max planck institut informatik

Problem 1 (KB Inference).

In this lab, we are working on rule mining from extracted knowledge. We provide a dataset, `got-fact.txt`, which contains a collection of facts extracted from infoboxes of the *Game of Thrones* universe from Wikia. Each fact is represented in a S-P-O triple form (`subject [tab] predicate [tab] object`).

From this dataset, mine rules and rank them based on support. Mine rules of the following three forms:

$P_1(X, Y), P_2(X, Z) \longrightarrow P_3(Y, Z)$
$P_1(X, Y), P_2(Y, Z) \longrightarrow P_3(X, Z)$
$P_1(X, Y) \longrightarrow P_2(Y, X)$

For example,

$\texttt{father}(S_0, S_1), \texttt{mother}(S_0, S_2) \longrightarrow \texttt{spouse}(S_2, S_1)$
$\texttt{lover}(S_0, S_1) \longrightarrow \texttt{lover}(S_1, S_0)$
$\texttt{father}(S_0, S_1) \longrightarrow \texttt{child}(S_1, S_0)$

Your program, called `run.py`, takes the file `got-fact.txt` as the input and prints out the top 10 rules in terms of support, along with their support, for each of the three patterns above (thus, 30 rules in total).

(*Bonus 1: Score rules also by PCA-confidence*)

(*Bonus 2: Consider a richer rule language, for example, all rules with two atoms in their body.*)

How to run: `python run.py got-fact.txt`

Your submitted files must include all necessary code and files, especially the main program file `run.py`. If you used any external libraries, please indicate them in a README file.

Please submit all necessary files, which are compressed into a zip file named:
Lab08_MatriculationNumber_Name.zip
to the email address: **cxchu@mpi-inf.mpg.de** with title of the email: **[IE]Lab08_MatriculationNumber_Name**

Deadline: 23:59 14.12.2019 (Saturday)