

Roaring bitmap : nouveau modèle de compression bitmap

Samy Chambi*, Daniel Lemire**, Robert Godin*, Owen Kaser***

*Département d'informatique, UQAM, 201, av. Président-Kennedy
Montréal, QC, H2X 3Y7 Canada
chambi.samy@gmail.com
godin.robert@uqam.ca

**LICEF, Université du Québec, 5800 Saint-Denis, Montréal, QC, H2S 3L5 Canada
lemire@gmail.com

***Computer Science and Applied Statistics, UNB Saint John,
Saint John, NB, E2L 4L5 Canada
o.kaser@computer.org

Résumé. Les index bitmap sont très utilisés dans les entrepôts de données et moteurs de recherche. Leur capacité à exécuter efficacement des opérations binaires entre bitmaps améliore significativement les temps de réponse des requêtes. Cependant, sur des attributs de hautes cardinalités, ils consomment un espace mémoire important. Ainsi, plusieurs techniques de compression bitmap ont été introduites pour réduire l'espace mémoire occupé par ces index, et accélérer leurs temps de traitement. Ce papier introduit un nouveau modèle de compression bitmap, appelé *Roaring bitmap*. Une comparaison expérimentale, sur des données réelles et synthétiques, avec deux autres solutions de compression bitmap connues dans la littérature : WAH (*Word Aligned Hybrid compression scheme*) et Concise (*Compressed 'n' Composable integer Set*), a montré que *Roaring bitmap* n'utilise que $\approx 25\%$ d'espace mémoire comparé à WAH et $\approx 50\%$ par rapport à Concise, tout en accélérant significativement les temps de calcul des opérations logiques entre bitmaps (jusqu'à 1100 fois pour les intersections).

1 Introduction

Le volume croissant des ensembles de données scientifiques et commerciales pousse les chercheurs à adopter des techniques d'indexation efficaces, permettant d'extraire rapidement des informations intéressantes. Les index bitmap sont parmi les types d'index les plus communément utilisés (O'Neil, 1987; Su et al., 2013). Plusieurs travaux proposent des solutions faisant recours à ces index pour améliorer les performances des entrepôts de données (Bouchakri et Bellatreche, 2011; Bellatreche et al., 2007; Boukhalfa et al., 2010; Aouiche et al., 2005; Stockinger et Wu, 2008). En effet, leur capacité à exécuter efficacement des opérations logiques entre bitmaps accélère considérablement les temps de réponse des requêtes OLAP multidimensionnelles.