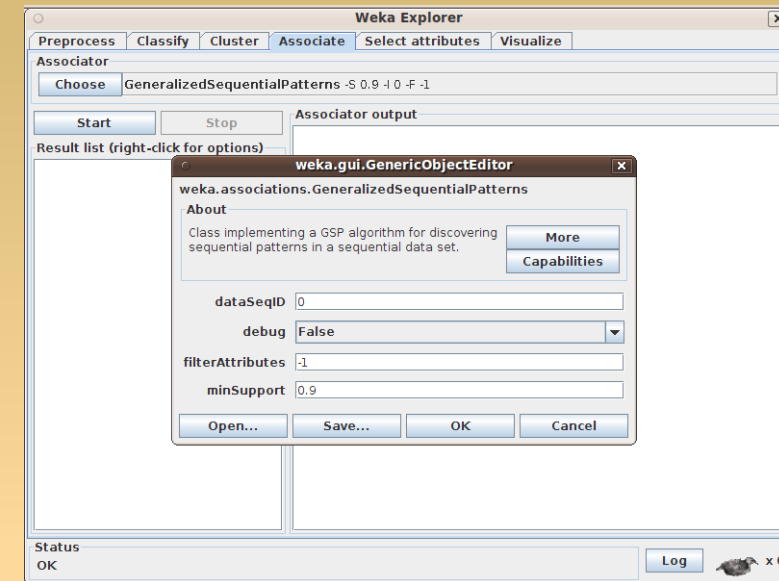


Strumenti per Sequential Pattern Mining

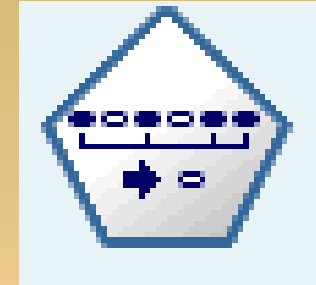
Weka

- Algoritmo GSP
- Formato file:
 - *seqID prodottoA prodottoB*
 - *1 A noB*
 - *1 noA B*
 - *2 A B*
- NB: anche il seqID deve essere nominale



Clementine

- Nodo "Sequenza"
- Formato del file:
 - Transazionale

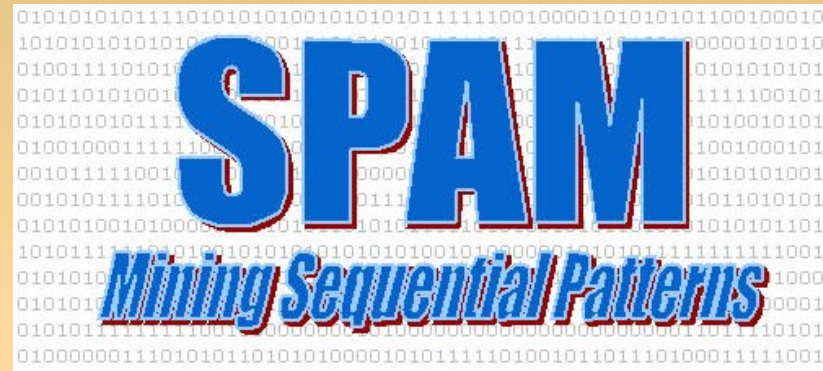


	user	time	item
1	A	0	coke
2	A	0	butter
3	A	6	milk
4	B	0	coke
5	C	0	butter
6	D	0	coke
7	D	1	butter
8	D	3	milk

- Problema: alla versione 11 sembra trovare solo sequenze del tipo $A \rightarrow B$ (2 items soli)

Tool standalone: SPAM

- <http://himalaya-tools.sourceforge.net/Spam>
- Sorgenti C++ da compilare
- Formato input:
 - Transazionale
 - Due alternative:
 - Transazionale normale con ID e item numerici
 - Transazionale "verticale" con ID e item stringhe



Tool standalone: SPAM

Formato transazionale normale

- Sintassi: (0.3 = supporto minimo)
 - `spam -sup 0.3 -fn FILE_INPUT -ascii -stdout`
- Input:
 - 0 0 100
 - 0 0 101
 - 0 6 102
 - 1 0 100
 - 2 0 101
 - 3 0 100
 - 3 1 101
 - 3 3 102

100 = coke
101 = butter
102 = milk

Tool standalone: SPAM

Formato transazionale "verticale"

- Sintassi: (0.3 = supporto minimo)
 - `spam -sup 0.3 -fn FILE_INPUT -ascii -str -stdout`
- Input:
 - A
 - 0
 - coke
 - A
 - 1
 - butter
 - ecc.

Tool standalone: Altre

- SLP Miner (Karypis)
 - <http://glaros.dtc.umn.edu/gkhome/pafi/overview>
 - Solo eseguibili x linux e SunOS
- Seqwog (Christian Borgelt)
 - <http://www.borgelt.net/seqwog.html>
 - Solo sequenze di item (no itemset) contigue
- Altri, tramite ricerche web