**Weekly Assignment 7**

**Saisrikar Paruchuri**

**44751575**

**INFS7410**

**Weekly Assignment-7**

**Unigram Language Model**

Unigram language model is used in information retrieval it can be treated as a combination of several one state finite automata. It splits the probabilities of terms in the context.

**Drawbacks of Unigram**

The assumption of unigram model is the term in a document are independent with each other, which ignores the latent semantic relationship among the terms in a document. So, any term that is there in a document it will retrieve that document in this process it will retrieve many documents and processing time is more and the results will be inefficient also it will take more space and the cost will be very high.

The unigram precision score shows that we are getting 65% of the root words correct. However, the unigram precision score with full words is about 52% for the retrieval. Thus, we are missing about 13% of the words although we seem to be getting their roots correct.

**Improving the Unigram**

To improve unigram there are many erroneous words which are easy to fix suggests some ideas to improve unigram precision. One can utilize a morpheme level spelling corrector that operates on segmented representations, and corrects such forms to possible morphologically correct words to form a lattice which can again be rescored to select the contextually correct one.