

Sujet de master recherche « Architectures logicielles distribuées » 2006–2007

Filtrage d'information à l'aide de profils utilisateurs

Encadrant principal : Guillaume RASCHIA
courriel : Guillaume.Raschia@univ-nantes.fr
tél. : 02 51 12 58 37

Objectif du stage

Le stage s'inscrit dans le contexte d'une Action Concertée Incitative (ACI) *Masses de Données* intitulée « Accès Personnalisé à des Masses de Données ».

Le projet APMD [1] se situe au cœur de la problématique générale d'accès efficace à de grands volumes d'informations. Ce projet fait suite aux deux Actions Spécifiques (AS) CNRS, sur la personnalisation de l'information (AS98) et sur le passage à l'échelle dans les systèmes de recherche d'information (AS91). Ces deux actions, menées durant l'année 2003 par deux groupes de réflexion différents, ont convergé vers une même conclusion : une des principales pistes de recherche favorisant le passage à l'échelle des systèmes de recherche d'information est l'exploitation généralisée de préférences définies dans les profils des utilisateurs. Dans cette perspective, l'objectif principal du projet APMD est de promouvoir l'usage de profils utilisateurs dans l'évaluation de requêtes en présence de grands volumes de données.

Plus spécifiquement, nous souhaitons étudier certaines options pour l'expression des profils utilisateurs et leur impact sur le filtrage de flux de données [2]. Nous nous plaçons dans le contexte général de la « recharge de données » pour les appareils mobiles, telle que définie par M. Cherniack, M. J. Franklin et S. Zdonik [3] et tentons de répondre partiellement au problème exposé en ces termes : “*a key challenge for data recharging [...] is the development of a suitable language and processing strategy for [...] highly expressive user profiles.*”

Nous nous intéressons plus particulièrement à la formulation de profils utilisateurs à partir de caractérisations linguistiques floues [4, 5, 6] dont le principal objet est d'offrir une interface symbolico-numérique adaptée à la fois à l'utilisateur final et au traitement automatique. Les prédicats flous qu'il est alors possible de construire à partir des profils sont utilisés pour filtrer l'information à la manière des systèmes de dissémination sélective, dont le paradigme le plus représentatif est la publication/souscription (*publish/subscribe*) [7]. Enfin, le filtrage de données réalisé à partir de profils n'est pas étranger au problème des requêtes continues, largement couvert dans la littérature sur les bases de données [8]. En effet, chacune des deux approches invoque le contexte des flux de données continus et les langages déclaratifs pour le traitement de requêtes continues tels que CQL [9] ou NiagaraCQ [10] permettent de formuler des requêtes qui opèrent un filtrage sur ces flux. Il est donc fondamental de positionner le travail de ce stage dans la double perspective de la personnalisation de l'information et du traitement des requêtes continues.

Travail à réaliser

Le travail attendu est, en premier lieu, une étude comparative des langages de requêtes continues, notamment du point de vue de leur pouvoir d'expression. L'adéquation de ces langages à la représentation de profils utilisateurs sera également abordée, dans le contexte ouvert de la recharge de données.

Une proposition pour introduire les prédicats flous (à base de caractérisations linguistiques) dans les langages de requêtes continues est souhaitée ; les prédicats flous étant alors une traduction directe dans ces langages des centres d'intérêts et des préférences de l'utilisateur.

Références

- [1] Aci masses de données, accès personnalisé à des masses de données (apmd). <http://apmd.prism.uvsq.fr/>.
- [2] Golab L. and Ozsu M. T. Issues in data stream management. *SIGMOD Record*, 32(2) :5–14, June 2003.
- [3] M. Cherniack, M. J. Franklin, and S. Zdonik. Expressing user profiles for data recharging. *IEEE Personal Communications : Special Issue on Pervasive Computing*, 8(4) :32–38, August 2001.
- [4] L.A. Zadeh. Concept of a linguistic variable and its application to approximate reasoning-I. *Information Systems*, 8 :199–249, 1975.
- [5] L.A. Zadeh. Concept of a linguistic variable and its application to approximate reasoning-II. *Information Systems*, 8 :301–357, 1975.
- [6] L.A. Zadeh. Concept of a linguistic variable and its application to approximate reasoning-III. *Information Systems*, 9 :43–80, 1976.
- [7] T. Eugster, P. A. Felber, R. Guerraoui, and A.-M. Kermarrec. The many faces of publish/subscribe. *ACM Computing Surveys*, 35(2) :114–131, 2003.
- [8] Shivnath Babu and Jennifer Widom. Continuous queries over data streams. *SIGMOD Record*, 30(3), September 2001.
- [9] Arvind Arasu, Shivnath Babu, and Jennifer Widom. The CQL continuous query language : Semantic foundations and query execution. Technical Report 2003-67, Stanford University, October 2003.
- [10] Jianjun Chen, David J. DeWitt, Feng Tian, and Yuan Wang. NiagaraCQ : a scalable continuous query system for Internet databases. pages 379–390, 2000.