

The background features a complex, abstract design. It includes a network of thin, intersecting lines in shades of red, orange, and grey, creating a web-like structure. Scattered throughout are numerous small, colored dots in green, blue, and orange. On the left side, there is a vertical strip with a grid of small, light-colored squares. The overall aesthetic is technical and data-oriented.

SPADE: Sequential Pattern Mining in Vertical Data Format

Sequential Pattern Mining in Vertical Data Format: The SPADE Algorithm

- A sequence database is mapped to: $\langle \text{SID}, \text{EID} \rangle$
- Grow the subsequences (patterns) one item at a time by Apriori candidate generation

| SID | Sequence |
|-----|---|
| 1 | $\langle a(\underline{abc})(\underline{ac})d(cf) \rangle$ |
| 2 | $\langle (ad)c(bc)(ae) \rangle$ |
| 3 | $\langle (ef)(\underline{ab})(df)\underline{cb} \rangle$ |
| 4 | $\langle eg(af)cbc \rangle$ |

$\text{min_sup} = 2$

Ref: SPADE (Sequential Pattern Discovery using Equivalent Class)
[M. Zaki 2001]

| SID | EID | Items |
|-----|-----|-------|
| 1 | 1 | a |
| 1 | 2 | abc |
| 1 | 3 | ac |
| 1 | 4 | d |
| 1 | 5 | cf |
| 2 | 1 | ad |
| 2 | 2 | c |
| 2 | 3 | bc |
| 2 | 4 | ae |
| 3 | 1 | ef |
| 3 | 2 | ab |
| 3 | 3 | df |
| 3 | 4 | c |
| 3 | 5 | b |
| 4 | 1 | e |
| 4 | 2 | g |
| 4 | 3 | af |
| 4 | 4 | c |
| 4 | 5 | b |
| 4 | 6 | c |

| a | | b | | ... |
|-----|-----|-----|-----|-----|
| SID | EID | SID | EID | ... |
| 1 | 1 | 1 | 2 | |
| 1 | 2 | 2 | 3 | |
| 1 | 3 | 3 | 2 | |
| 2 | 1 | 3 | 5 | |
| 2 | 4 | 4 | 5 | |
| 3 | 2 | | | |
| 4 | 3 | | | |

if $\text{SID}_a = \text{SID}_b$ & $\text{EID}_a < \text{EID}_b$

| ab | | | ba | | | ... |
|-----|---------|--------|-----|---------|--------|-----|
| SID | EID (a) | EID(b) | SID | EID (b) | EID(a) | ... |
| 1 | 1 | 2 | 1 | 2 | 3 | |
| 2 | 1 | 3 | 2 | 3 | 4 | |
| 3 | 2 | 5 | | | | |
| 4 | 3 | 5 | | | | |

↓ drawing $\text{ab}_{\text{EID}(b)} = \text{ba}_{\text{EID}(b)}$

| aba | | EID(a) | | ... |
|-----|---------|---------|---------|-----|
| SID | EID (a) | EID (b) | EID (a) | ... |
| 1 | 1 | 2 | 3 | |
| 2 | 1 | 3 | 4 | |

↓ then
合并