Software Development for Information Systems
Project 2

Alexios Spiliotopoulos - 1115201700147
Fanourios-Stylianos Psathopoulos - 1115201400320

Winter semester 2020-2021

1 Compilation

For the compilation of the main application and the unit tests we use a single Makefile. First, cd into the project's main directory (where the Makefile is located).

- To compile the main application, use **make** or **make app**
- To compile the unit tests, use **make test**

Object files are located in *build* directory and executables inside *bin* directory. To remove these items, simply run **make clean**.

2 Execution

2.1 main application

To execute the main app, use: ./bin/app [execution type] *dataset relations* [-s *n*] [-f *f*] [-p *Positive/Negative/Both*]

- (optional) execution type, if provided, it must be either "-test" or "-validate"
- dataset, pathname of the datasets directory
- relations, pathname to the relations file
- (optional) -s *n*, default: of json files in "data_path", else specify the size of the clique's hashmap
- (optional) -f *f*, the number of max features for TFIDF
- (optional) -p *Positive/Negative/Both*, print only positive relations, only negative or all types of relations

Example use:
To train the model, use:
./bin/app Datasets Datasets/sigmod_large_labelled_dataset.csv -f 1000

To test the model, use:
./bin/app -test Datasets Datasets/sigmod_large_labelled_dataset.csv

2.2 main application

To execute the unit tests use:

```
./bin/test
```

3 Project Structure and Files

- Project 2
  - bin
  - build
  - Dataset
    - sigmod_large_labelled_dataset.csv
    - sigmod_medium_labelled_dataset.csv
    - stopwords.txt
    - camera_specs
      - 2013_camera_specs
        - buy.net
          - 4233.json
          - ...
          - 6785.json
        - ...
        - www.wexphotographic.com
          - 146.json
          - ...
          - 626.json
  - include
    - acutest.h
    - classifier.h
    - clique.h
    - dataset_parsing.h
    - heap.h
    - list.h
    - map.h
    - types.h
    - util.h
    - vector.h
    - vectorizer.h
  - modules
    - heap.c
    - list.c
    - map.c
    - vector.c
  - src
    - classifier.c
    - clique.c
    - dataset_parsing.c
    - main.c
    - util.c
    - vectorizer.c
  - tests
    - datasetY *#contains 2 csv files*
    - example *#contains a few json files*
    - tests.c
  - Makefile
  - README.pdf

4 Dataset and Text Vectorization Techniques

5 Logistic Regression Classifier

5.1 Training

5.2 Testing