**Knowledge-base for Word Prediction**

Ohad Abramovitz: 307880112

Ori Suchy: 204051114

The task:

Map-Reduce system for calculating the conditional probability of each trigram found in a given corpus.

This system input corpus is Hebrew 3-Gram dataset of Google Books Ngrams.

The output of the system is a list of word trigrams and their conditional probabilities.

The list is ordered:

1. By ascending
2. By the probability , descending.

The probabilities are calculated as suggested by [Thede & Harper](https://dl.acm.org/doi/10.3115/1034678.1034712):

* is the number of times occurs.
* is the number of times sequence ) occurs.
* is the number of times sequence occurs.
* is the total number of word instances in the corpus.
* is the number of times occurs.
* is the number of times sequence occurs.