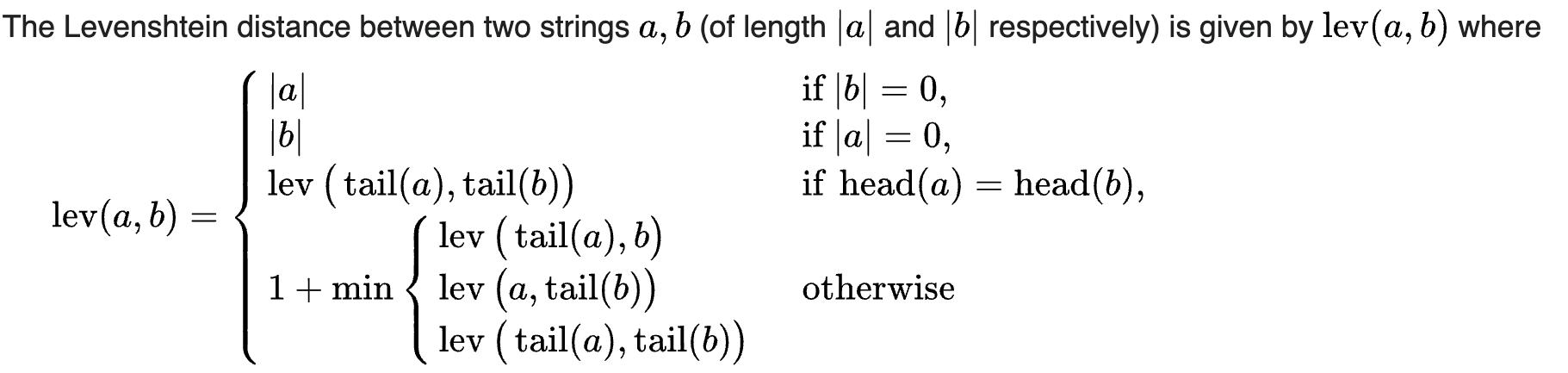
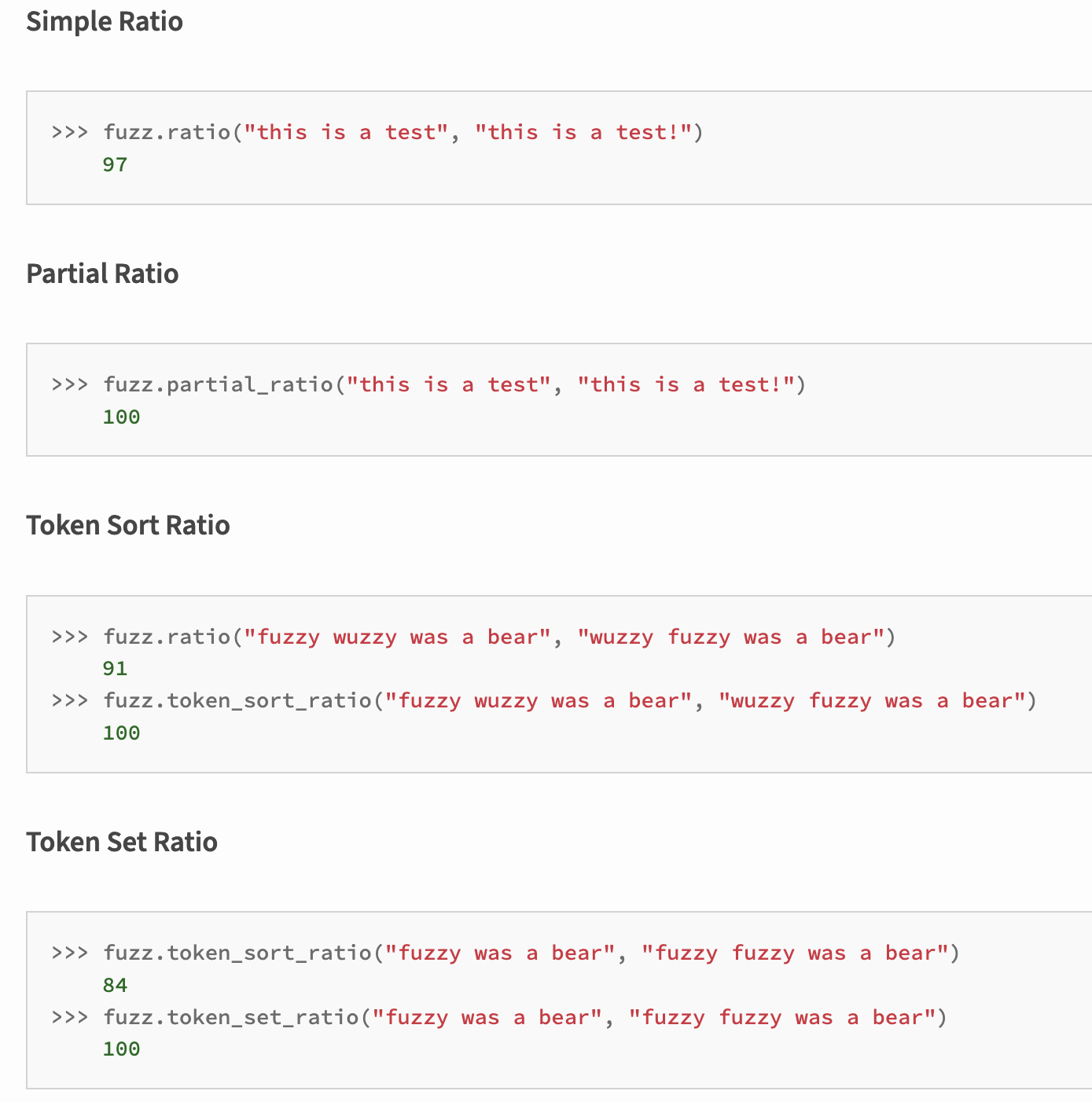
**Fuzzy**

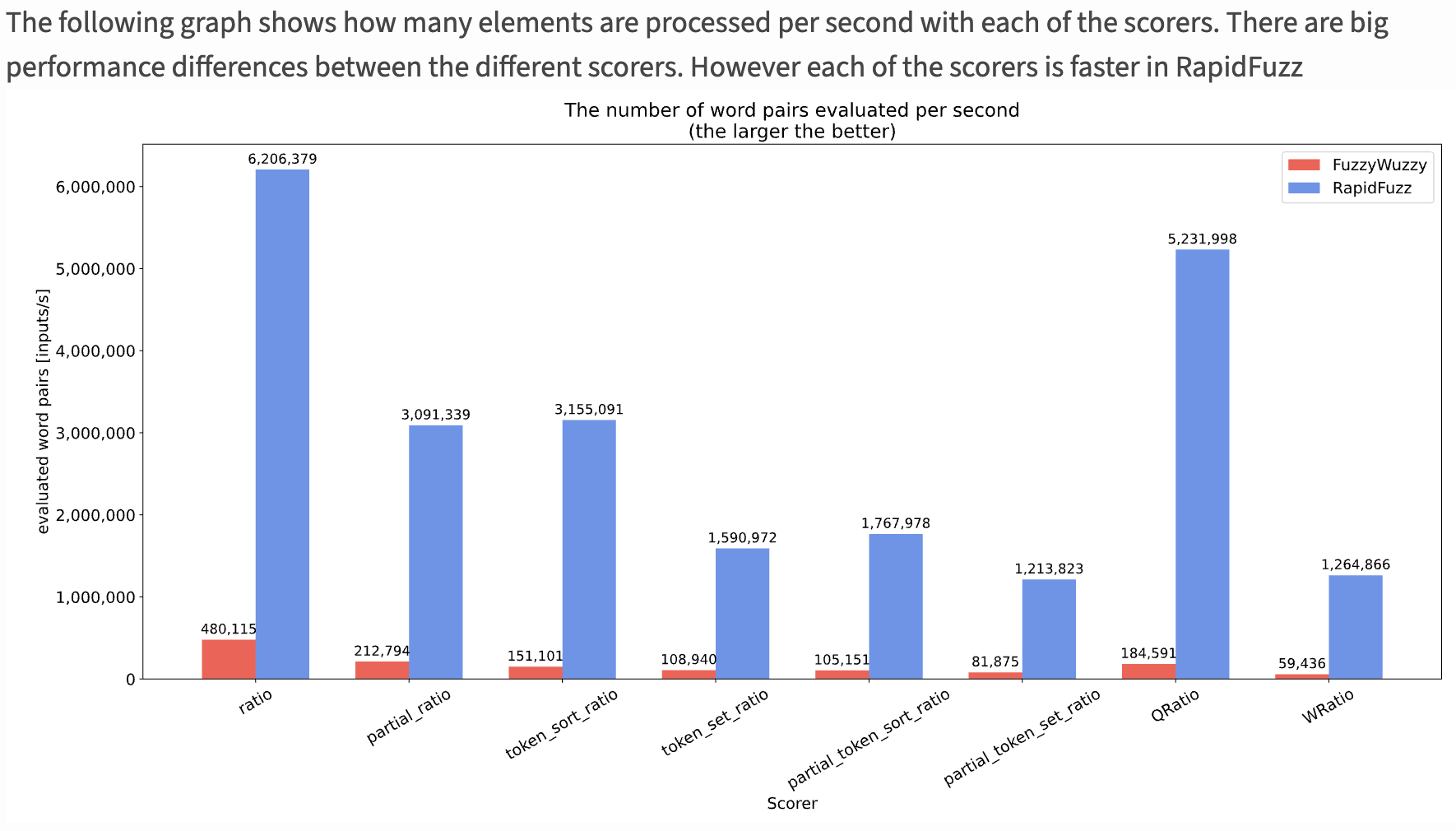
* (Python library) [FuzzyWuzzy](https://pypi.org/project/fuzzywuzzy/)
  + uses the Levenshtein distance to calculate string similarity



minimum number of single-character edits (insertions, deletions or substitutions) required to change one word into the other

* + Not specific to name matching, but for text similarity
  + Simple ratio, partial ratio, token sort ratio, and token set ratio available



* + Does not support unicode
* [RapidFuzz](https://pypi.org/project/rapidfuzz/1.5.0/)
  + Diff from FuzzyWuzzy: mostly written in C++, different license
  + Same formulas: simple ratio, partial ratio, token sort ratio, and token set ratio
  + Provides the same results as in FuzzyWuzzy, but with a faster speed
  + Supports unicode
  + But less mature than FuzzyWuzzy
* [Jellyfish](https://pypi.org/project/jellyfish/)

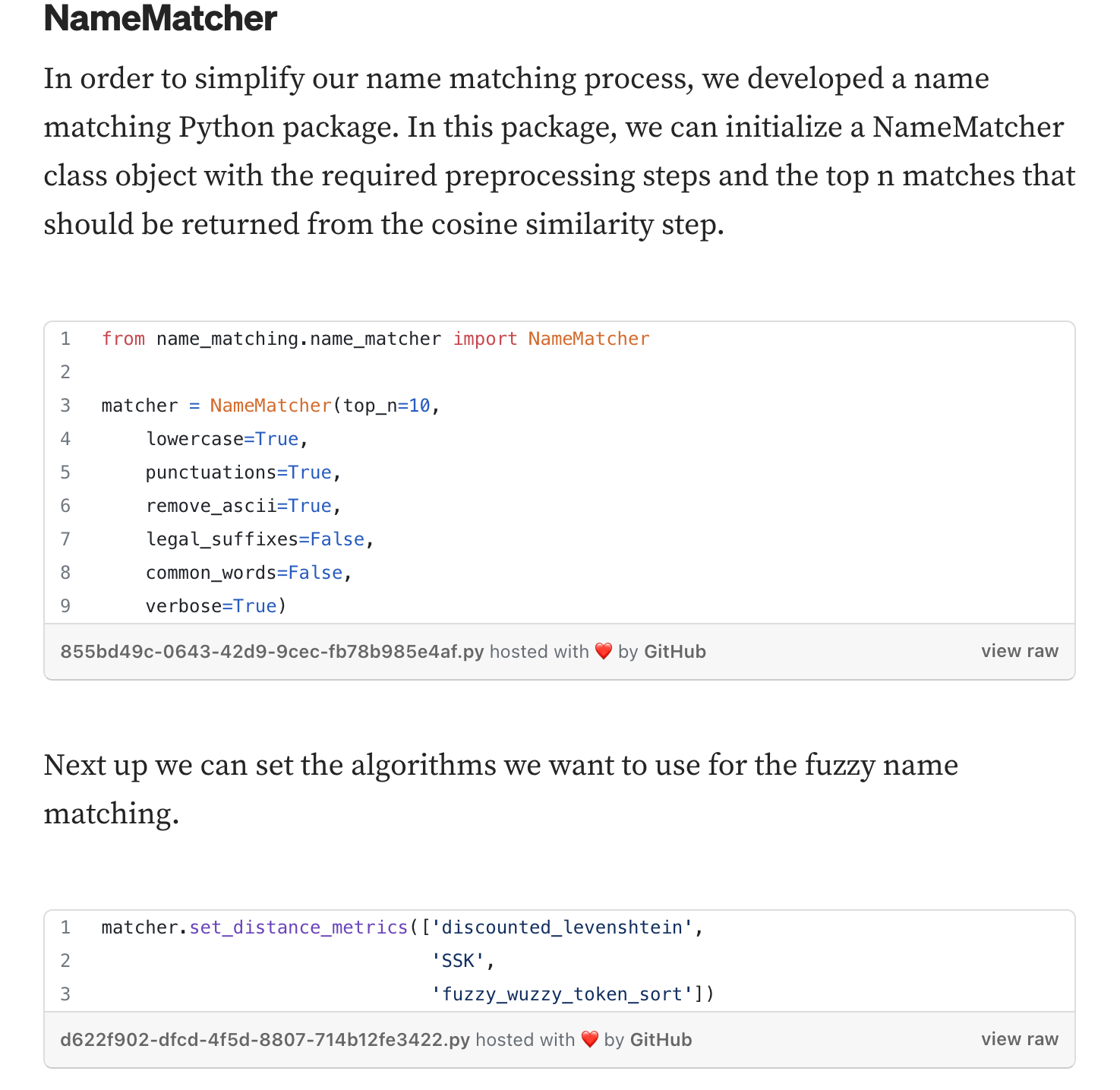
approximate & phonetic matching of strings

String comparison: Levenshtein Distance, Damerau-Levenshtein Distance, Jaro Distance, Jaro-Winkler Distance, Match Rating Approach Comparison, Hamming Distance

Phonetic encoding: American Soundex, Metaphone, NYSIIS (New York State Identification and Intelligence System), Match Rating Codex



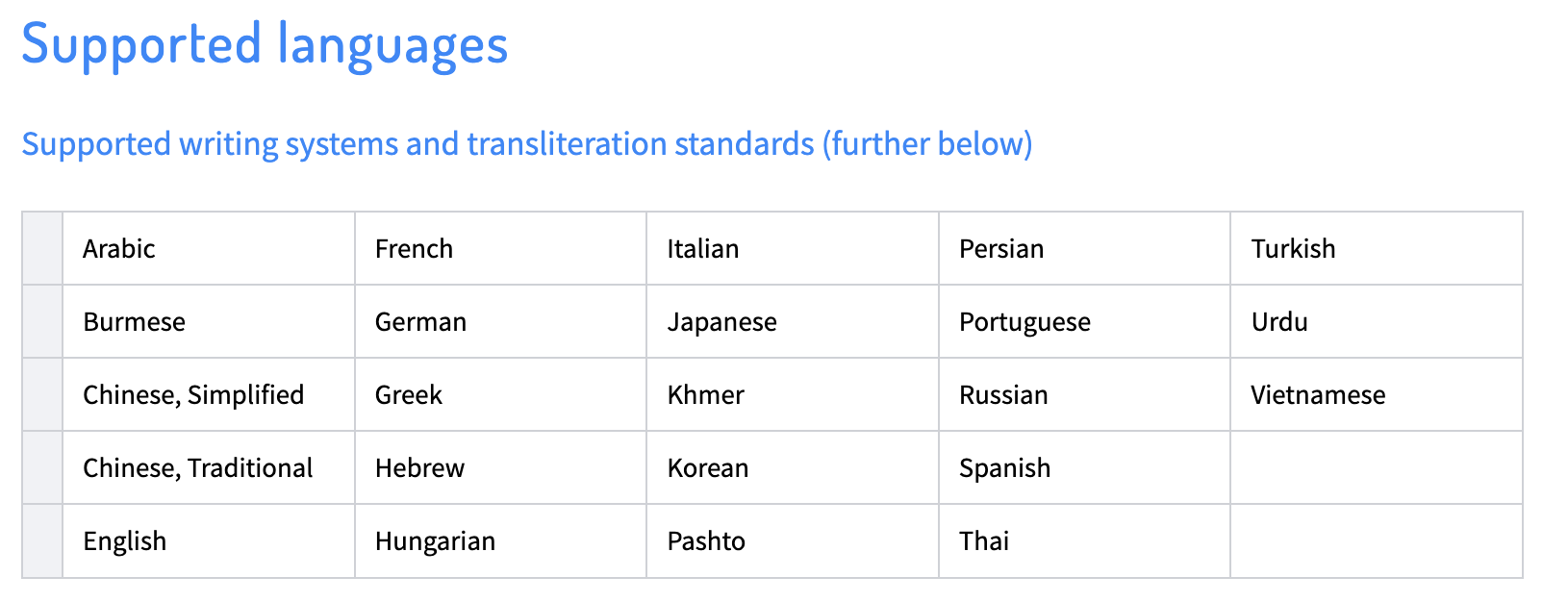
[**NameMatcher**](https://pypi.org/project/name-matching/0.8.2/)

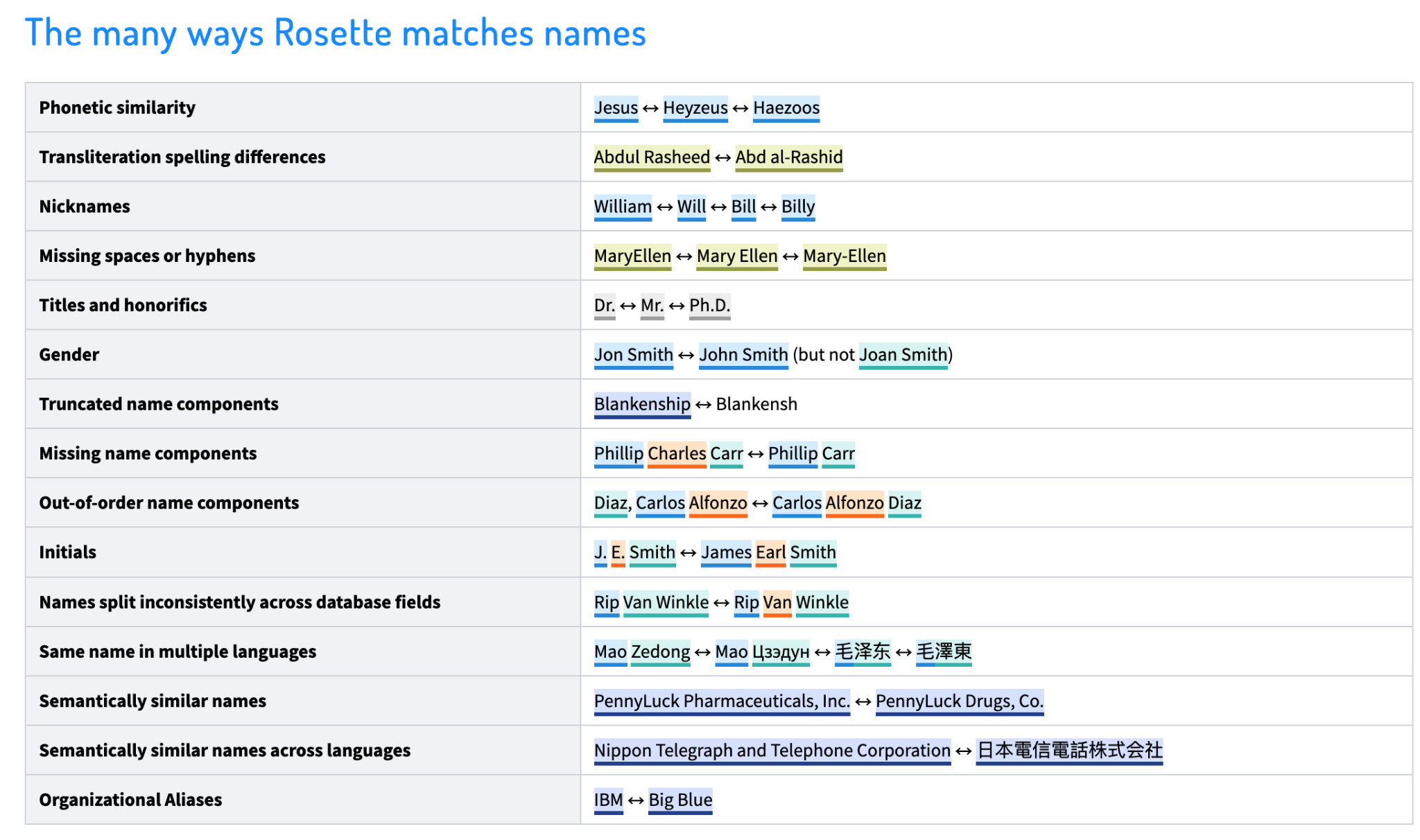


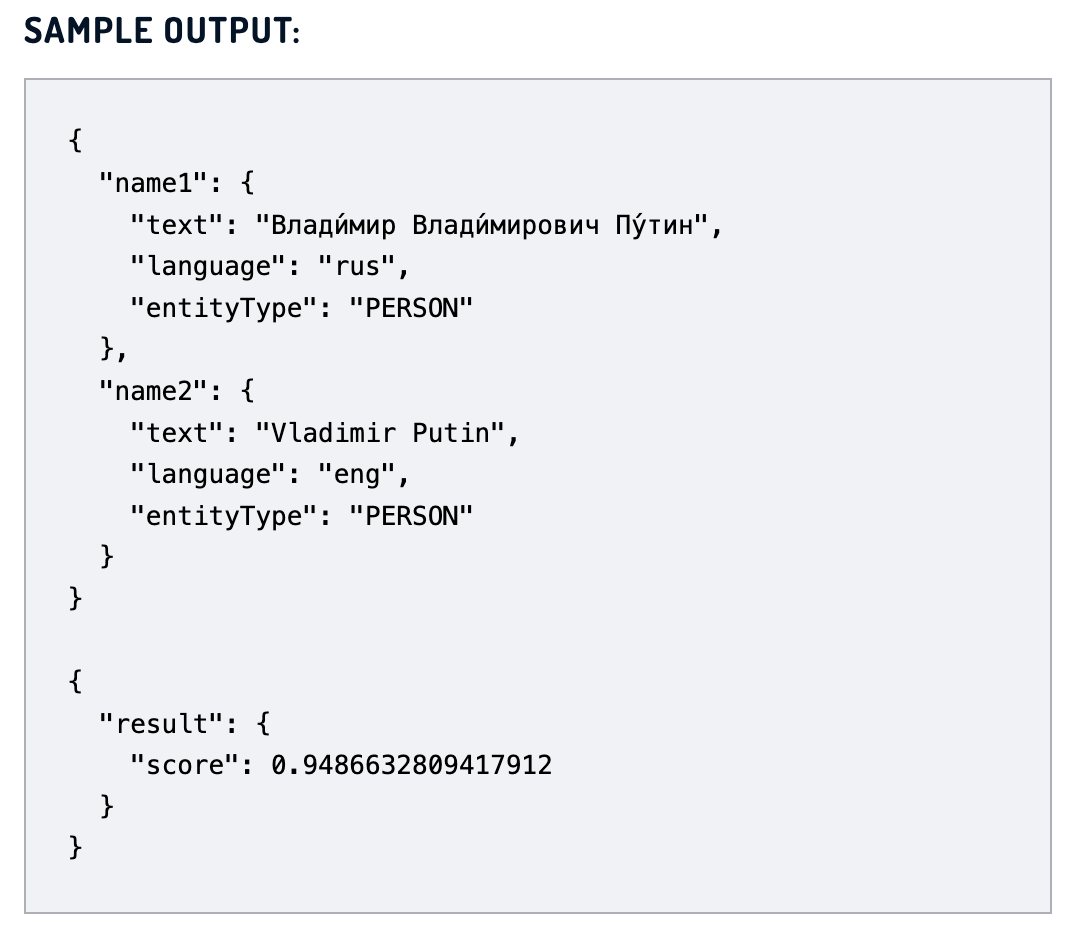
Source: [Company Name Matching](https://medium.com/dnb-data-science-hub/company-name-matching-6a6330710334)

[**Rosette**](https://www.rosette.com/capability/name-indexer/#overview)

* Matches names of people, locations, and organizations
* Cross-lingual



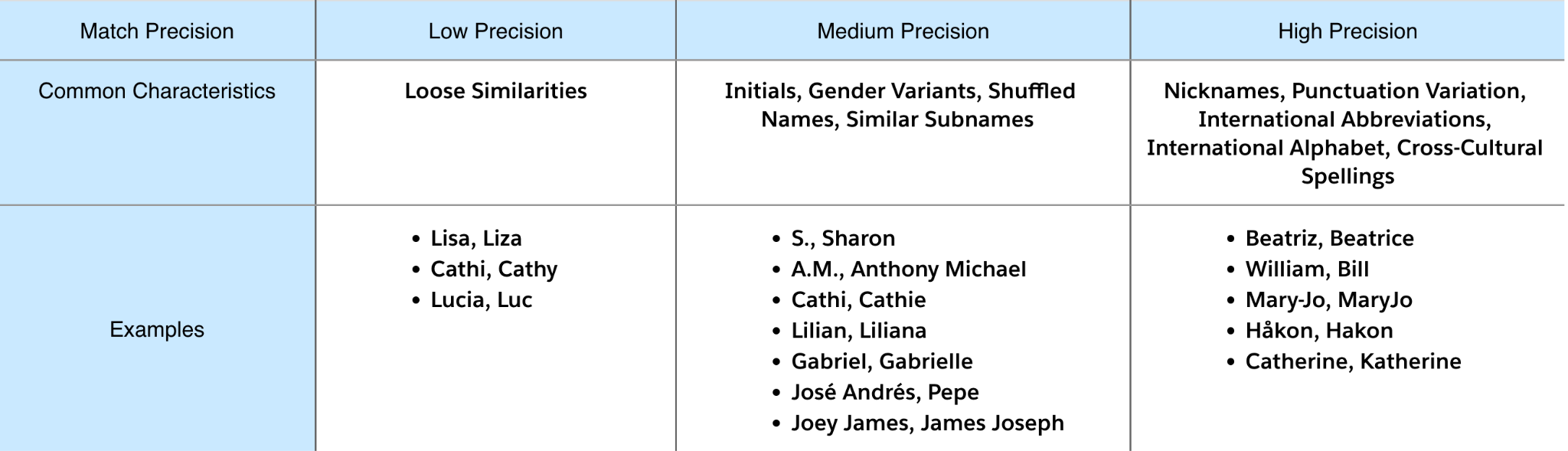
* Very comprehensive ways to match
* Ranks results by match score



* Paid service, Rosette Cloud 30 days free trial available

[**AI-Based Fuzzy**](https://engineering.salesforce.com/ai-based-identity-resolution-the-key-for-linking-diverse-customer-data/)

By Salesforce AI & Salesforce Data Cloud [[paper](https://arxiv.org/pdf/2111.10497.pdf)]



multilingual DistilBERT + MLP