

Innovation & Technologie

Référence : AFS/WP/TEKI/3.0c

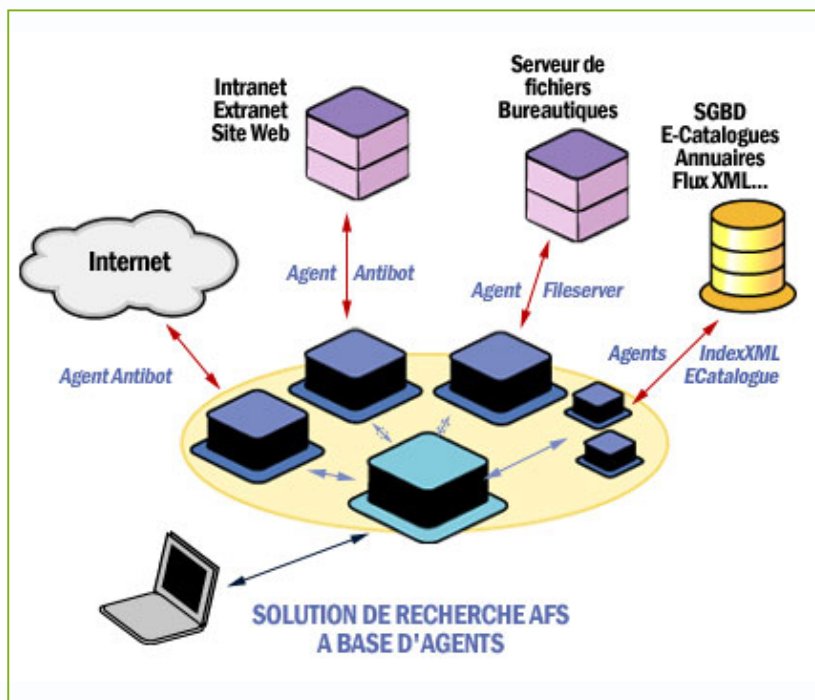
PRINCIPE

Toutes les solutions d'accès à l'information proposées par Antidot reposent sur le même **socle technique innovant** et plusieurs fois primé : **Antidot Finder Suite (AFS)**.

Ce fond technologique, unique de par sa conception, exploite le principe d'**agents¹ logiciels** organisés en réseau.

Ainsi, la mise en œuvre d'une solution d'accès à l'information, moteur de recherche ou autre, consiste à identifier les différentes sources de données puis à lister les contraintes fonctionnelles, organisationnelles et techniques associées afin de déployer et de configurer les agents adaptés à la problématique.

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



Une fois les agents déployés, il suffit de mettre en place le **module central** chargé de recevoir les questions provenant directement des utilisateurs ou d'applications déjà en place.

Ce module se charge de transmettre les questions aux différents agents constituant la solution et de récupérer leurs réponses afin de constituer la **réponse finale** en appliquant les règles de filtrage, de tri, de dédoublement et de catégorisation. C'est également au niveau de ce module que sont précisées les règles de **sécurité** et les **règles métiers** permettant de fournir des réponses adaptées et contextuelles selon les questions, le contenu des réponses et les profils des utilisateurs.

Le cœur d'AFS est un moteur de traitement de données XML qui permet d'agréger des flux XML à la volée en leur appliquant des règles de traitement. Le résultat final est lui-même un flux XML facile à intégrer dans des applications existantes ou à traduire vers d'autres langages (HTML, WAP, ...).

AVANTAGES

Cette approche totalement révolutionnaire introduit une **rupture technologique** par rapport aux solutions classiques basées sur un index unique. L'architecture d'AFS à base d'agents **permet** en effet :

- d'apporter **des réponses plus pertinentes** car chaque agent traite au mieux la source de données à laquelle il est attaché en prenant en compte ses spécificités.
- de garantir la **stabilité**, la **robustesse** et l'**évolutivité** de la solution car elle permet d'ajouter, d'enlever ou de modifier un agent sans toucher et remettre en cause les agents existants.
- de s'adapter aux contraintes d'exploitation (**scalabilité**) puisque par simple ajout d'agents, la solution peut traiter un plus **grand nombre de données** (plusieurs dizaines ou centaines de millions), supporter **plus d'utilisateurs** (des centaines de requêtes par secondes), ou offrir des fonctions de **haute disponibilité**.
- de **pérenniser l'investissement** puisqu'il suffit d'ajouter un agent pour traiter une nouvelle source : à aucun moment la solution ne se trouve dépassée.
- de déployer **des solutions uniques et sur mesure** de façon **simple et rapide** car un très grand nombre d'**agents prêts à l'emploi** et répondant aux besoins les plus courants sont disponibles.

DES AGENTS PRETS A L'EMPLOI

AntiBot

Cet agent est un indexeur nouvelle génération permettant l'analyse et la recherche dans des collections de documents web et bureautiques, sur internet ou dans les intranet. Il détecte automatiquement la langue des documents, les pages miroirs, les doublons, suit automatiquement les liens pour découvrir toutes les pages composant un périmètre donné (les URLs dynamiques sont supportées, et il est possible d'écrire des scripts de parcours pour les sites dynamiques).

Basé sur une analyse syntaxique complète du contenu des documents et des liens entrants et sortants de chaque document, AntiBot répond avec une pertinence qui le place parmi les produits leaders du marché - ceci dans des temps de réponse très courts, même en présence de collections contenant plusieurs dizaines de millions de documents....

Bien d'autres agents permettent d'indexer serveurs de fichiers, email, annuaires, liens sponsorisés, mots clés, aide au référencement, ... ou de s'interfacer avec des systèmes déjà existant (GED, ERP, ...) afin de les questionner et de consolider leurs données offrant ainsi une porte d'entrée unifiée dans l'information.

IndeXML / BD

Un agent générique qui permet d'indexer les données structurées (flux XML, bases de données, ...) en conservant toutes les informations de structure de façon à maximiser la pertinence et la consolidation des données entre elles. En plus des fonctionnalités classiques supportées par tous les agents de la solution AFS (support des thésaurus, la recherche floue, ...), cet agent offre des fonctions de recherche multicritères et une catégorisation fine et contextuelle des réponses.

eCatalogue

Un agent qui permet d'indexer des catalogues de produits avec une prise en compte des prix, des volumes, de la disponibilité, des références et des descriptions de façon à gérer les requêtes multicritères. En plus de la recherche floue et de la suggestion orthographique, cet agent intègre le concept de dictionnaire métier pour une recherche plus souple.

FONDEMENTS TECHNOLOGIQUES

AFS met en œuvre les technologies les plus avancées dans de multiples domaines mathématiques et informatiques :

- **Théorie de la compilation** : grâce à ses **parseurs à responsabilité** (par opposition aux classiques parseurs gloutons sur grammaires régulières), AFS indexe les données dans leur format natif (HTML, PDF, ...) tout en étant résistant aux documents mal-formés. Lors de l'analyse lexicale, AFS travaille sur des arbres décorés qui étiquettent les relations syntaxiques de façon à lever les ambiguïtés sémantiques propres au format. Ces méthodes permettent de gagner un ordre de grandeur de pertinence.
- **Traitement du langage** : pour l'**analyse morphologique**, AFS utilise des **méthodes stochastiques** de type processus markovien, ainsi que des **algorithmes transformationnels** afin de réaliser l'étiquetage morphosyntaxique des documents indexés et d'extraire les concepts signifiants.
- **Calcul matriciel** : recherche de points fixes, calcul de **valeurs singulières** et de **valeurs propres**, ... Ces différentes techniques issues du calcul algébrique sont utilisées pour l'amélioration de la pertinence, pour la classification ou même l'analyse des relations entre documents sur le web (détection des documents de référence, des sites pivots, ...).
- **Compression et encodage** : grâce à des algorithmes spéciaux d'encodage des informations, AFS utilise des représentations particulièrement compactes (5 bits par documents en moyenne), ce qui permet de traiter facilement de très grandes volumétries (plusieurs dizaines ou centaines de millions de documents).
- **Classification, clusterisation et filtrage** de contenu utilisent des outils d'**apprentissage non supervisés** (non supervised learning algorithms) de type **bayésien** ou **vectorel**.

Au-delà de ces aspects technologiques, AFS est un produit extrêmement avancé en termes d'**architecture logicielle**. Entièrement écrit en **C++** (plus de **500 000 lignes de code**) et auto-paramétré en XML, AFS est composé de classes regroupées en modules qui s'assemblent simplement et qui peuvent se dériver et s'enrichir par simple héritage ou surcharge.

MOYENS D'ETUDES & RECHERCHE

AFS est issu de 20 années homme de recherche et développement et Antidot consacre une partie importante de ses moyens aux études et à l'innovation afin de conserver son avance technologique et d'assurer ses clients qu'ils disposent toujours d'une solution de recherche intégrant les fonctionnalités les plus abouties. Toutes les activités de R&D sont réalisées sur le site d'Aix-en-Provence.



Décembre 2002, Antidot a été labellisée « **Entreprise Innovante** » par l'Anvar (Agence Nationale pour la Valorisation de la Recherche) pour l'originalité et les performances de sa technologie à base d'agents logiciels et bénéficie également de son soutien financier pour sa politique d'innovation.

La technologie AFS a également été nommée par un jury d'experts internationaux parmi 437 candidatures issues de 27 pays européens à l'**IST-Prize 2003**.

Le Prix Européen IST est la distinction la plus prestigieuse pour les produits ou services innovants dans le domaine des Technologies de la Société de l'Information européenne. Ce prix est organisé par Euro-CASE (European Council of Applied Sciences and Engineering).

