# שלבי האלגוריתם:

## ALL against all app. Matching

FileA, FileB, minDistanceMatch, maxErrorMatch -> output1

## UNION

output1, maxDistanceForUnion -> output2

**newMaxDistance = (len(rec1) + len(rec2)) \* maxDistanceForUnion – התחשבות באורך הסגמנטים**

For every two matches in the first alg. לא יעיל, אם נמיין אפשר לחסוך הרבה חישובים –

### Term Frequency (maybe change to TF-IDF )

FileA, FileB -> wordCountDict

## Local Align

FileA, FileB, output2, localAlignPadRatio, wordCountDict -> outFile

First pads segments with PadRatio

### words\_water

Based on [Smith**–**Waterman](http://en.wikipedia.org/wiki/Smith%E2%80%93Waterman_algorithm) local alignment alg.

seq1, seq2, wordCountDict, words\_mismatch\_penalty, words\_gap\_penalty

### water

seq1, seq2, match\_award, mismatch\_penalty, gap\_penalty