

Aux représentants des médias

COMMUNIQUE DE PRESSE

Un projet européen de gestion de données du web coordonné par l'UniNE

Neuchâtel, le 23 octobre 2012. L'Institut d'informatique de l'Université de Neuchâtel (IIUN) coordonne depuis début octobre un nouveau projet de recherche européen consacré à la gestion de grandes quantités de données du web à l'intention des petites et moyennes entreprises. Maître-assistant à l'IIUN au sein de l'équipe du Prof. Pascal Felber, Etienne Rivière assure la coordination du projet LEADS*, regroupant quatre universités et trois entreprises réparties sur 5 pays. Le projet est doté d'un budget total de 4,25 millions d'euros pour une durée de trois ans. La subvention européenne revenant à l'IIUN s'élève à 900'000 francs.

Chaque jour, le web s'enrichit d'un contenu équivalent à huit fois le catalogue total des bibliothèques des Etats-Unis. Chaque minute, 35 heures de vidéo sont chargées sur YouTube. Toutes ces données publiques constituent le fonds de commerce de grandes multinationales comme Google ou Amazon qui tiennent des archives et des collections de toute la planète. Mais extraire et traiter ces informations coûte cher et nécessite de gigantesques capacités de stockage et de calcul, nécessitant des centres de données abritant des milliers de serveurs.

Une entreprise dont la spécialité n'est pas nécessairement l'informatique, mais qui désire extraire du web des tendances et des opinions pour un produit récemment mis sur le marché, n'a pas les moyens de le faire elle-même. Le projet LEADS répond à ce besoin : d'importantes quantités de données pourront être compilées grâce au partage des collections, des stockages et des requêtes de nombreux utilisateurs. Cette approche s'inscrit dans l'esprit de l'informatique dématérialisée, ou « cloud computing » en anglais.

Une des forces du concept réside dans sa capacité à assurer la confidentialité des données privées. Encryptées par une clé qui n'est pas connue de l'infrastructure LEADS, ces données pourront être soumises à des traitements et requêtes, qui sont eux-mêmes encryptés. « Cet aspect est particulièrement novateur, car il met en œuvre des techniques de cryptage qui n'ont été proposées que récemment ou qui seront développées au sein du projet. Ceci concerne typiquement des comparaisons, des tris, ou des extractions, qui constituent l'essence des opérations de traitement de grandes masses de données », indique Etienne Rivière.

L'autre atout de LEADS est de favoriser la fédération d'un grand nombre de micro centres de calcul, distribués géographiquement (et appelés « micro-clouds »), de façon à ce que les opérations se déroulent dans des lieux proches des clients, ou de la source de données. « Effectuer un traitement sur la partie francophone du Web, mais en stocker les résultats au Brésil, n'a aucun sens, illustre le coordinateur du projet. Ceci générerait d'énormes flux de données transatlantiques et très peu de connexions depuis l'Amérique latine. Situer ce même stockage dans une région à forte densité francophone serait nettement plus avantageux. » A l'UniNE, c'est précisément la prise en compte de la localité des accès et des traitements que les chercheurs de l'équipe du Prof. Felber vont étudier. Ils utiliseront pour leurs expérimentations des serveurs offerts ce printemps à l'IIUN par la société Yahoo!, en association avec les infrastructures des autres membres du consortium.

Contact :

Dr Etienne Rivière, Institut d'informatique, coordinateur du projet LEADS
Tél. +41 32 718 2729 ; etienne.riviere@unine.ch