All Against All Textual Approximate Matching Tool

# Required Features

* Document vs. Document online matching
* Document vs. offline corpus matching
* Language independent (UNICODE)
* Document / Paragraph / Ngram level match
* Word/letter mismatch – penalty relational to inverse frequency
* Client for watching the matches
* Plugins:
  + Stop words List
  + Prefix list
  + Morphology

# Accuracy Test

* Get review of a humanitarian experts (Tibetan and Bible).
* Generate synthetic text with known errors and check recall/precision