**[coat, dress, jacket, jumper, pants, shirt, skirt, winter jacket]**

**이미지와 xml 파일들이 저장되는 폴더입니다. 이미지들은 이전에 검토 작업을 하며, 고령자들에게 어울리지 않는 이미지를 걸러냈으며, floral, striped, dot attribute가 제대로 tag되지 않은 이미지를 삭제했습니다.**

**[Python\_files]**

1. **module.py**

**module.py**

**: 함수들이 저장되어 있는 파일입니다.**

1. **preprocessing.py (실행하지 않아도 됩니다)**

**사실상, 해당 파일은 한 번만 실행하면 되며, 이미 한 차례 실행하였습니다.**

* 이미지 파일의 이름 바꾸기. (Rename Image files)

원본 이미지 이름은 Heathered\_Pocket\_Babydoll\_Dress/img\_00000012.jpg와 같이 직관적이지 않습니다. 따라서, Heathered\_Pocket\_Babydoll\_Dress\_img\_00000012.jpg와 같은 형태로 바꾸었습니다. 이 작업은 한 번만 실행하면 되는데, **이미 실행했기 때문에 함수를 호출하는 부분은 주석 처리 했습니다.**

* 기존 데이터 셋의 Annotation의 text file들을 쉬운 parsing을 위해 변경(Change text files) 했고,

**./Anno/changed/list\_attr\_img.txt,**

**./Anno/changed/list\_attr\_cloth.txt,**

**./Anno/changed/list\_category\_cloth.txt,**

**./Anno/changed/list\_category\_img.txt,**

**./Anno/changed/list\_bbox.txt**

으로 저장했습니다.

* Text file을 csv file 로 parsing 했습니다.

**./csv/attr\_combined.csv**

**./csv/category\_combined.csv**

**./csv/attr\_category\_combined.csv**

**./csv/bbox.csv**

* 현재 이미지 파일들은 원본 이미지 파일들 중에서 몇 차례 검토하여 추려 낸 이미지 파일입니다.(get data of img files) 따라서, 삭제된 이미지 파일에 대한 정보는 필요 없으므로, 존재하는 이미지 파일들에 대해서만 관리할 수 있도록 구현했고, **./csv/df\_category\_attr\_combined\_from\_imgfile.csv**로 저장했습니다.

1. **merge\_by\_image\_attributes.py**

**image의 attribute를 random sampling을 통해 직접 이미지를 확인 후, 비슷한 attribute끼리 합치는 작업과 category 특성에 따라 attribute를 유추할 수 있는 경우에 attribute를 지정하는 작업을 했습니다. explore\_attributes.py에서 확인 가능합니다.**

* winter jacket category는 winter attribute를 가지고, skirt category와 dress category는 woman attribute를 가지는 것은 자명합니다.
* module.py의 plotRandomImage\_by\_attribute(attribute, img\_num, df)로 직접 random sampling을 통해 attribute에 대해 분석하고, 비슷한 attribute는 합쳤습니다. 예를 들어, ['dotted ', 'dot ', 'dots ', 'polka dot ']는 plotRandomImage\_by\_attribute로 확인을 한 결과, 모두 dotted와 비슷한 attribute라고 판단 되어 dotted attribute로 합쳤습니다.
* 결과는 **./csv/merged\_by\_img\_attributes.csv** 에 저장했습니다.

1. **merge\_by\_imagd\_name.py**

**이미지 파일 이름에 attribute 정보가 들어간 경우, 해당 attribute를 뽑아내는 작업을 했습니다.**

* 결과는 **./csv/merged\_by\_img\_name.csv** 에 저장했습니다.

1. **convert\_to\_xml\_based\_csv.py**

**xml로 변환하기 위한 csv를 만드는 작업입니다.**

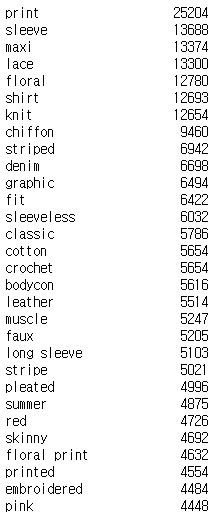
* 결과는 **./csv/final\_csv\_for\_xml.csv** 에 저장했습니다.

1. **convert\_to\_xml.py**

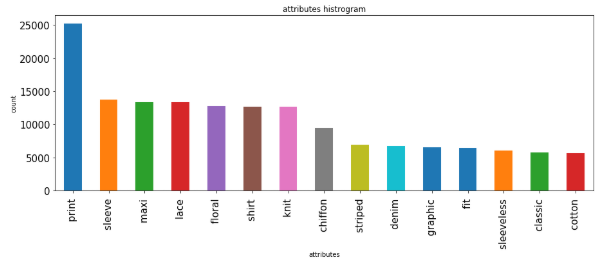
**final\_csv\_for\_xml.csv의 각 row data를 읽어, xml file을 생성하는 파일입니다.**

**Analyze**

**Before merging attributes(left) After merging(right)**



**Before merging attributes**



**After merging**

