**build\_vocab.zip** codes

**data100k.zip** Raw data

**Trends\_Glossary.json** Gold standard glossary/vocabulary.

**maxRecall100K**  all the terms found in 100K documents.

We are working with 4 classes which are as follows.

ART : article\_types

ATTR : merged attribute\_types and attribute\_values

FSN : fashion\_strings

FBR : fabric

As there were some issues with brands (BRD), we left that out.

trainBIO, testBIO : sentence is separated by a newline and each line represents word, tag.

train3010, test3010 : corresponding sentences are pickled in these files