Done by:

Saif Ullah 19L-0953 and Abdullah Asim 19L-1160

**Same Context Word Finder**

**Introduction:**

Our main goal is to find words of same context. For example if we give input king then we will get queen, palace and all other words similar to king.

We have used word2vec to achieve our goal.

**Our Work:**

Basically word2vec is trained only on wikipedia data and it is trained on english data. We have decided to extend this idea by training it on different dataset and in urdu instead of english. We will give input an urdu word and output will also be in urdu.

**Implementation:**

We have gathered data from websites like hamariweb, urduwiki and bbc urdu. We have scraped data from these websites and then cleaned it (removing words other than urdu).

Then we have created word embeddings of the data scraped and trained our model on the embeddings.

Finally when we give an urdu word as input we get words of same context in urdu

along with their cosine distances.

Scrapping is done through python library Beautiful Soup.

**Output:**

Text

Description automatically generated

Following is the output of our model we get words of same context along with their cosine distances.