

Web Scrapping

Microsoft Academic

Semantic Scholar

Web of Science

Damerau-Levenshtein  
distance

String similarity

n-gram based similarity  
& distance

String profile

Clean & Merge

Remove duplicates

Remove unrelated  
items