents et de répondre à plusieurs millions de requêtes par jour. De plus, AFS a été conçue pour fonctionner sur des machines de type PC ou serveur d'entrée de gamme afin de garantir la maîtrise des coûts d'exploitation.

AFS est une solution extrêmement **robuste** qui a été conçue pour fonctionner dans les environnements les plus exigeants en termes de **stabilité**, de continuité de service et de robustesse. Elle est utilisée par les plus grands sites et portails français avec une disponibilité supérieure à 99,99% par an.

10. Intégration

Que ce soit de façon autonome ou au sein d'une chaîne applicative plus complexe, **AFS est extrêmement simple à intégrer** puisqu'il se présente sous forme d'un **Web Service** interrogeable avec les protocoles les plus classiques : HTTP (en GET ou en POST) ou XML-RPC.



Les requêtes des utilisateurs peuvent soit être transmises directement à AFS qui se chargera de générer une réponse complète et exploitable par le programme client, soit transiter par une application intermédiaire (portail web, application métier) qui récupérera les réponses et pourra à son tour appliquer les traitements nécessaires.

Ce deuxième cas est le plus adapté aux environnements ayant des contraintes de sécurité importante puisque ce mécanisme permet non seulement de masquer les paramètres de recherche, mais également de consolider la stratégie de sécurité avec celle de l'application.

AFS propose également des **modules d'intégration** prêts à l'emploi pour certains environnements comme les **CMS** (Content Management System) courants du marché (et en particulier les CMS Open Source comme SPIP, eZ Publish ou Typo3) ou pour des environnements applicatifs tels que **Java** ou **.Net** de Microsoft.

Une solution polymorphe et partageable

Il est possible de **partager le moteur** entre **plusieurs sites** ou **applications** et chacun d'eux aura tout loisir d'intégrer les résultats et d'appliquer la feuille de style qui lui convient. Les critères de pertinence, de tri ou de catégorisation peuvent aussi être fonction du site depuis lequel la requête est soumise (comme accentuer la priorité des documents du site ou alors présenter d'un côté les documents du site local, puis ensuite ceux disponibles sur les autres sites).

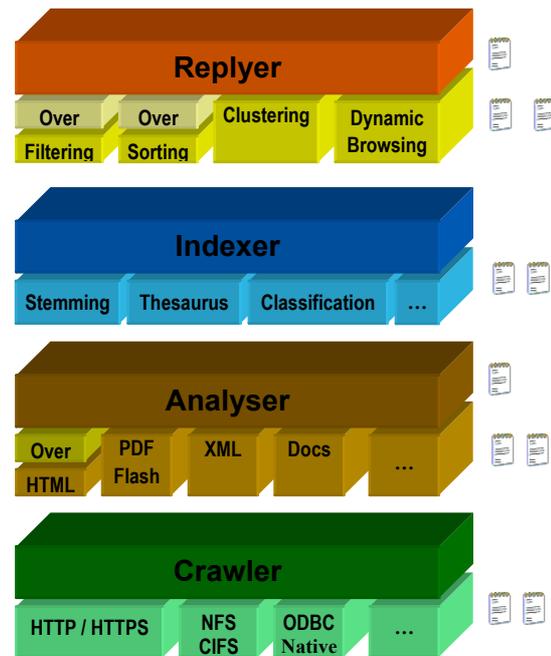
En plus d'être partagée entre plusieurs sites avec **chacun leurs réglages**, la solution AFS peut être utilisée sur un même site avec un comportement contextuel selon l'endroit du site depuis lequel une requête est formulée.

Architecture et évolutivité

AFS est principalement développé en C++, langage objet connu pour sa performance et il bénéficie d'une architecture logicielle très évoluée.

Chaque fonctionnalité est implémentée dans un module configurable par des règles externes, et les différentes couches peuvent être enrichies par un système de **plug-ins** (modules dynamiques).

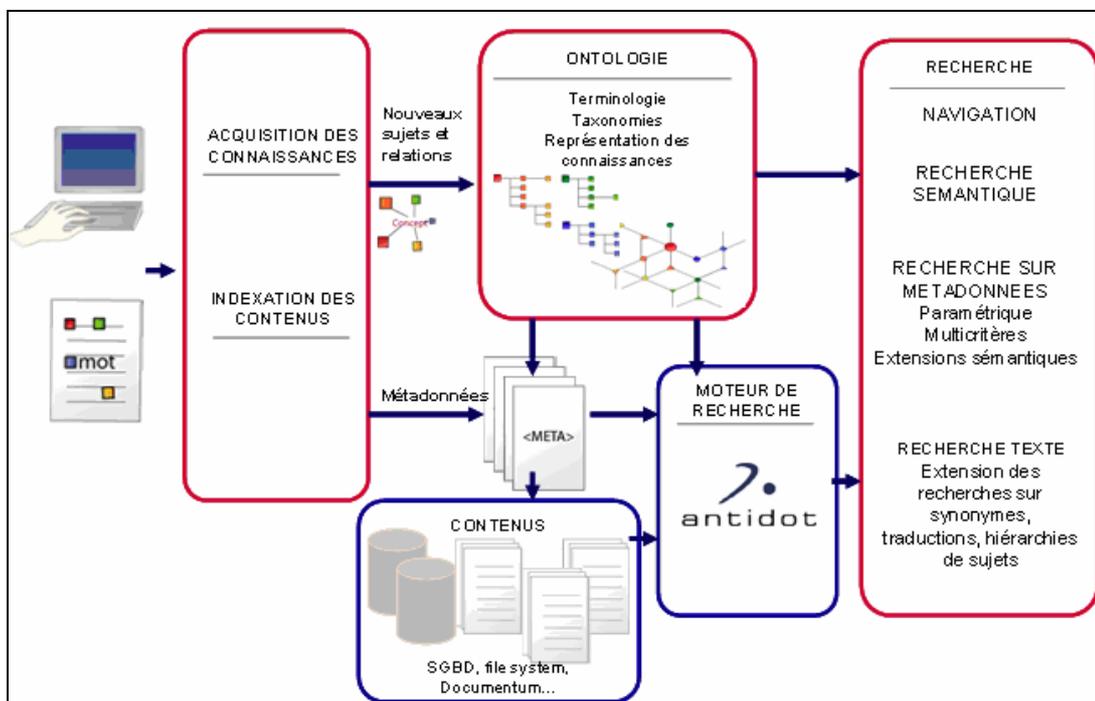
Par ailleurs, chaque classe est modifiable par surcharge. Ces mécanismes permettent de modifier et d'adapter AFS aux cas les plus complexes et les plus invraisemblables.



Analyse de l'information

AFS est conçu pour s'interfacer avec des solutions logicielles d'extraction et d'analyse d'information. Lors de l'indexation AFS peut en effet leur transmettre les données extraites des sources afin de celles-ci soient analysées et que la connaissance extraite soit injectée dans des systèmes spécifiques de gestion de connaissance avec lesquels AFS est capable de s'interfacer.

De telles applications complètes permettent d'associer récupération de l'information, analyse avancée, indexation des données et des connaissances extraites, injection dans des systèmes de gestion de connaissance, mise à jour dynamique d'ontologies.



AFS peut donc facilement s'insérer dans des architectures globales d'analyse d'information comme celles qui peuvent être définies avec UIMA ou Gate.

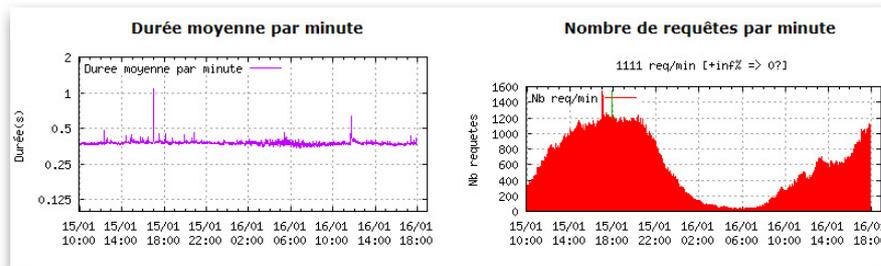
11. Back-office

Au-delà des fonctionnalités principales liées à l'indexation et à la recherche, AFS dispose d'un back-office offrant de nombreux outils et interfaces permettant une exploitation simple et efficace.

Statistiques quantitatives

AFS fournit un ensemble d'indicateurs numériques et de graphes ce qui permet de surveiller son comportement et de visualiser la progression des indicateurs dans le temps :

- nombre de requêtes par minutes,
- temps de réponses,
- nombre de requêtes et de réponses par jour.



Toutes les requêtes traitées par le système sont loggées et archivées avec tous les paramètres et indicateurs associés (agents ayant répondu, temps de réponse, ...). Ces logs sont au format CLF (Common Log Format, le format utilisé par Apache), afin d'être exploitables facilement.

Analyse qualitative

AFS intègre un outil, appelé « Hit-parade », permettant l'**étude des requêtes** c'est-à-dire des **mots clés recherchés** par les utilisateurs. Il détermine un classement quotidien, hebdomadaire et mensuel des mots clés les plus recherchés, des réponses les plus cliquées. Pour chaque mot clé, cet outil donne accès à l'ensemble des requêtes saisies contenant ce mot.

Il permet par ailleurs de déterminer des catégories thématiques de mots clés (par exemple on pourra créer la thématique "automobile" qui contiendra l'ensemble des mots clés correspondants à cette thématique comme "moteur", "freins", ...) et donc de profiler les recherches des utilisateurs de façon à fournir des informations en adéquation avec leurs intérêts.

En plus des requêtes des utilisateurs, cet outil permet d'obtenir des statistiques sur les réponses cliquées par les utilisateurs et d'identifier ainsi les informations qui semblent les plus pertinentes au regard de leurs requêtes.

The screenshot shows the 'Hit parade' interface. At the top, there are navigation tabs: 'Home', 'Statistiques', 'Mots clés', and 'Hit parade'. Below this, there are sub-tabs: 'Hit Parades', 'Les mots', 'Les sites', and 'Les requêtes'. The main content area is divided into two sections. On the left, there is a calendar for August 2004, with the 25th highlighted. Below the calendar, there are options to 'Calculer seulement les 200 premiers' and a dropdown for 'Catégorie à analyser : Global'. On the right, there is a table titled 'les mots les plus recherchés' with columns for 'Mot', 'Occurrences', 'Public', and 'Requetes'. The table lists 15 items, with 'Star academy' at the top.

Mot	Occurrences	Public	Requetes
1 Star academy		Normal	Liste des requetes
2 Jennifer		Indéfini	Liste des requetes
3 academie		Normal	Liste des requetes
4 starac		Normal	Liste des requetes
5 casting		Normal	Liste des requetes
6 sex bomb		Normal	Liste des requetes
7 vidéos		Normal	Liste des requetes
8 jeremy		Normal	Liste des requetes
9 houcine		Normal	Liste des requetes
10 armande altai		Normal	Liste des requetes
11 million		Normal	Liste des requetes
12 ile de la tentation		Normal	Liste des requetes
13 emma daumas		Normal	Liste des requetes
14 qui veut gagner des millions		Normal	Liste des requetes
15 24		Normal	Liste des requetes

Optimisation du référencement

AFS peut-être utilisé de façon automatique pour générer des pages mettant en avant la richesse des contenus, y compris ceux qui ne sont pas accessibles par la navigation, afin de maximiser leur **indexation par les moteurs de recherche Internet** et donc **d'augmenter le référencement** et la visibilité du site auprès des moteurs de recherche Internet (Google, Yahoo!, ...) permettant ainsi d'augmenter sensiblement le trafic.



Agent mots-clés

Le back office AFS donne accès à un ensemble d'interfaces graphiques permettant de configurer et d'administrer simplement les agents.

La copie d'écran suivante illustre l'agent mots-clés qui permet d'associer des réponses choisies par l'administrateur à certains mots clés ou expressions.

Campagnes		Par mots-clés	Par réponses	Filter	Go	Rafraîchir
campagnes (7 enregistrements)						
Campagnes	Réponses	Affichages	Clicks	Etats		
AirFlex	du 24/03/2005 au 25/05/2005	159 / 1000	7 / 400	●	<input type="checkbox"/>	<input type="checkbox"/>
isolation mur etancheite isolation phonique air environnement	Airflex : poussez les murs...avec l'isolant réflecteur ultramine (10 mm) et ultraperformant. Avec ses 10 mm ultra isolants et ultra souples, Airflex permet un gain de pl ...					
Ciel	du 25/03/2005 au 30/04/2005	5 / 500	4 / 4	●	<input type="checkbox"/>	<input type="checkbox"/>
logiciel devis facture	Ciel Multi-Devis du Bâtiment Réalisez simplement vos devis, factures et suivis de chantiers selon les spé ...					
Ecolechezsoi	du 24/03/2005 au 25/05/2005	81	0	●	<input type="checkbox"/>	<input type="checkbox"/>
formation	Ecole Chez Soi : Former Autrement au service du Bâtiment et des Travaux Publics Une formation à distance ou une formation par correspondance dans le Bâtimen ...					
Ecolechezsoi	du 24/03/2005 au 20/04/2005	0	0	●	<input type="checkbox"/>	<input type="checkbox"/>
formation former btp	Ecole Chez Soi Formations 100% btp et immobilier ? Consultez-nous...					
France Air	du 24/03/2005 au 25/05/2005	140 / 100	2 / 200	●	<input type="checkbox"/>	<input type="checkbox"/>
climatisation ventilation air formation	Anticipez l'été avec le tout nouveau mono split Finesse ! Finesse... le choix du confort, de l'élégance et de la discrétion... A décou ...					

Configuration

Le back-office d'AFS inclut les outils nécessaires à la configuration, au suivi et à la maintenance de la solution : déclaration et paramétrage des sources à indexer, gestion des dictionnaires métiers, référentiel projet, suivi des actions effectuées à travers le back-office, ...

CONTACTS

ANTIDOT SAS
AGENCE COMMERCIALE
40 rue Camille Roy
F-69007 LYON
TEL : +33(0) 4 37 28 37 64
FAX : +33(0) 4 42 28 61 03

WEB : WWW.ANTIDOT.NET
MAIL : INFO@ANTIDOT.NET

ANTIDOT SAS
SIÈGE SOCIAL
29, rue Jean Monnet
F-13410 LAMBESC
TEL : +33 (0) 4 42 28 61 26
FAX : +33 (0) 4 42 28 61 03