cmput497_a1_yonael_zichun3

Prerequisites

- Python 3.7+
- virtualenv

Setup

```
# Setup python virtual environment
$ virtualenv venv --python=python3
$ source venv/bin/activate

# Install python dependencies
$ pip install -r requirements.txt
```

Run

Simply run `python main.py`, output `TSV` files are saved into `output/` folder.

By default, the program assume wiki files are under the `data` directory, and write output file to `output` directory. However, you may change it if you want, see below for advance usage.

```
usage: main.py [-h] [--input INPUT] [--output OUTPUT]

Extract relations from wiki files.

optional arguments:
  -h, --help       show this help message and exit
  --input INPUT    Provide path to directory of input wiki files
  --output OUTPUT  Provide path to save output TSV files
```

More

Run the command below to check how many facts we extracted are missing compare to the sample data set.

The script added a bunch of special handling for things like `musicComposer -> music`, `producer -> producers`, but still requires some manual work to double check.

```
$ python check.py > coverage_report.txt
```

Notes

- For evidence, we try to keep it as short as possible. e.g, `plainlist` only show the first line which contains the `predicate`, but not the objects.

Author

- Yonael Bekele - yonael
- Michael Lin - zichun3