zhhailon / prefixspan

Imported from Google Code. https://code.google.com/p/prefixspan

1 commit **1** branch **0** releases **1** contributor

zhhailon first commit Latest commit `751423f` on 12 Mar 2013

Main.cpp	first commit	4 years ago
Makefile	first commit	4 years ago
Prefixspan.cpp	first commit	4 years ago
Prefixspan.h	first commit	4 years ago
README	first commit	4 years ago
data	first commit	4 years ago

README

```
PrefiSpan --- An Implementation of Prefix-projected Sequential Pattern mining

Author: Yasuo Tabei <tabei@cb.k.u-tokyo.ac.jp>
```

```
          Dept of Computational Biology,
          Graduate School of Frontier Science,
          University of Tokyo

 License: GPL2 (Gnu General Public License Version 2)

 Reference:
  PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth
  Jian Pei, Jiawei Han, Behzad Mortazavi-asl, Helen Pinto, Qiming Chen, Umeshwar Dayal and Mei-
chun Hsu
  IEEE Computer Society, 2001, pages 215

 Requirements:
   C++ compiler with STL (Standard Template Library).

 Install:
  % make

 Usage:
     ./lcm [options] data

     option:
         -min_sup NUM:   set minimum support        (default: 1)
         -max_pat NUM:   set maximum pattern length (default: infinity)

  Format of input data:

  3 1 3 4 5
  2 3 1
  3 4 4 3
  1 3 4 5
  2 4 1
  6 5 3

  Each line corresponds to the each transaction which has a sequence of items separated by single
space.


 Format of results:

  itemsets
  ( ids ) freq
```

```
itemsets
( ids ) freq
itemsets
( ids ) freq
...

Here is an example:

1
( 0 1 3 4 ) : 4
1 3
( 0 3 ) : 2
1 3 4
( 0 3 ) : 2
1 3 4 5
( 0 3 ) : 2
1 3 5
( 0 3 ) : 2
1 4
( 0 3 ) : 2
...

This result means:

FREQUENT SEQUENCE   : TRANSACTION ID : FREQUENCY
1                       0 1 3 4          4
1 3                     0 3              2
1 3 4                   0 3              2
1 3 4 5                 0 3              2
1 3 5                   0 3              2
1 4                     0 3              2
...

Each line represents the frequent sequences
whose frequency is no less than min_sup (-min_sup option) and the size of sequences
is less than or equal max_pat (-max_pat option).
```