Introduction

Many Applications

- Querying purchase patterns for marketing
- Stock market analysis
- Studying meteorological data

What’s needed:

- Expressive query language for finding complex patterns over data streams
- Efficient and scalable implementation: “Query Optimization”

SQL-TS

A query language for finding complex patterns in sequences

- Minimal extension of SQL—only the from clause affected
- A new Query optimization technique based on extensions of the Knuth, Morris & Pratt (KMP) string-search algorithm

Optimized String Search: KMP

After failing, use the information acquired so to:

- back track to shift(j), rather than i+1, and
- only check pattern values after next(j)

shift and next

- Success for first j-1 elements of pattern. Failure for jth element (when input is at i)
- Any shift less than shift(j) is guaranteed to lead to failure, and
- Match elements in the pattern starting at next(j)

Matrix θ and φ

Input tested on pj is now tested against pk

\[
\begin{align*}
\text{pj succeeded}_j & \iff \begin{cases} 
1 & \text{if } p_j \Rightarrow p_k, \text{ and } p_k \neq \text{False} \\
0 & \text{if } p_j \Rightarrow \neg p_k, \\
U & \text{otherwise}
\end{cases} \\
\text{pj failed} & \iff \begin{cases} 
1 & \text{if } \neg p_j \Rightarrow p_k, \text{ and } p_k \neq \text{True} \\
0 & \text{if } \neg p_j \Rightarrow \neg p_k, \\
U & \text{otherwise}
\end{cases}
\end{align*}
\]

Combing values of these lower triangular matrices (j ≥ k)

We derive the values of next(j) and shift(j)

Nested Star Patterns vs. Star Patterns

Handling Nested Star Patterns

State model

Adjacency Matrix

Example

Calculating next and shift

shift(j): \[ \min \{ s \mid \exists \text{ path from } \theta \text{ to } \phi \text{ in } G_j \} \]

next(j): If shift(j)<j-1 then

\[ \text{next}(j) = \min \{ s \mid \exists \text{ path from } \theta \text{ to } \phi \text{ in } G_j \text{ and there is no fork in the path from } \theta \text{ to } \phi \} \]

else next(j)= j-shift(j)

SELECT X.first.magnitude, X.first.time, U.previous.time FROM earthquake

WHERE X.region=“Los Angeles”

AND 1.8<X.magnitude

AND X.magnitude<2.5

AND Y.magnitude<1.1

AND Z.magnitude<1.1

AND U.magnitude<3.5

SELECT X.next.date, X.next.price, S.previous.price FROM quote

WHERE X.name=“Intel”

AND X.price>X.previous.price

AND 30<Y.price

AND Y.price<40

AND Z.price<Z.previous.price

AND T.price>T.previous.price

AND U.price<U.previous.price

AND V.price<25

Experimental result

Pattern of consecutive interval of period of high fluctuation (more than 1% up and down) followed by a period of steady raise.

Improving and Extending OPS*

Buffer Size

Data Compression

Multi Dimensional Data Streams

Pattern Dependent Search