## Distributed Representations of Words and Phrases and their Compositionality summary:

**Paper extends skipgram model in several ways to improve its efficiency.**
**Subsampling frequent words.**
**Negative sampling is done to remove noise.**
These both improved computational efficiency.
Words that appear in phrases frequently are phrased together.
Phrase vectors are used so as to capture idiomatic phrases. Phrases are used as individual tokens.
Hierarchal softmax is computationally faster than regular softmax due to binary trees. Here, Huffman tree is used.
Vectors are somehow capturing more linear properties (+/-) like Russia + river = Volga river.
We could use much larger data to train due to computational efficiency which resulted in higher accuracy.

## Efficient Estimation of Word Representations in Vector Space:

Hierarchal softmax, Huffman binary tree are used which ease computation.
RNNs further ease computation.
Parallel training of neural networks done (distbelief).
New log linear models are used which are computationally easing due to less nonlinearity while still having nonlinearity.
1st architecture- Continuous bag of words model is used where future and history words are used to classify middle word.
2nd architecture- Continuous skipgram model is used with nearby past and future words.
Biggest - big + small = smallest
France is to Paris and Germany is to Berlin (semantic relationships).
5 types of semantic and 9 types of syntactic questions have been created for test.
Training data on large dataset for one epoch is giving better result then training thrice on small dataset.
Weighted combination of both models is giving higher accuracy in Microsoft sentence completion test.