Customer Name Fuzzy Matching Package

Flow chart of current version of 'Fuzzy Matching' algorithms. The package is continuously updated, tested, and validated. Current uses include matching records to B2B customer hierarchies (using customer names and country specifically), account hierarchy cleaning, mapping error rate assessment, matching customer names from ad-hoc sources, product taxonomy clean-up, automated sic code (industry) attribution, person-party matching, email address and domain name matching as well as USCS calculations

Prepare & index source and reference tables

Source names, countries

Target names, countries

Both organization and site level customer names, countries

Patterns / rules

Clean and standardize business names

Build standard name based on patterns/rules with regular expression

Standard source names, countries

Standard target names, countries

Find Unique Shortest Common String (USCS)

Add matched customers to the hierarchy

Clusters, Classification, Naïve Bayes

Matched

Unmatched

Process matching

Unique words or string look ups

Distances/Similarities:
Levenshtein
Jaro
Jaro-Winkler
Jaccard
Manhattan

Match results

Unique match?

Validate match results

Weight is calculated based on country, state, and LCS

Calculate weight or confidence

No

Yes

Build clustering or classification

Match result and statistics data