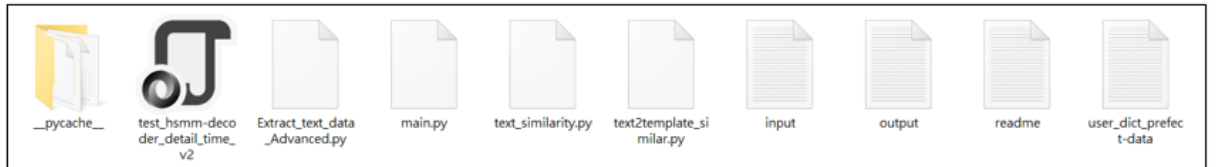
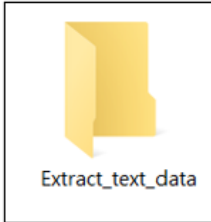


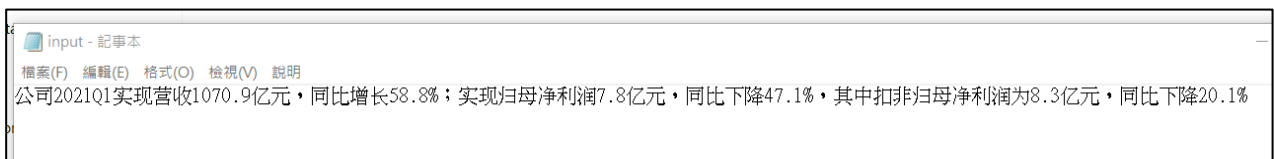
## 協助抽取指標數據



路徑: cd ./data/taiwan/Extract\_text\_data

環境: docker exec -it taiwan-new /bin/bash

## 使用方式



在 input.txt 中輸入待提取出數據的句子

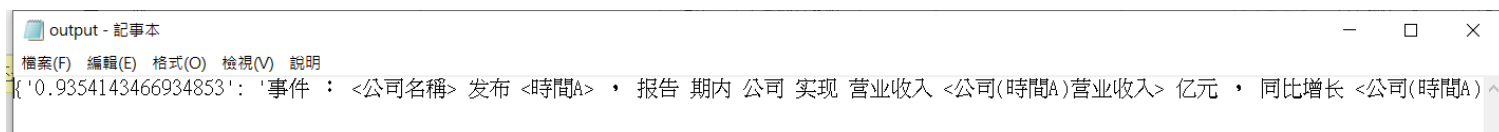
執行: python main.py

```
root@hczq-1os:/data/taiwan/Extract_text_data# python3 main.py
Building prefix dict from the default dictionary ...
Dumping model to file cache /tmp/jieba.cache
Loading model cost 1.092 seconds.
Prefix dict has been built successfully.
Processing.....
Processing.....
Processing.....
Processing.....
Processing.....
Processing.....

input_context: 事件：公司发布2021年半年报，报告期内实现营业收入14.7亿元，同比增长36.9%

[[[0.9354143466934853, '事件：<公司名稱>发布<時間A>，報告期內公司實現營業收入<公司(時間A)營業收入>億元，同比增長<公司(時間A)營業收入同比>，實現歸母淨利潤<公司(時間A)歸母淨利潤>億元，同比增長<公司(時間A)歸母淨利潤同比><eos>'], [0.9354143466934853, '事件：<公司名稱>发布<時間A>，報告期內公司實現營業收入<公司(時間A)營業收入>億元，同比增長<公司(時間A)營業收入同比>，實現歸母淨利潤<公司(時間A)歸母淨利潤>億元，同比增長<公司(時間A)歸母淨利潤同比><eos>'], [0.9166984970282112, '事件：<公司名稱>发布<時間A>，報告期內公司實現營業收入<公司(時間A)營業收入>億元，同比增長<公司(時間A)營業收入同比>，實現歸母淨利潤<公司(時間A)歸母淨利潤>億元，同比下降<公司(時間A)歸母淨利潤同比><eos>'], [0.9166984970282112, '事件：<公司名稱>发布<時間A>，報告期內公司實現營業收入<公司(時間A)營業收入>億元，同比下降<公司(時間A)營業收入同比>，實現歸母淨利潤<公司(時間A)歸母淨利潤>億元，同比增長<公司(時間A)歸母淨利潤同比><eos>'], [0.9166984970282112, '事件：<公司名稱>发布<時間A>，報告期內公司實現營業收入<公司(時間A)營業收入>億元，同比<公司(時間A)營業收入同比>，實現歸母淨利潤<公司(時間A)歸母淨利潤>億元，同比增長<公司(時間A)歸母淨利潤同比><eos>']]]
root@hczq-1os:/data/taiwan/Extract_text_data#
```

## 在 output.txt 中輸出結果



## 附註: readme.txt

1. 將欲做提取數據的句子，輸入至 input.txt 裡
2. 執行 main.py 程式(會用到 import jieba)
3. 結果將會輸出在 output.txt 裡

附註說明-輸出結果:

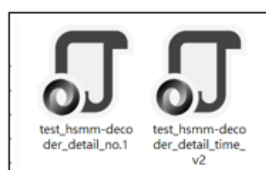
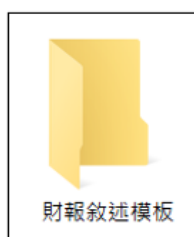
\*若能有效提取出數據，會輸出對應指標關係。如:

{'時間 A': '2021Q1', '公司(時間 A)營收': '1070.9', '公司(時間 A)營收同比': '58.8%', '公司(時間 A)归母净利润': '7.8', '公司(時間 A)归母净利润同比': '47.1%', '公司(時間 A)扣非归母净利润': '8.3', '公司(時間 A)扣非归母净利润同比': '20.1%'}

\*反之，則會輸出相似度前幾名的模板。格式為:(相似度分數,模板) 如:

{'0.9354143466934853': '事件 : <公司名稱> 发布 <時間 A> , 报告 期 内 公 司 实 现 营 业 收 入 <公司(時間 A)营业收入> 亿元 , 同比增长 <公司(時間 A)营业收入同比> , 实现 归母净利润 <公司(時間 A)归母净利润> 亿元 , 同比增长 <公司(時間 A)归母净利润同比> <eos>'}

## 上傳財報敘述模板



路徑: [cd ./data/taiwan-cuda/財報敘述模板](#)

### test\_hsmm-decoder\_detail\_time\_v2.json

```
{
  "3": {
    "1.references": [
      [
        53 items
      ]
    ],
    "2.state2words_references": [
      [
        7 props
      ]
    ],
    "3.selected_template": [
      [
        "<時間A> , 公司]25,[<時間B> 公司 实现 营收]69,[<時間A)营收> 亿元 , 同比->]108,[<公司(時間A)收入同比> , 归母净利润]44,[<時間A)公司归母净利润> 万元 , 同比下降]193,[<時間A)归母净利润同比>]19,[<eos>]240,
      ]
    ],
    "4.predictions_result": [
      [
        "<時間A> , 公司<時間B>公司实现营收<時間A)营收>亿元 , 同比->公司(時間A)收入同比> , 归母净利润<時間A)公司归母净利润>万元 , 同比下降<時間A)归母净利润同比><eos>"
      ]
    ]
  }
}
```

### test\_hsmm-decoder\_detail\_no.1.json

```
{
  "5": {
    "4.props": [
      [
        4 props
      ]
    ],
    "6": {
      "1.references": [
        [
          50 items
        ]
      ],
      "2.state2words_references": [
        [
          3 props
        ]
      ],
      "3.selected_template": [
        [
          "公司 发布<時間>]207,[2019 年 上半年]262,[报告 显示 公司]152,[总收入 <总收入> 亿元]272, [ , 同比+<营业收入同比>]30,[ , 实现]40,[归母净利润 <归母净利润> 亿]162,[ , 同比增长 <归母净利润同比>]203,[<eos>]92,
        ]
      ],
      "4.predictions_result": [
        [
          "公司发布<時間>2019年上半年报告显示公司总收入<总收入>亿元 , 同比+<营业收入同比> , 实现归母净利润<归母净利润>亿 , 同比增长<归母净利润同比><eos>"
        ]
      ]
    }
  }
}
```