

Organisation de log de requêtes OLAP sous forme de site web

Sonia Colas, Patrick Marcel, Elsa Negre

Université François Rabelais Tours, Laboratoire d'Informatique, France
{prénom.nom}@univ-tours.fr

Résumé. Vu comme une simple collection de requêtes, un log de requêtes d'un serveur OLAP est une structure peu exploitable. Dans cet article, nous proposons d'organiser un log de requêtes d'un serveur OLAP sous la forme d'un site web. Cela a plusieurs avantages, comme la compréhension rapide de ce qui a été fait lors des sessions d'analyse précédentes ou comme l'aide aux futures sessions d'analyse. La technique utilisée a fait ses preuves pour la réorganisation de sites web. Le log est transformé en un graphe pondéré de requêtes, les poids reflétant une similarité entre requêtes. Cette similarité est calculée en tenant compte à la fois des faits interrogés et de la présence ou non des deux requêtes au sein de la même session. Ces poids sont ensuite ajustés par une heuristique utilisant des fourmis artificielles pour trouver un équilibre entre ces deux facteurs. Enfin le log est présenté sous la forme d'un site web dont la structure est l'arbre couvrant minimal de ce graphe. Cette technique est implémentée en java pour le traitement de logs de requêtes MDX. Des expériences ont été conduites sur des logs synthétiques pour tester la qualité des solutions obtenues selon des critères utilisés dans le domaine du web et des interfaces homme-machine.

1 Introduction

Le travail de Khoussainova et al. (2009) a récemment mis en avant une lacune importante des SGBD actuels, à savoir le besoin d'outil de gestion de requêtes. De tels outils permettraient notamment de faciliter l'exploration et l'analyse de gros volumes de données dans un environnement multi-utilisateurs, par exemple en organisant les logs de requêtes pour mettre rapidement en évidence les analyses déjà conduites, de manière à aider les explorations ultérieures.

Le présent travail constitue un pas dans cette direction, dans le contexte de l'analyse OLAP d'entrepôts de données. Nous proposons ainsi d'organiser les logs de requêtes d'un serveur OLAP sous la forme de sites web. Cette organisation donne la possibilité à l'utilisateur d'explorer un ensemble de requêtes pour parcourir un cube de données, et ce grâce au moyen de navigation très répandu qu'est le site web. Afin de proposer un site web répondant aux critères d'utilisabilité (Nogier (2005)), la structure du site correspond à un arbre de requêtes plutôt qu'à un graphe. Nous adaptons une technique qui a fait ses preuves pour la réorganisation de sites web (Colas et al. (2008)). Cette adaptation repose sur une similarité entre requêtes basée sur le TF/IDF. Plus précisément, le log est d'abord vu comme un graphe pondéré de requêtes, les poids reflétant la similarité entre requêtes. Cette similarité est calculée en tenant compte à la fois des faits interrogés et du fait que les requêtes se trouvent au sein de la même session. Ces poids sont ensuite ajustés par une heuristique utilisant des fourmis artificielles pour trouver un équilibre