**\_0\_crawling\_module.py**

Crawling을 하기 위한 함수들이 구현되어 있는 파일입니다.

**\_1\_man\_cralwer.py**

<http://www.chictopia.com/> 는 전체 post들을 man Gallery 와 woman Gallery로 분류해 놓았는데, 이 중 man Gallery의 post들만 crawling하는 파일입니다. Crawling된 결과는 **crawling\_result/text** 폴더 안에**result\_man.txt**로, **crawling\_result/json** 폴더 안에 **result\_man.json**로 각각 저장됩니다.

**\_2\_woman\_crawler.py**

<http://www.chictopia.com/> 는 전체 post들을 man Gallery 와 woman Gallery로 분류해 놓았는데, 이 중 woman Gallery의 post들만 crawling하는 파일입니다. 하지만, Woman Gallery는 post의 수가 man 에 비해 월등히 많기 때문에, Garment(category) 별로 random 하게 크롤링할 수 있도록 구현했습니다. 예를 들어, dress category의 경우, 4322 pages가 있는데, 이 중 10%의 페이지만을 랜덤하게 크롤링하고 싶고, 최대 1000개의 post만을 얻고 싶다면,

crawling\_module.random\_sampler('dress',4322, 0.1, 1000)

와 같이 crawling\_module 파일의 함수를 호출하면 됩니다.

Crawling된 결과는 **crawling\_result/text** 폴더 안에**result\_dress.txt**로, **crawling\_result/json** 폴더 안에 **result\_dress.json**로 각각 저장됩니다.

**\_3\_combine\_jsons.py**

각각 카테고리 별로 크롤링 된 woman 의 category를 한 파일로 합치기 위한 파일입니다.

합쳐진 결과는 **crawling\_result/temp** 폴더의 **woman\_combined.txt**과 **woman\_combined.json** 에 저장됩니다.

**\_4\_delete\_duplicates.py**

크롤링된 result\_man.json 파일과 3번에서 합쳐진 woman\_combined.json 파일의 중복되는 post를 제거하는 파일입니다. 각각 **crawling\_result/final** 폴더의 **man\_final.json**과 **woman\_final.json** 로 저장됩니다.

**\_5\_download\_img.py**

이미지를 다운로드하는 파일입니다. **man/img** 폴더와 **woman/img** 폴더 안에 저장됩니다.

**\_6\_jsons\_to\_xml.py**

Json 파일을 XML 파일로 변환하는 파일입니다. **man/img** 폴더와 **woman/img** 폴더 안에 저장됩니다.

**해당 크롤링 프로그램의 결과의 예시는 sample 폴더 안에 저장했습니다.**

Woman의 category와 page 수는 다음과 같습니다.

[**coat**](http://www.chictopia.com/browse/people/clothes-coat?g=1)

[coat](http://www.chictopia.com/browse/people/clothes-coat?g=1) (11789)

[**dress**](http://www.chictopia.com/browse/people/clothes-dress?g=1)

[dress](http://www.chictopia.com/browse/people/clothes-dress?g=1) (39711)

[**vest**](http://www.chictopia.com/browse/people/clothes-vest?g=1)

[vest](http://www.chictopia.com/browse/people/clothes-vest?g=1) (5509)

[**shirt**](http://www.chictopia.com/browse/people/clothes-shirt?g=1)

[shirt](http://www.chictopia.com/browse/people/clothes-shirt?g=1) (23051)

[blouse](http://www.chictopia.com/browse/people/clothes-blouse?g=1) (13794)

[t-shirt](http://www.chictopia.com/browse/people/clothes-t-shirt?g=1) (10495)

[sweater](http://www.chictopia.com/browse/people/clothes-sweater?g=1) (15328)

[top](http://www.chictopia.com/browse/people/clothes-top?g=1) (25298)

[**jumper**](http://www.chictopia.com/browse/people/clothes-jumper?g=1)

[jumper](http://www.chictopia.com/browse/people/clothes-jumper?g=1) (2820)

[**jacket**](http://www.chictopia.com/browse/people/clothes-jacket?g=1)

[jacket](http://www.chictopia.com/browse/people/clothes-jacket?g=1) (20374)

[blazer](http://www.chictopia.com/browse/people/clothes-blazer?g=1) (11840)

[**pants**](http://www.chictopia.com/browse/people/clothes-pants?g=1)

[pants](http://www.chictopia.com/browse/people/clothes-pants?g=1) (17232)

[jeans](http://www.chictopia.com/browse/people/clothes-jeans?g=1) (18002)

[shorts](http://www.chictopia.com/browse/people/clothes-shorts?g=1) (15912)

[panties](http://www.chictopia.com/browse/people/clothes-panties?g=1) (244)

[leggings](http://www.chictopia.com/browse/people/clothes-leggings?g=1) (5339)

[**skirt**](http://www.chictopia.com/browse/people/clothes-skirt?g=1)

[skirt](http://www.chictopia.com/browse/people/clothes-skirt?g=1) (31883)

[**scarf**](http://www.chictopia.com/browse/people/clothes-scarf?g=1)

[scarf](http://www.chictopia.com/browse/people/clothes-scarf?g=1) (7922)

[**bag**](http://www.chictopia.com/browse/people/clothes-bag?g=1)

[bag](http://www.chictopia.com/browse/people/clothes-bag?g=1) (36526)

[purse](http://www.chictopia.com/browse/people/clothes-purse?g=1) (5119)

[**shoes**](http://www.chictopia.com/browse/people/clothes-shoes?g=1)

[shoes](http://www.chictopia.com/browse/people/clothes-shoes?g=1) (37678)

[sneakers](http://www.chictopia.com/browse/people/clothes-sneakers?g=1) (4144)

[boots](http://www.chictopia.com/browse/people/clothes-boots?g=1) (26952)

[flats](http://www.chictopia.com/browse/people/clothes-flats?g=1) (3555)

[wedges](http://www.chictopia.com/browse/people/clothes-wedges?g=1) (4476)

[sandals](http://www.chictopia.com/browse/people/clothes-sandals?g=1) (5305)

[loafers](http://www.chictopia.com/browse/people/clothes-loafers?g=1) (1304)

[clogs](http://www.chictopia.com/browse/people/clothes-clogs?g=1) (453)

[**hat**](http://www.chictopia.com/browse/people/clothes-hat?g=1)

[hat](http://www.chictopia.com/browse/people/clothes-hat?g=1) (15831)

[glasses](http://www.chictopia.com/browse/people/clothes-glasses?g=1) (2610)