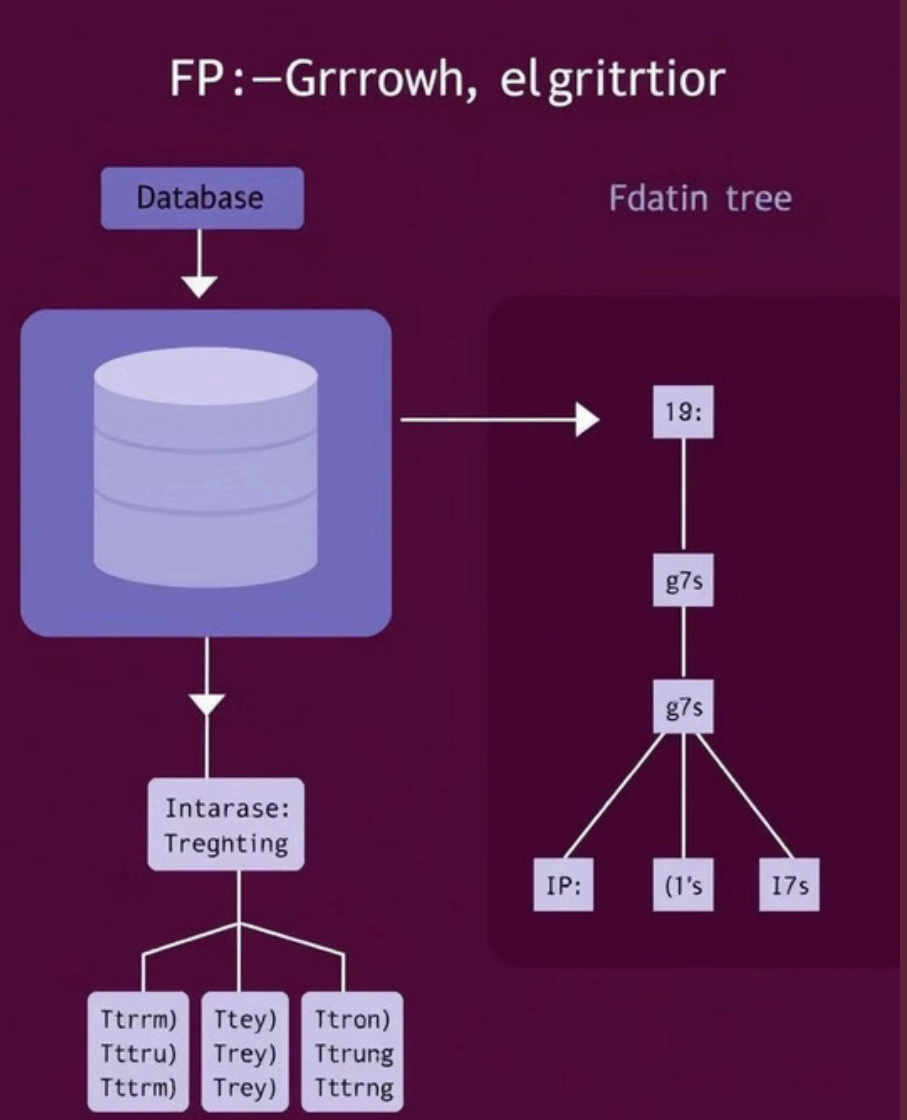
FP-Growth in Data Mining

This presentation introduces the FP-Growth algorithm. It's a key technique for discovering frequent patterns in data. We'll cover its core concepts, step-by-step implementation, and essential data structures. We'll also highlight its advantages over traditional methods.

by Mounish V V



Core Concepts of FP-Growth

FP-Growth uses a Frequent Pattern (FP) Tree. It's a compact data structure for frequent items. This method compresses data, preserving associations. This reduces database scans and speeds up discovery. Let's explore how it works.

FP Tree

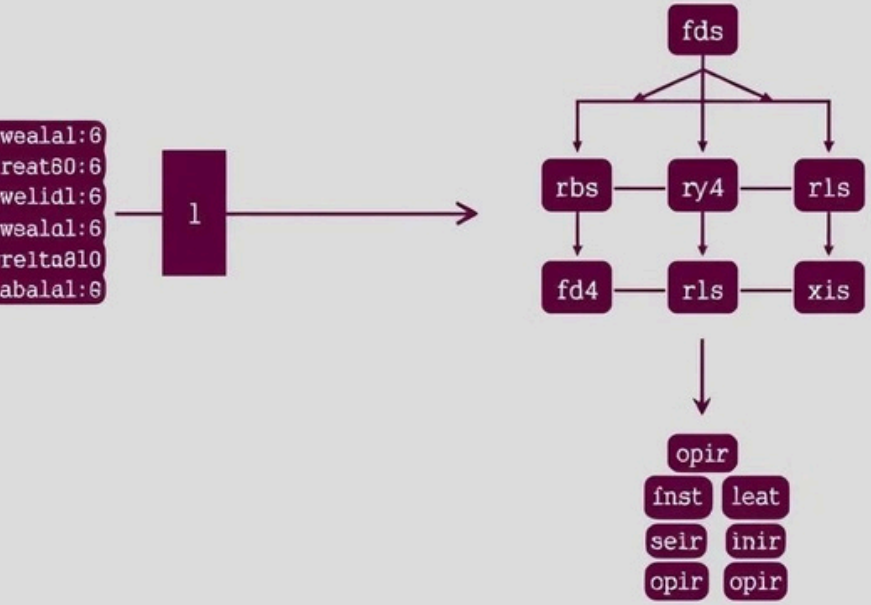
Compactly represents frequent items.

Reduced Scans

Minimizes database scans for speed.

Faster Discovery

Accelerates pattern identification.



FP-Growth Algorithm Steps

The FP-Growth algorithm has three key steps. First, scan the dataset for frequent items. Next, construct the FP-Tree from the data. Finally, mine the FP-Tree recursively. This finds frequent itemsets efficiently.

Scan Dataset

Identify frequent items and counts.

Construct FP-Tree

Create a tree for frequent items.

Mine FP-Tree

Recursively find itemsets.

Data Structures for Implementation

Implementing FP-Growth involves specific data structures. The FP-Tree Node stores item labels and counts. The Header Table links items and support counts. These structures facilitate efficient mining.

FP-Tree Node

- Item name and label.
- Transaction count.
- Parent/child pointers.
- Node links.

Header Table

- Item label.
- Support count.
- Pointer to first node.

FP-Growth Advantages

1 **Efficiency**

Faster than Apriori algorithm. No need to generate candidate sets.

2 **Compact Representation**

FP-Tree provides condensed, relevant data storage.

3 **Reduced Database Scans**

Scans database only twice during the entire process.

Real-World Applications

Market Basket Analysis

Identify products frequently purchased together. Helps in product placement and promotions.

Web Usage Mining

Analyze user navigation patterns to improve website design and recommendations.

Bioinformatics

Discover frequent gene sequences and patterns in biological datasets for research.

Thank You

Thank you for your time. We have covered FP-Growth's core ideas.