**Analyse du Fonctionnement :**

Scrutari est un moteur de recherche destiné à effectuer des recherches sur un nombre précis de sites (les sites « abonnés »). Sa particularité est de baser sa recherche sur les méta-données transmises par les sites abonnés et non sur les documents des site eux-mêmes (pages HTML, fichier PDF, etc.), contrairement aux moteurs de recherche classiques. L'objectif est de sacrifier l'exhaustivité au bénéfice de la pertinence. En particulier, Scrutari permet de maitriser l'algorithme de recherche et donc de maitriser l'ordre des résultats (par exemple, pour mettre en avant les documents qui sont estimés les plus importants).

Le système Scrutari est divisé en trois parties bien distinctes :

* **Le format ScrutariData** de description des méta-données d'un site. Ce format fait la part belle aux titres et sous-titres et à l'indexation par des mots-clés. C'est un format défini en XML ; il peut donc être implémenté pour n'importe quel site et quelque soit le langage de programmation (comme le sont les flux RSS).

Un fichier au format ScrutariData est un fichier xml avec comme racine l'élément <base>. Cet élément doit contenir :

* un et un seul élement <base-metadata> (qui contient les détails sur la base),
* un ou plusieurs éléments <corpus> (qui contiennent les informations sur les fiches),
* zéro ou plusieurs éléments <thesaurus> (qui contiennent les informations sur les mots-clés),
* zéro ou plusieurs éléments <indexation-group> (qui contiennent les informations sur les liens entre fiches et mots-clés).

Ce qui donne le format suivant : (l'ordre « traditionnel »)

<?xml version='1.0' encoding='UTF-8'?>

<base>

<base-metadata>...</base-metadata>

<corpus>...</corpus>

<corpus>...</corpus>

<corpus>...</corpus>

...

<thesaurus>...</thesaurus>

<thesaurus>...</thesaurus>

<thesaurus>...</thesaurus>

...

<indexation-group>...</indexation-group>

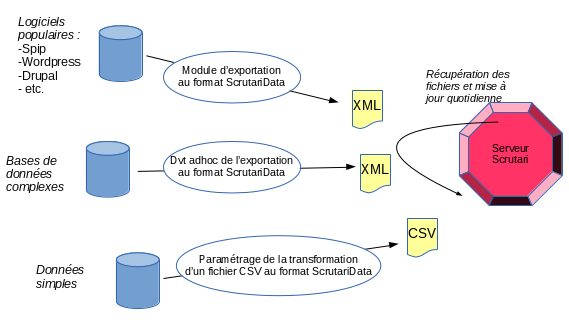
<indexation-group>...</indexation-group>

<indexation-group>...</indexation-group>

...

</base>

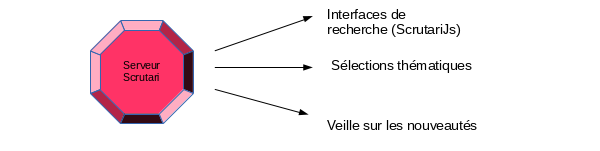
* Le serveur Scrutari qui recueille les méta-données au format ScrutariData transmises par les sites abonnés au serveur et qui traite les requêtes envoyées par les clients du serveur. En plus des recherches libres classiques (i.e. des mots saisis directement par l'utilisateur), le serveur Scrutari propose également des outils d'analyse des méta-données (comparaison entre thésaurus, statistiques, etc.). Le serveur Scrutari est écrit en Java et s'appuie sur le moteur de servlets Tomcat.



La valorisation des données se fait par l’intermédiaire de « clients » (comme celui disponible sur ce site via la fonction « Rercherche Scrutari ») qui soumettent les recherches des internautes au serveur Scrutari ; ce dernier renvoie alors les résultats dans un format brut (en termes techniques, via une API au format JSON) que le client se charge de mettre en forme.

La recherche peut être simple (un seul champ de texte libre) ou avancée avec des filtres et des options de recherche, comme propose le client [ScutariJs](http://scrutarijs.coredem.info/).

D’autres formes de valorisation sont disponibles comme des flux de syndication pour suivre les nouveautés sur les différents sites.



* Les clients de Scrutari qui communiquent avec le serveur Scrutari via des requêtes HTTP classiques en utilisant un format XML spécifique ou un format simplifié en texte brut (le protocole SimpleScrutari). Les clients Scrutari peuvent donc être développés dans n'importe quel langage de programmation. Deux clients sont actuellement disponibles : un client simple en PHP conçu pour s'intégrer facilement dans un site web et un client en XUL conçu comme une extension de Firefox.

On voit que ces trois parties du système Scrutari communiquent entre elles via des formats texte (XML ou format SimpleScrutari) suivant le protocole HTTP. Aussi n'y a-t-il aucune contrainte quant au langage de programmation utilisé du moment que ce langage est capable d'écrire du texte, d'analyser du XML et de communiquer via HTTP. Le choix du langage se fera donc à la convenance du développeur. On pourra ainsi tout aussi bien développer son propre client que se passer du serveur Scrutari pour lire directement les fichiers XML au format ScrutariData fourni par les sites.

Groupement d’université

Groupement de gamers