

Outil d'indexation et de recherche de fichiers

Projet du module "Génie Logiciel"

Année 2012-2013

1 Le but du projet

Ce projet est une étude de cas concrète pour la conduite d'un projet informatique. Son but principal est de vous mettre dans la situation d'un projet assez important, qui doit être effectué en équipe et en respectant les contraintes de spécification et implémentation vues dans le cours de "Génie logiciel".

Par conséquent, l'accent sera mis sur la spécification du projet, sa documentation, son architecture, sa fiabilité, sa facilité d'interfaçage avec d'autres projets ou plate-formes, le travail en groupe. Peu de contraintes seront posées sur la partie implémentation : vous pouvez choisir le langage de programmation pour chaque module.

2 Le sujet

Sur vos disques durs se trouvent des centaines de giga-octets de fichiers contenant des images, du code, des vidéos, courriers, etc. Pour trouver un fichier contenant une information bien précise, par exemple, les notes de cours en "Génie logiciel", vous devez soit être très bien organisés, soit lancer un outil de recherche sur votre disque, par exemple "find" sur les systèmes Linux. Malheureusement, ces outils sont assez lourds si vous avez à balayer tout le disque.

Un système d'indexation et de recherche de fichier¹ pallie cette lourdeur et permet de retrouver quasi immédiatement les fichiers correspondant aux critères recherchés. Ces critères, aussi nommées "méta-données", peuvent être le nom du fichier (ou une partie de ce nom), le contenu du fichier (quand il est lisible), le format du fichier, la date de sa dernière modification, ou toute information significative pour le type de fichier en question. Le résultat de la recherche est affiché très rapidement grâce à une pré-indexation des fichiers du disque qui est effectuée à tout moment par une tâche de fond. Plusieurs systèmes d'indexation existent, le plus connu est sûrement le Google Desktop², qui est une adaptation du moteur de recherche à un système de fichiers local, indépendant du système d'exploitation. Les systèmes propriétaires (Mac OS X ou Windows) disposent de

1. http://fr.wikipedia.org/wiki/Moteur_de_recherche_de_bureau

2. http://fr.wikipedia.org/wiki/Google_Desktop

leur propres outils d'indexation. GNOME a également développé le système d'indexation Tracker³ qui est un logiciel libre.

Les performances d'un tel système peuvent être mesurées selon différents critères : les formats de fichiers dont le contenu est examiné, la charge supplémentaire générée par la tâche de fond de pré-indexation, la complexité des requêtes encodant les critères de recherche, la portabilité sur différents systèmes de fichiers, etc.

En général, un système d'indexation est composé de trois parties :

1. le moteur d'indexation qui est exécuté en tâche de fond,
2. la base de donnée contenant les résultats du moteur d'indexation et
3. le moteur de recherche qui assure l'interface entre la base de données et l'utilisateur.

3 L'organisation

Les équipes Le projet doit être effectué *en groupe de 5 ou 6 étudiants*. Chaque groupe forme une équipe qui doit désigner un chef d'équipe. Le chef d'équipe doit, en plus d'autres tâches, répartir le travail entre les membres de l'équipe afin d'aboutir à un produit logiciel final. Donc, le chef du projet n'est pas "seulement" responsable de distribuer les tâches à ses coéquipiers mais aussi il doit s'assigner des tâches. Attention, chaque membre d'une équipe sera noté selon sa participation au projet, le chef du projet doit veiller à une répartition équitable des tâches.

Les équipes peuvent communiquer pour échanger de l'expérience sur les modules du projet. L'échange de code ou de documentation est interdit. Des séances de TD seront dédiées à un travail inter-équipes afin d'établir des interfaces communes à toutes les équipes, pour les composants développés. Le but étant pour chaque projet d'être capable de fonctionner indépendamment des autres (i.e., avec tous les composants développés par l'équipe) mais *aussi* avec au moins deux composants développés par deux autres groupes différentes.

Le rendu du projet sera composé

- du cahier de charges,
- de la spécification fonctionnelle,
- de la spécification architecturale,
- de la spécification d'interface de chaque module du projet,
- du plan de test et de validation,
- du code implementé et des résultats du plan de test,
- le journal du projet constitué de compte-rendus hebdomadaires.

Au moins une des spécifications (fonctionnelle, architecturale ou d'interface) doit être faite en utilisant les types abstraits de données.

Les documents et le code devront être déposés sur l'espace numérique de l'enseignement. Vous devez vous inscrire sur cet espace numérique et sur sa liste de diffusion.

3. <http://projects.gnome.org/tracker/>, <http://en.wikipedia.org/wiki/MetaTracker>

Dates et échéances

- le 1^{er} mars 2013 : le rendu du cahier de charges.
- le 1^{er} avril 2013 : le rendu des spécifications d'architecture, d'interface et fonctionnelle.
- le 15 avril 2013 : le rendu du plan de test.
- le 13 mai 2013 : le rendu complet du projet.
- La soutenance des projets aura lieu pendant la session d'examens, du 13 au 24 mai 2013.