

Yahoo place le moteur d'indexation Anthelion en Open Source

Si l'actualité de Yahoo est concentrée sur l'avenir de la société, elle vient de prendre une décision originale en plaçant en Open Source, Anthelion, un moteur d'indexation conçu pour l'analyse des données structurées des pages HTML. Il prend en charge différents types d'annotations sémantiques, comme RDFa, Microformats et Microdata. Ces dernières sont toutes compatibles avec la syntaxe pour les données structurées, schema.org, un projet réunissant les moteurs de recherche Google, Yahoo et Bing.

Sur un blog, les équipes de Yahoo explique qu'Anthelion fournit un ensemble de marqueur pour faire des recherches dans des pages spécifiques. Et de citer l'exemple d'une annotation sur un film avec deux attributs qui sont le titre et les acteurs. Le système est une extension prête à l'emploi pour Apache Nutch Crawler (un moteur de recherche Open Source basée sur Lucene). Il peut être exécuté sur une seule machine ou sur un cluster Hadoop.

Une stratégie Open Source en mode pompier

Pour plus d'informations techniques, Yahoo renvoie à [un document de travail](#) réalisé par Peter Mika et Roi Blanco de Yahoo Labs, et Robert Meusel de l'Université allemande de Mannheim. Les auteurs estiment que les technologies d'indexation peuvent apporter un plus grand nombre de réponses pertinentes sur certaines requêtes. Pour être complet, Anthelion est disponible sur le [GitHub de Yahoo](#) et sous licence Apache 2.0.

Anthelion est au centre de beaucoup de services de Yahoo en dehors de son moteur de recherche traditionnel. Mail, Finance, Messenger, Flickr et Tumblr. Le fait que la firme dirigée par Marissa Mayer décide de partager le code source d'Anthelion dans un domaine aussi concurrentiel que la recherche web est assez significatif.

La firme a fait [le choix du « reverse spin off »](#). Afin de se protéger contre une lourde taxation, qui aurait pu atteindre jusqu'à 10 milliards de dollars. Ce ne sont pas les actions Alibaba qui seront cédées à une nouvelle entreprise (Aabaco) mais toutes les activités historiques de la firme. La fin des activités Internet aux dires de certains, qui pousseraient la direction actuelle à placer en Open Source des technologies clés pour continuer leur développement.

A lire aussi :

[Traitement choc de SpringOwl pour Yahoo : 9000 licenciements et départ de Marissa Mayer](#)

[RSSI chez Yahoo, un poste en CDD ?](#)

© Sergej Khackimullin – Fotolia.com