

Le projet Frutch, moteur de recherche francophone "Open Source"

[Retour au sommaire de la lettre](#)

Jérôme Charron, à l'origine de nombreuses initiatives dans le domaine de la recherche d'information sur le Web francophone depuis de nombreuses années, a lancé un projet baptisé "Frutch", basé sur la mise en place d'un moteur de recherche "Open Source" pour le monde de la francophonie. Il nous en dit plus, dans cet interview, sur ce projet...

Déjà initiateur de la liste de discussion Motrech en 1998 (<http://motrech.free.fr/>), puis d'un blog (<http://motrech.blogspot.com/>) et d'un Wiki (<http://frutch.free.fr/wikini>) sur le sujet, auteur d'une thèse ayant pour sujet "Méthodes et outils d'exploration multilingue sur internet en vue d'une veille technologique sur un domaine restreint" (http://www.chez.com/jcharron/jeronimo/sujet_these.html), Jérôme Charron, a lancé le projet Frutch (<http://frutch.free.fr/>) qui a pour ambition de constituer une alternative transparente aux moteurs de recherche commerciaux actuels (Google, Yahoo, MSN...). Jérôme a bien voulu répondre à nos questions à propos de ce projet ambitieux. Merci à lui...



- Jérôme Charron, bonjour. Pouvez-vous vous présenter en quelques mots à nos lecteurs ?

En fait, je suis tombé très tôt dans l'informatique, puis durant mes études, j'ai rapidement croisé le chemin de la recherche documentaire. Et par la suite, Internet est venu se mêler de tout ça. C'est donc tout naturellement que je me suis passionné pour les moteurs de recherche. Pendant ma thèse, concernant la recherche multilingue sur Internet, j'ai créé la liste motrech. Après une période de presque 5 ans où je souhaitais m'éloigner de ce domaine, j'ai replongé de plus belle dans les méandres des moteurs à travers la liste motrech, le blog motrech, et Frutch.

- En quoi consiste le projet Frutch ?

Il y a souvent un malentendu à ce sujet, et je souhaite donc le dissiper tout de suite : Frutch n'est pas pour le moment une initiative visant à mettre en place un moteur de recherche en ligne (même si cela est notre but ultime, nous n'en avons pour le moment pas la capacité). Ainsi, à la base, le projet Frutch consiste à tisser un lien entre les francophones s'intéressant au moteur open source Nutch. C'est une approche très pragmatique : il est souvent beaucoup plus facile de débiter sur un sujet ou d'échanger des propos complexes dans sa langue maternelle. Frutch pourrait à terme devenir une sorte de représentant francophone de Nutch. Mais comme je le disais précédemment, en fonction des opportunités de partenariat ou de financement qui se présenteraient, Frutch pourrait très bien se décliner en deux instances, comme c'est de plus en plus le cas dans le monde de l'open source : une instance communautaire (.org), ainsi qu'une instance commerciale (.com).

- Quel est le but de Frutch ? Proposer un moteur de recherche alternatif aux Google, Yahoo! et MSN ou proposer une technologie "disponible" pour des moteurs qui voudraient se monter ?

Le but premier de Frutch est donc de regrouper les compétences francophones en matière de recherche d'information afin de tenter de bousculer un peu l'ordre établi aujourd'hui sur le marché des moteurs de recherche.

En fait, les deux approches que vous mentionnez cohabitent. Frutch est une initiative très jeune, et il est donc à la fois intéressant et inévitable me semble-t-il de contribuer activement au développement de Nutch sous forme de code et de documentation.

Cela nous permet de rentrer au cœur de Nutch, de le tester, et de bien comprendre son fonctionnement, son déploiement et son administration, ainsi que d'orienter son évolution. Dans un second temps, si des opportunités de financement ou de partenariat se présentent nous mettrons certainement en place un moteur de recherche alternatif (vous remarquerez certainement que le nom de domaine frutch.com a déjà été réservé).

- Qui participe à ce projet actuellement ?

Nous ne sommes aujourd'hui qu'une petite dizaine d'inscrits sur la liste Frutch et les contributeurs au wiki Frutch se comptent sur les doigts d'une main. Mais notre groupe de travail est encore très jeune (créé en Février) et encore peu connu.

Nous sommes composé de développeurs ayant une expérience significative dans le domaine des moteurs de recherche, de curieux ayant fait quelques tests de Nutch, mais aussi de professionnels ayant travaillé ou travaillant encore dans des moteurs commerciaux.

- Pouvez-vous nous en dire plus sur Nutch ?

Nutch est une initiative de Doug Cutting (créateur de la librairie d'indexation/recherche Lucene). Il partage l'approche, que j'essaye (trop) souvent de soutenir, consistant à dire que le moment est venu de proposer une solution alternative, transparente, libre et gratuite aux outils commerciaux. Frutch (et donc Nutch) est à Google ce que Linux est à Windows : une alternative gratuite et ouverte aux solutions commerciales.

- Où en êtes-vous actuellement sur le projet Frutch ?

Nous n'intervenons aujourd'hui que dans la phase développement open source, c'est à dire en contribuant à Nutch: participation sur les listes de discussion de Nutch, don de code à Nutch (j'ai par exemple récemment donné un module de détection de type de document, et je travaille actuellement sur le plugin de détection de langue), développement de notre base de connaissance francophone sur le wiki de Frutch.

- De quoi avez-vous besoin aujourd'hui pour faire avancer au mieux le projet ?

De passionnés prêts à consacrer du temps au développement, à la documentation, aux tests de Nutch. Si nous voulons pouvoir lancer la deuxième grande phase de Frutch, et pouvoir proposer un moteur en ligne, nous avons essentiellement besoin de financements et/ou de matériels.

- Un moteur "open source" a-t-il sa chance aujourd'hui face aux "mastodontes" américains ?

Il me semble que oui ! Le moteur en lui même (au sens du service) est nécessairement plus ou moins commercial puisqu'il faut financer les machines, la bande passante, l'électricité, la maintenance. Ce n'est cependant pas le côté commercial ou financier qu'il faut retenir, mais la finalité : contrairement à Google, Yahoo! ou MSN dont la finalité est le bénéfice, la finalité de ce moteur serait d'apporter un service (le seul objectif commercial est donc de s'autofinancer, et non pas de dégager des bénéfices).

Ce moteur a donc tout intérêt à utiliser des technologies open source, afin de réduire ses coûts, d'augmenter la visibilité et la transparence de ses technologies, et de bénéficier de la forte dynamique des communautés open source pour faire évoluer ses technologies.

- Quelles sont les prochaine étape prévues ? Quelles sont les grandes échéances attendues dans un futur proche ?

Premièrement l'angle open source qui consiste à travers une communauté francophone à participer à l'effort collaboratif de Nutch. Sous cet angle, les étapes sont simples : parvenir à franciser au plus vite l'intégralité de la documentation de Nutch, et continuer à contribuer au code de Nutch. Deuxièmement, l'angle commercial qui sera très dépendant de l'intérêt que peuvent porter de potentiels investisseurs au projet. Cependant, l'échéance principale de ce point de vue consiste à mettre au point un Frutch-mini, concurrent direct du Google-Mini. D'après nos premières estimations, nous pouvons fournir un tel système de type "plug and search" à un prix 4 à 5 fois moins élevé que celui de Google-Mini pour des fonctionnalités équivalentes.

Merci Jérôme et bonne chance pour ce projet !