PROJECT ASSOCIATION RULES

HPC Lab Project

- Frequent Pattern Mining
- To Be Delivered:
  - Sequential implementation of one of the algorithms discussed
  - Parallel implementation
    - Multi-threaded or Distributed or GPU
  - Report discussing performance figures of the proposed parallel implementation
    - Varying threads/cores/processors
    - Varying algo's parameters/strategies
    - Varying parallelism strategy…
- Bonus:
  - SIMD Vectorization
  - Cache analysis

HPC Lab Project

- The report should be short, ~5 pages
- I expect to find the following flow:
  - I attacked the problem with strategy A
    - Data layout and parallel decomposition
  - I expected to find these results
  - Experimental results are different
  - I tried to understand why
  - This lead me to strategy B
  - (… repeat …)

Datasets

- One possible repo is here:
  - http://fimi.uantwerpen.be/data/

HPC Lab Project

- You may deliver your code and report via moodle, or you can send me a link to your git repo, 3 days before the exam date
- Git Branching model:
  - http://nvie.com/posts/a-successful-git-branching-model/
- C++ coding guidelines:
  - https://lefticus.gitbooks.io/cpp-best-practices/content/03-Style.html
- Markdown can be used for the report
  - https://bitbucket.org/tutorials/markdowndemo
- English writing:
  - https://faculty.washington.edu/heagerty/Courses/b572/public/StrunkWhite.pdf
  - https://www.publishingcampus.elsevier.com/websites/elsevier_publishingcampus/files/Skills%20training/Elements_of_Style.pdf
  - http://services.unimelb.edu.au/__data/assets/pdf_file/0009/471294/Using_tenses_in_scientific_writing_Update_051112.pdf
  - https://pingpong.chalmers.se/public/pp/public_courses/course08583/published/1510227352918/resourceId/4156227/content/Zobel%20-%20Writing%20for%20computer%20science%203rd%20edition.pdf

References

- In-depth slides:
  - http://www.cs.uoi.gr/~tsap/teaching/2012f-cs059/material/datamining-lect4.pptx
- Chapter 6 from Mining Massive Datasets by Anand Rajaraman and Jeff Ullman.
  - http://infolab.stanford.edu/~ullman/mmds.html
- Fp-growth:
  - Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach
  - https://www2.cs.sfu.ca/~jpei/publications/dami03_fpgrowth.pdf

The End!