

# Nouveaux modèles d'index bitmap compressés à 64 bits

Samy Chambi\*, Daniel Lemire\*\*, Robert Godin\*

\*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

**Résumé.** Les index bitmap sont très utilisés dans les entrepôts de données et moteurs de recherche pour accélérer les requêtes d'interrogation. Leurs principaux avantages résident en leur forme compacte et leur capacité à tirer profit du traitement parallèle de bits dans les CPU (*bit-level parallelism*). Dans l'ère actuelle du *Big Data*, les collections de données deviennent de plus en plus volumineuses. Les bibliothèques d'index bitmap compressés introduites à ce jour, telles que : *Roaring bitmap*, *WAH* ou *Concise*, ne supportent que des bitmaps d'au plus  $2^{32} \approx 4$  milliards d'entrées et sont souvent impraticables sur de telles masses de données. Ce travail propose trois nouveaux modèles d'index bitmap compressés supportant jusqu'à  $2^{64}$  entrées. Des expériences comparant les performances des nouveaux modèles avec celles de la solution *OpenBitSet* utilisée par le moteur de recherche Apache Lucene, ainsi que d'autres collections Java ont montré que les approches proposées ont permis de calculer des opérations logiques jusqu'à  $\approx 6$  millions de fois et jusqu'à  $\approx$  soixante-trois milles fois plus vite qu'*OpenBitSet* et les structures Java, respectivement, tout en consommant beaucoup moins d'espace mémoire.

## 1 Introduction

Les quantités d'informations générées de nos jours ne cessent de croître à une vitesse phénoménale (Richard Benjamins, 2014). Pour indexer de telles masses de données, la majorité des bibliothèques de représentation d'index bitmap proposées à ce jour s'avèrent impraticables dans de telles situations, car elles ne peuvent être appliquées que sur des ensembles de données ne dépassant pas les  $2^{32} \approx 4$  milliards d'entrées, nous citons comme exemple, la bibliothèque ConciseSet 2.2 (Colantonio, 2010) implémentant les modèles de compression bitmap *WAH* (Wu et al., 2006) et *Concise* (Colantonio et Di Pietro, 2010) en 32 bits, la bibliothèque JavaE-WAH (Lemire et Kaser, 2010) implémentant le modèle de la technique de compression bitmap à 32 bits *EWAH* (Lemire et al., 2010), et la bibliothèque *RoaringBitmap* (Roaring's team, 2014) mettant en œuvre le modèle de la solution de compression bitmap précédemment proposée : *Roaring bitmap* (Chambi et al., 2014, 2016). Les ingénieurs du moteur de recherche Apache