Weibo Hotlist Crawl code

- ## Pipeline

  > get mid list —> get comment list —> get csv

- ## Function

  > - get mid list
  >
  >   > input page
  >   >
  >   > reutrn mid_list ## in that page ##
  >   >
  >   > Ajax : base_url + urlencode(params)
  >   >
  >   > request —> response.json() —> json.get —> re.findall
  >
  > - get flow list
  >
  >   > input mid & page
  >   >
  >   > return flow_list ## in that page ##
  >   >
  >   > Ajax
  >   >
  >   > request —> response.json() —> json.get —> re.findall
  >
  > - get comment list
  >
  >   > input mid & page
  >   >
  >   > return comment_list ## from page 1 to page ##
  >   >
  >   > ————**we must get comment from page 1**————
  >   >
  >   > ————**we need get max_id by json.get()**————
  >   >
  >   > initial :
  >   >
  >   > request —> response.json() —> json.get —> re.findall
  >   >
  >   > for :
  >   >
  >   > request —> response.json() —> json.get —> re.findall
  >
  > - main body
  >
  >   > - get mid list —> get comment list
  >   > - json file as comment.json flow.json
  >
  > - cut.py
  >
  >   > json load file comment.json
  >   >
  >   > json loads file json.dict
  >   >
  >   > open().read() ChineseStopWords.txt
  >   >
  >   > cut comments with HanLP.segment into words

```
every 6000 lines a file
mywriter = csv.writer
mywriter.writerow()
```