

Executor: Sergei Kirjanov, 128217FAY

See sources at https://github.com/rewlad/ttu_andm

== Work Flow ==

Perl script was created.

Fixes: non-letters to underscore, letters to lower case.

Apriori was briefly overviewed, and useful options was found:

```
-t#      target type (default: s)
-tr      target type r: association rules
-s#      minimum support   of a      set/rule      (default: 10%)
-c#      minimum confidence of a      rule          (default: 80%)
```

Rule, for example:

```
D <- A B (50, 100)
```

means that:

(A and B) is in 50% of rows, (>-s)

D is in all rows where is (A and B), (>-c)

Top sets and rules was found.

Default filter parameters needed to be lowered to find rules.

```
1  #!/bin/bash
2  perl ts.pl < tshekid_office2003.csv > input2.txt
3  ./apriori -ts input2.txt output-s.txt
4  ./apriori -tr -s3.5 input2.txt output-r-s3.txt
5  ./apriori -tr -c60 input2.txt output-r-c60.txt

1  use strict;
2  use utf8;
3  my $prev_c;
4  binmode STDIN,":utf8";
5  binmode STDOUT,":utf8";
6  for(<>){
7      /^(\\d+),("[^"]+)",,/_or_print STDERR "undef:$_"_and_next;
8      my($cur_c,$txt)=($1,$2);
9      $txt=~s/\\W/_/g;
10     print $cur_c eq $prev_c ? " " : "\n";
11     $prev_c=$cur_c;
12     print lc $txt;
13 }
```

kommid (10.2941)
konserveeritud (11.7647)
pakendatud_valmistoidud (11.2745)
karastusjoogid (13.7255)
keefir (13.7255)
võid__margariinid (11.7647)
jogurtitooted (12.2549)
munad (14.2157)
külmutatud_pooltooted (13.7255)
õlled (14.7059)
kilekotid (13.7255)
juurviljad (20.098)
pant_0_5 (17.6471)
mahlad (17.6471)
pant_1_0 (20.5882)
kohupiimatooted (24.0196)
kohupiimatooted puuviljad (11.2745)
kohupiimatooted piim (10.7843)
köögiviljad (26.4706)
koored (24.0196)
koored leivad (10.7843)
koored piim (13.7255)
koored saiad (12.2549)
koored lihaproduktid (10.7843)
juustud (25.9804)
juustud leivad (12.2549)
juustud piim (10.2941)
juustud saiad (10.7843)
juustud lihaproduktid (13.7255)
leivad (28.4314)
leivad piim (14.2157)
leivad saiad (14.2157)
leivad lihaproduktid (14.7059)
puuviljad (34.8039)
puuviljad piim (12.7451)
puuviljad saiad (11.2745)
puuviljad lihaproduktid (15.1961)
piim (37.2549)
piim saiad (18.6275)
piim lihaproduktid (15.6863)
saiad (37.7451)
saiad lihaproduktid (17.6471)
lihaproduktid (46.0784)

```

pant_1_0 <- muud_alkohoolsed_joogid (3.92157, 87.5)
lihaproduktid <- hakklihad (3.92157, 87.5)
piim <- lihafileed (4.90196, 80)
lihaproduktid <- soolased_määrded (3.92157, 100)
piim <- siidrid saiad (3.92157, 87.5)
lihaproduktid <- maitseained saiad (5.39216, 90.9091)
piim <- võid__margariinid leivad (4.41176, 88.8889)
saiaid <- võid__margariinid leivad (4.41176, 88.8889)
piim <- jogurtitooted saiaid (4.90196, 90)
lihaproduktid <- jogurtitooted saiaid (4.90196, 80)
pant_0_5 <- õlled kohupiimatooted (3.92157, 87.5)
saiaid <- õlled leivad (4.41176, 88.8889)
saiaid <- võid__margariinid leivad piim (3.92157, 87.5)
piim <- võid__margariinid leivad saiaid (3.92157, 87.5)
lihaproduktid <- jogurtitooted piim saiaid (4.41176, 88.8889)
saiaid <- jogurtitooted piim lihaproduktid (4.90196, 80)
piim <- jogurtitooted saiaid lihaproduktid (3.92157, 100)
lihaproduktid <- kilekotid piim saiaid (3.92157, 87.5)
pant_0_5 <- õlled leivad saiaid (3.92157, 87.5)
õlled <- pant_0_5 leivad saiaid (3.92157, 87.5)
saiaid <- õlled pant_0_5 piim (4.41176, 88.8889)
pant_0_5 <- õlled piim saiaid (4.90196, 80)
õlled <- pant_0_5 piim saiaid (4.90196, 80)
saiaid <- pant_0_5 leivad piim (3.92157, 87.5)
piim <- pant_0_5 leivad saiaid (3.92157, 87.5)

lihaproduktid <- jogurtitooted (12.2549, 60)
lihaproduktid <- külmutatud_pooltooted (13.7255, 60.7143)
lihaproduktid <- kilekotid (13.7255, 64.2857)
pant_0_5 <- õlled (14.7059, 63.3333)
lihaproduktid <- õlled (14.7059, 60)
piim <- koored saiaid (12.2549, 64)
saiaid <- leivad piim (14.2157, 65.5172)
piim <- leivad saiaid (14.2157, 65.5172)

```