**Describe the process of scraping, provide urls used for scrapping and develop plain corpus.**

The loop runs on a specified range and then beautiful soup is used to parse the html (articles) corresponding to the numbers in range. Then the data obtained is cleaned by removing all html tags, punctuation marks and numerical values since we are looking for roman Urdu words. The links used are given below:

<https://medium.freecodecamp.org/how-to-scrape-websites-with-python-and-beautifulsoup-5946935d93fe>

<https://stackoverflow.com/questions/16121001/suggestions-on-get-text-in-beautifulsoup>

**Run skip gram word embedding algorithm with:**

1. **k=2,5,10,15,20**

CONTEXT\_SIZE = 2

EMBEDDING\_DIM = 2

EPOCHS = 5

Epoch 0 Loss: 261641.21875

Beginning epoch 1

Epoch 1 Loss: 259683.51562

Beginning epoch 2

Epoch 2 Loss: 258116.34375

Beginning epoch 3

Epoch 3 Loss: 256806.00000

Beginning epoch 4

Epoch 4 Loss: 255678.90625

CONTEXT\_SIZE = 5

EMBEDDING\_DIM = 2

EPOCHS = 10

Epoch 0 Loss: 712534.06250

Beginning epoch 1

Epoch 1 Loss: 670855.31250

Beginning epoch 2

Epoch 2 Loss: 654694.25000

Beginning epoch 3

Epoch 3 Loss: 646409.62500

Beginning epoch 4

Epoch 4 Loss: 640835.93750

Beginning epoch 5

Epoch 5 Loss: 636606.93750

Beginning epoch 6

Epoch 6 Loss: 633237.62500

Beginning epoch 7

Epoch 7 Loss: 630440.00000

Beginning epoch 8

Epoch 8 Loss: 628069.25000

Beginning epoch 9

Epoch 9 Loss: 626031.56250

CONTEXT\_SIZE = 10

EMBEDDING\_DIM = 2

EPOCHS = 10

Epoch 0 Loss: 1388236.62500

Beginning epoch 1

Epoch 1 Loss: 1302604.00000

Beginning epoch 2

Epoch 2 Loss: 1277561.75000

Beginning epoch 3

Epoch 3 Loss: 1263383.62500

Beginning epoch 4

Epoch 4 Loss: 1253772.25000

Beginning epoch 5

Epoch 5 Loss: 1246742.62500

Beginning epoch 6

Epoch 6 Loss: 1241430.12500

Beginning epoch 7

Epoch 7 Loss: 1237222.75000

Beginning epoch 8

Epoch 8 Loss: 1233812.50000

Beginning epoch 9

Epoch 9 Loss: 1231008.00000

CONTEXT\_SIZE = 15

EMBEDDING\_DIM = 2

EPOCHS = 10

Epoch 0 Loss: 2057541.37500

Beginning epoch 1

Epoch 1 Loss: 1931330.25000

Beginning epoch 2

Epoch 2 Loss: 1896187.50000

Beginning epoch 3

Epoch 3 Loss: 1877029.37500

Beginning epoch 4

Epoch 4 Loss: 1864735.12500

Beginning epoch 5

Epoch 5 Loss: 1856215.12500

Beginning epoch 6

Epoch 6 Loss: 1849955.87500

Beginning epoch 7

Epoch 7 Loss: 1845183.37500

Beginning epoch 8

Epoch 8 Loss: 1841484.75000

Beginning epoch 9

Epoch 9 Loss: 1838495.12500

CONTEXT\_SIZE = 20

EMBEDDING\_DIM = 2

EPOCHS = 10

Epoch 1 Loss: 2541476.75000

Beginning epoch 2

Epoch 2 Loss: 2500793.25000

Beginning epoch 3

Epoch 3 Loss: 2478519.25000

Beginning epoch 4

Epoch 4 Loss: 2464712.75000

Beginning epoch 5

Epoch 5 Loss: 2455343.00000

Beginning epoch 6

Epoch 6 Loss: 2448592.00000

Beginning epoch 7

Epoch 7 Loss: 2443462.50000

Beginning epoch 8

Epoch 8 Loss: 2439401.50000

Beginning epoch 9

Epoch 9 Loss: 2436073.50000

1. **embedding dimension=100,200,300**

CONTEXT\_SIZE = 2

EMBEDDING\_DIM = 100

EPOCHS = 10

Epoch 0 Loss: 286755.81250

Beginning epoch 1

Epoch 1 Loss: 269985.53125

Beginning epoch 2

Epoch 2 Loss: 258952.54688

Beginning epoch 3

Epoch 3 Loss: 250973.20312

Beginning epoch 4

Epoch 4 Loss: 244771.20312

Beginning epoch 5

Epoch 5 Loss: 239728.26562

Beginning epoch 6

Epoch 6 Loss: 235476.17188

Beginning epoch 7

Epoch 7 Loss: 231787.03125

Beginning epoch 8

Epoch 8 Loss: 228523.10938

Beginning epoch 9

Epoch 9 Loss: 225593.12500

CONTEXT\_SIZE = 2

EMBEDDING\_DIM = 200

EPOCHS = 10

Epoch 0 Loss: 284255.34375

Beginning epoch 1

Epoch 1 Loss: 262333.25000

Beginning epoch 2

Epoch 2 Loss: 248253.56250

Beginning epoch 3

Epoch 3 Loss: 238080.81250

Beginning epoch 4

Epoch 4 Loss: 230127.62500

Beginning epoch 5

Epoch 5 Loss: 223586.43750

Beginning epoch 6

Epoch 6 Loss: 218014.81250

Beginning epoch 7

Epoch 7 Loss: 213148.81250

Beginning epoch 8

Epoch 8 Loss: 208826.32812

Beginning epoch 9

Epoch 9 Loss: 204941.98438

CONTEXT\_SIZE = 2

EMBEDDING\_DIM = 300

EPOCHS = 10

Epoch 0 Loss: 281219.37500

Beginning epoch 1

Epoch 1 Loss: 255363.57812

Beginning epoch 2

Epoch 2 Loss: 239372.95312

Beginning epoch 3

Epoch 3 Loss: 227796.10938

Beginning epoch 4

Epoch 4 Loss: 218648.40625

Beginning epoch 5

Epoch 5 Loss: 211063.84375

Beginning epoch 6

Epoch 6 Loss: 204597.23438

Beginning epoch 7

Epoch 7 Loss: 198975.20312

Beginning epoch 8

Epoch 8 Loss: 194019.29688

Beginning epoch 9

Epoch 9 Loss: 189606.40625