

Dictionnaire informatique

Introduction

Actuellement nous avons les pieds plongés dans l'ère de l'échange d'information à travers le monde. Internet foisonne d'articles en tout genre, dont nombre d'entre eux sont intéressants... mais souvent inaccessibles par le commun des mortels : illisible ! Trop de jargon !

Qui osera dire ne s'être jamais intéressé aux ornithorynques d'Australie ? Et Pourtant, leur reste bien inconnu pour beaucoup d'entre nous. Je vois dans vos yeux que vous avez maintes fois essayé de consulter des archives sur la physiologie des ornithorynques et que vous avez abandonné, usés et fatigués par votre incompréhension, devant la foule de mots techniques des biologistes avertis.

Nous vous proposons ici de développer un logiciel qui résoudra ce terrible problème.

En effet, la première partie de votre travail sera de vous mettre dans la peau d'un orn... d'un informaticien, côté serveur. Oui, cet informaticien caché derrière les pages HTML de votre surf Internet. Cet informaticien qu'on ne voit pas mais qui donne le ton du Net désormais mondial !!

Ce n'est pas peu dire ! Vous allez être parmi les vétérans de la programmation Web, mais aussi parmi les pionniers de la clarification, de la simplification et de l'ouverture au plus grand nombre du Net !

Quoi ? Oui c'est vrai, je m'emballe. Bon. Le projet donc.

Nous simulerons un fonctionnement client-serveur. Votre programme sera du côté serveur : c'est-à-dire du côté où l'on travaille.

Le client donnera une page HTML (un article par exemple), le serveur, lui, (donc votre programme) se chargera de rajouter des informations pertinentes à cette page.

Plus en détail : votre programme connaîtra les définitions de mots du jargon utilisé dans l'article, et les rajoutera dans l'article.

Un exemple

Voici un article du Monde (<http://www.lemonde.fr/article/0,2320,111061,00.html>)

Microsoft est régulièrement victime de failles dans ses logiciels :

- en 1997, des étudiants américains avaient dévoilé des faiblesses dans la sécurité du navigateur Internet Explorer fonctionnant avec Windows 95 et NT. Ils avaient dissimulé

des raccourcis lançant des programmes derrière des liens hypertextes. Croyant accéder à une page de la Toile, les utilisateurs activaient un logiciel sur leur machine...

Malgré la publication rapide d'une "rustine" par Microsoft, d'autres étudiants découvraient deux autres bogues permettant des manoeuvres similaires, contraignant l'entreprise de Redmond à colmater ces nouvelles brèches;

- en 1999, les boîtes aux lettres électroniques de quarante millions d'utilisateurs de Hotmail, la messagerie gratuite rachetée par Microsoft, ont été librement accessibles par n'importe qui pendant le week-end des 28 et 29 août, à la suite d'une attaque de pirates ayant réussi à désactiver l'ensemble des mots de passe ;

- le 14 avril, le Wall Street Journal révélait que les ingénieurs de Microsoft avaient introduit dans un outil de création de pages Internet fonctionnant avec Windows NT 4.0 un mot de passe ("Les ingénieurs de Netscape sont des mauviettes") donnant accès à des données confidentielles (numéros de cartes bancaires) présents sur des milliers de sites de la Toile;

-début mai, le virus ILOVEYOU semait la panique dans des milliers d'ordinateurs en accédant au carnet d'adresses des victimes et en envoyant de lui-même des messages lui permettant de se reproduire via Internet. Le programme pouvait agir sur n'importe quel fichier du disque dur en le renommant ou en le modifiant.

Vous l'avez compris, ma grand-mère non.

Voici l'article passé à la moulinette de votre futur programme :

Microsoft est régulièrement victime de failles dans ses logiciels(1) :

- en 1997, des étudiants américains avaient dévoilé des faiblesses dans la sécurité du navigateur(2) Internet Explorer fonctionnant avec Windows(3) 95 et NT. Ils avaient dissimulé des raccourcis(4) lançant des programmes(5) derrière des liens hypertextes(6). Croyant accéder à une page de la Toile(7), les utilisateurs activaient un logiciel(1) sur leur machine...

Malgré la publication rapide d'une "rustine" par Microsoft, d'autres étudiants découvraient deux autres bogues(8) permettant des manoeuvres similaires, contraignant l'entreprise de Redmond à colmater ces nouvelles brèches;

- en 1999, les boîtes aux lettres électroniques de quarante millions d'utilisateurs de Hotmail, la messagerie gratuite rachetée par Microsoft, ont été librement accessibles par n'importe qui pendant le week-end des 28 et 29 août, à la suite d'une attaque de pirates ayant réussi à désactiver l'ensemble des mots de passe ;

- le 14 avril, le Wall Street Journal révélait que les ingénieurs de Microsoft avaient introduit dans un outil de création de pages Internet fonctionnant avec Windows(3) NT 4.0 un mot de passe ("Les ingénieurs de Netscape sont des mauviettes(9)") donnant accès à des données confidentielles(numéros de cartes bancaires) présents sur des milliers de sites de la Toile(7);

-début mai, le virus ILOVEYOU semait la panique dans des milliers d'ordinateurs en accédant au carnet d'adresses des victimes et en envoyant de lui-même des messages lui permettant de se reproduire via Internet. Le programme(5) pouvait agir sur n'importe quel fichier du disque dur en le renommant ou en le modifiant.

(1) **logiciel** : il est possible de remplir son ordinateur avec des logiciels : ils permettent d'écrire des textes, de dessiner, d'aller sur internet, ou toute autre fonction que peut remplir l'ordinateur.

(2) **navigateur** : logiciel qui donne accès aux sites (par exemple des pages de texte avec quelques images) présents sur internet.

(3) **windows** : système informatique compliqué qui fait marcher l'ordinateur(enfin...en général)

(4) **raccourci** : raccourci : petit chemin qui permet d'arriver à un logiciel de l'ordinateur directement, en évitant certains endroits.

(5) **programme** : les informaticiens écrivent des programmes pour commander l'ordinateur, lui faire faire certaines tâches. Un logiciel est un programme, il commande l'imprimante par exemple.

(6) **lien hypertexte** : sur une page trouvée sur internet, on peut cliquer sur des

liens hypertexte (certains mots, phrases, images, etc...), ils envoient l'utilisateur sur une autre page.

(7) **toile** : autre nom pour "internet". Vient de l'anglais "web" (www signifie "world wide web" : large toile mondiale).

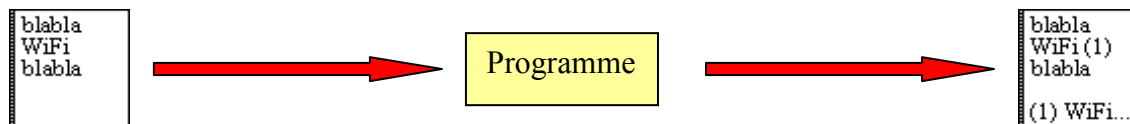
(8) **bogue** : problème présent dans un logiciel. Les bogues (ou "bugs") sont difficiles à repérer, à identifier et parfois à corriger.

(9) **mauviette** : quelqu'un qui se dégonfle.

Et ma grand-mère comprend. Pratique. Surtout pour les grand-mères.

Comprendre

Voici un schéma qui symbolise le fonctionnement de notre serveur.



L'article est modifié dans le serveur, il ressort modifié.

Votre programme est dans le serveur, c'est lui qui fait le boulot.

Nous vous proposons la voie qui nous paraît la plus simple : un fichier en entrée pour l'article, un fichier en sortie pour l'article modifié ;

Et votre programme qui fonctionne de cette manière :

- lancer le programme
- lui fournir l'article, il le renvoie modifié
- lui fournir un second article, il le renvoie modifié
- etc.
- arrêter le programme

Bon je vois que vous commencez à avoir une idée plus claire du projet.

Vous avez sûrement déjà pensé à une partie du programme : celle qui modifie le fichier texte. C'est la partie immergée de l'iceberg.

En effet, il faut maintenant penser à la recherche des mots, puis à la recherche des définitions... et donc le stockage des définitions.

Pour que votre programme soit réaliste, il faut, il faut qu'il soit rapide et pas trop lourd (pas d'utilisation intensive de la mémoire).

Nous vous proposons ici une structure de données qui répond assez bien à cette demande. Comme pour la technologie des bases de données, elle met en œuvre un index, qui va être utilisé pour les recherches. Cet index est construit de façon optimale, pour répondre de façon rapide aux requêtes. Cet index référence les données, sans les charger en mémoire.

Pour être plus précis, la structure de l'index stockera de façon particulière les mots, mais les définitions seront, elles, sur le disque dur, et non en mémoire (les définitions seront stockées dans des fichiers).

L'index est un arbre dont chaque nœud possède 26 fils, pour les 26 lettres de l'alphabet (non pas un arbre binaire, mais un arbre 26-aire).

Exemple :

Si l'on possède la définition de "amour", "banc" et "bécotte", on a l'arbre suivant :

