



Lesen der .csv-
Dateien

Übertragung in
DataSet mit
Person-Objekten

Erstellung der n-
Gramme
(Tokens)

Aufnahme der
Tokens in Bloom
Filter

Berechnung der
LSH-Keys

Finden von
Kandidaten-
Paaren

Entfernung von doppelten Kandidaten-
Paaren

Berechnung der
Ähnlichkeiten für
alle Kandidaten-
Paare

Ausgabe aller
Paare mit
hinreichendem
Ähnlichkeitswert