Review Paper: “Mining Frequent Patterns without Candidate Generation”

Author: Jiawei Han, Jian Pei, Yiwen Yin

*Mining frequent patterns* pada database sering diteliti dalam penilitian *Data Mining.* Penelitian sebelumnya seringkali menggunakan pendekatan Apriori, tetapi pendekatan apriori yang menggunakan *Candidate Set Generation* masih terlalu mahal. Dalam Penelitian ini, diberikan ide baru dengan struktur pohon *frequent pattern (FP-tree)*, dan mengembangkan metode Mining berbasis *FP-tree*, yaitu FP-Growth.

Langkah-langkah yang dipakai untuk meningkatkan efisiensi adalah :

1. Database di kompresi menjadi data struktur yang lebih padat dan kecil , ini dilakukan agar menghindari pemindaian database yang terlalu mahal dan berulang-ulang.
2. Mengadopsi metode pertumbuhan fragment pattern agar tidak menghasilkan set kandidat dalam jumlah yang banyak dan juga mahal.
3. Metode berbasis partisi dan *Divide and Conquer* digunakan untuk memecah tugas mining menjadi tugas yang lebih kecil agar mengurangi ruang pencarian.

Percobaan performa yang kami lakukan menunjukkan metode FP-growth adalah metode yang effisien dan dengan mudah diperluas untuk memining frequent patterns baik panjang maupun pendek dalam database yang besar, mengalahkan algorithma berbasis *Candidate pattern generation* lain yang ada saat ini.

Beberapa saran yang dibuat :

1. Saat database berukuran besar dan tidak memungkinkan untuk membuat memori utama berdasarkan FP-tree, salah satu alternatif nya adalah dengan mempartisi database tersebut menjadi beberapa bagian lalu membuat fp-tree dan memining di setiap bagian tersebut. Alternatif lainnya yaitu dengan membuat Disk-resident FP-tree.
2. Saat menambahkan transaksi baru kedalam database selama batas yang dibuat tidak berubah menjadi batas yang tidak di inginkan maka pembuatan ulang fp-tree untuk database tidak diperlukan.