

Namensnormalisierung

Anwendung von Metaphone und Kölner Phonetik

Status: Working Paper | Version 1.0

Publikationsdatum: Mai 2026

Autor: Ronald Voortman

Urheberrecht & Nutzungsbedingungen: Die in diesem Dokument veröffentlichten Forschungsergebnisse, Analysen und methodischen Aufarbeitungen unterliegen dem Urheberrecht von Ronald Voortman (Vortmes - Donkey Books). Die private Nutzung und das Zitieren im wissenschaftlichen Kontext unter Angabe der Quelle sind ausdrücklich erwünscht. Eine Vervielfältigung, Verbreitung oder öffentliche Wiedergabe zu kommerziellen Zwecken bedarf der ausdrücklichen schriftlichen Genehmigung des Autors. Die zugrunde liegenden Primärquellen sind gemäß den Richtlinien der jeweiligen Archivinstitutionen zu verwenden. Für die Korrektheit und Vollständigkeit der Daten wird keine Gewähr übernommen. Der

Inhaltsverzeichnis

ZUSAMMENFASSUNG	3
1. EINLEITUNG	4
2. HISTORISCHE HERAUSFORDERUNGEN	5
2.1 VARIABLE SCHREIBWEISEN	5
2.2 MEHRSPRACHIGER KONTEXT	5
2.3 PATRONYME	5
2.4 UNVOLLSTÄNDIGE DATEN	5
3. DIE DATENGRUNDLAGE	7
4. METHODISCHE GRUNDPRINZIPIEN	8
5. NAMENSNORMALISIERUNG	9
6. TEXTUELLE ÄHNLICHKEIT	10
7. PHONETISCHER VERGLEICH	11
7.1 METAPHONE	11
7.2 KÖLNER PHONETIK	11
8. BERECHNUNG DER ÜBEREINSTIMMUNG	12
8.1 GEWICHTUNG DER MERKMALE	12
8.2 ZEITLICHE KOMPONENTE	12
9. DER MATCHING-PROZESS	13
10. FALLBEISPIELE	14
FALL A: VOERMAN UND VORTMANN	14
FALL B: VELDE UND FELDE	14
FALL C: MIGRATIONSKETTEN	14
11. GRENZEN DER METHODE	15
KEINE AUTOMATISCHE IDENTIFIKATION	15
HÄUFIGE NAMEN	15
FEHLENDE DATEN	15
12. NUTZEN FÜR DIE HISTORISCHE FORSCHUNG	16
13. ZUKÜNFTIGE ENTWICKLUNGEN	17
14. SCHLUSSFOLGERUNG	18

Eine hybride Methode zur Verknüpfung historischer Personenvermerke aus dem Artland (1550–1850)

Zusammenfassung

Historische Personendaten sind häufig über unterschiedliche Quellen und Archive verstreut. Dadurch ist es für Forschende schwierig festzustellen, ob einzelne Erwähnungen tatsächlich dieselbe Person betreffen. Dieses Problem wird durch wechselnde Schreibweisen, mehrsprachige Quellen, Patronyme und unvollständige Informationen zusätzlich erschwert.

Dieses Whitepaper beschreibt eine Methode, die für eine wachsende Datenbank mit derzeit mehr als 4.000 Personenvermerken von Migranten aus dem Artland zwischen etwa 1550 und 1850 entwickelt wurde. Ziel der Methode ist nicht die automatische Identifizierung von Personen, sondern die Unterstützung von Forschenden durch die Sichtbarmachung möglicher Übereinstimmungen zwischen Datensätzen.

Der Ansatz kombiniert Textvergleich, Namensnormalisierung, phonetische Verfahren (Metaphone und Kölner Phonetik) sowie ein gewichtetes Bewertungssystem. Dadurch können potenzielle Verbindungen zwischen historischen Quellen erkannt werden, die mit herkömmlichen Suchmethoden oft verborgen bleiben.

1. Einleitung

Die historische Migrationsforschung wird häufig dadurch erschwert, dass Personen im Laufe ihres Lebens in sehr unterschiedlichen Quellen erscheinen. Ein Migrant aus dem Artland kann beispielsweise in folgenden Quellen auftauchen:

- Registern der Niederländischen Ostindien-Kompanie (VOC)
- Aufgebotsregistern
- Bürgerbüchern
- kirchlichen Attestationen
- Mitgliederverzeichnissen von Gemeinden
- Notariatsakten
- Steuerregistern

Jede Quelle dokumentiert lediglich einen einzelnen Abschnitt eines Lebensweges.

Die zentrale Forschungsfrage lautet daher:

Wie können historische Einträge aus unterschiedlichen Quellen miteinander verknüpft werden, wenn Schreibweise, Sprache und Kontext voneinander abweichen?

Diese Frage bildet die Grundlage der hier vorgestellten Methode.

2. Historische Herausforderungen

2.1 *Variable Schreibweisen*

Bis weit ins 19. Jahrhundert hinein existierte keine einheitliche Rechtschreibung von Personennamen.

Dieselbe Person kann beispielsweise erscheinen als:

- Meyer / Meier
- Voerman / Vortmann
- Velde / Felde
- Dyk / Dieck
- Hake / Haacke

Eine exakte Suche nach Namen übersieht daher zahlreiche potenzielle Übereinstimmungen.

2.2 *Mehrsprachiger Kontext*

Migranten aus dem Artland bewegten sich häufig in einem niederländischen Umfeld.

Deutsche Namen wurden dabei oftmals niederländisiert oder phonetisch von niederländischen Schreibern erfasst.

Dadurch konnten innerhalb verschiedener Quellen stark voneinander abweichende Schreibweisen derselben Person entstehen.

2.3 *Patronyme*

In vielen historischen Quellen finden sich Patronyme.

Beispiele:

- Berend Jans
- Berent Johannis
- Johann Berends

Patronyme stellen wichtige Identifikationsmerkmale dar, weisen jedoch ebenfalls erhebliche Schreibvariationen auf.

2.4 *Unvollständige Daten*

Historische Quellen enthalten häufig nur begrenzte Informationen.

Oft fehlen Angaben wie:

- Geburtsdatum
- Alter
- Beruf
- genauer Herkunftsort

Dadurch muss die Identifikation häufig anhand weniger Merkmale erfolgen.

3. Die Datengrundlage

Die Methode wurde für eine Datenbank entwickelt, die derzeit mehr als 4.000 Personenvermerke aus unterschiedlichen historischen Quellen umfasst.

Der untersuchte Zeitraum erstreckt sich über rund drei Jahrhunderte:

1550–1850

Für jeden Datensatz werden – soweit vorhanden – unter anderem folgende Informationen gespeichert:

- Familienname
- Patronym
- Vorname
- Herkunftsort
- Jahr der Erwähnung
- Quellengattung

Die Datenbank wird kontinuierlich erweitert.

4. Methodische Grundprinzipien

Das System basiert auf drei grundlegenden Prinzipien:

1. Historische Namensvarianten müssen erkannt werden.
2. Die Methode soll nachvollziehbar und transparent bleiben.
3. Die Ergebnisse sind Hinweise, keine automatischen Identifikationen.

Die endgültige Bewertung bleibt stets Aufgabe der Forschenden.

5. Namensnormalisierung

Der erste Schritt besteht in der Normalisierung historischer Schreibweisen.

Viele Unterschiede folgen wiederkehrenden Mustern.

Beispiele:

Variante		Normalisiert		
			th	t
ij	i		oe/ö	o
y	i		ue/ü	u
ei	i		ae/ä	a
ie	i		oh	o
ck	k		ß/sz	s
ph	f		c	k
dt	t			

Dadurch werden viele künstliche Unterschiede zwischen Namen reduziert.

Beispiele:

- Voormann → Vorman
- Dieck → Dik
- Velde → Felde

Die ursprüngliche Schreibweise bleibt erhalten; für die Vergleichsberechnung wird jedoch die normalisierte Form verwendet.

6. Textuelle Ähnlichkeit

Nach der Normalisierung wird die Ähnlichkeit zwischen Namen berechnet.

Dabei wird untersucht, wie stark zwei Zeichenfolgen miteinander übereinstimmen.

Diese Methode eignet sich besonders für:

- kleine Schreibvarianten
- Schreibfehler
- Abkürzungen

Für historische Datensätze allein reicht sie jedoch nicht aus.

Deshalb wird zusätzlich eine phonetische Vergleichsebene eingesetzt.

7. Phonetischer Vergleich

7.1 *Metaphone*

Metaphone ist ein Verfahren, das Wörter in einen phonetischen Code umwandelt.

Namen mit ähnlicher Aussprache erzeugen häufig denselben oder einen sehr ähnlichen Code.

Beispielsweise können:

- Meyer
- Meier

als verwandt erkannt werden, obwohl sie unterschiedlich geschrieben werden.

Metaphone eignet sich besonders für allgemeine Lautähnlichkeiten.

7.2 *Kölner Phonetik*

Für deutschsprachige Namen wird zusätzlich die Kölner Phonetik eingesetzt.

Dieses Verfahren wurde speziell für deutsche Namensstrukturen entwickelt.

Da das Artland historisch Teil eines deutschsprachigen Raums ist, bietet die Methode erhebliche Vorteile.

Die Kölner Phonetik betrachtet nicht die Buchstaben selbst, sondern die zugrunde liegende Lautstruktur eines Namens.

Dadurch können Übereinstimmungen erkannt werden, die bei rein schriftlichen Vergleichen verborgen bleiben.

Wenn zwei Namen nach den Regeln der Kölner Phonetik nahezu identisch klingen, erhält diese Übereinstimmung zusätzliches Gewicht in der Gesamtbewertung.

Gerade bei der Erforschung frühneuzeitlicher Migration erweist sich dies als besonders wirkungsvoll.

8. Berechnung der Übereinstimmung

Das System stützt sich nicht auf einen einzelnen Vergleichswert.

Stattdessen werden mehrere Merkmale kombiniert.

8.1 Gewichtung der Merkmale

Derzeit werden folgende Gewichte verwendet:

Merkmal	Gewicht
Familiennamen	35
Patronym	25
Vorname	20
Herkunftsort	10
Jahr	10

Dadurch erhalten identifizierende Merkmale einen größeren Einfluss als Kontextinformationen.

8.2 Zeitliche Komponente

Der zeitliche Abstand zwischen zwei Erwähnungen wird ebenfalls berücksichtigt.

Ein Unterschied von wenigen Jahren wirkt sich kaum aus.

Größere Zeitabstände führen zu einer niedrigeren Bewertung.

Dadurch wird verhindert, dass Personen mit ähnlichen Namen aus völlig unterschiedlichen Zeiträumen als starke Übereinstimmungen erscheinen.

Gleichzeitig bleibt genügend Spielraum, um lange Lebensläufe und komplexe Migrationswege abzubilden.

9. Der Matching-Prozess

Der Ablauf erfolgt in mehreren Schritten:

1. Auswahl eines Kandidatendatensatzes
2. Normalisierung der Namen
3. Berechnung der textuellen Ähnlichkeit
4. Berechnung der phonetischen Ähnlichkeit
5. Vergleich der Herkunftsangaben
6. Vergleich der Jahresangaben
7. Berechnung einer gewichteten Gesamtbewertung
8. Anzeige potenzieller Übereinstimmungen

Die Endbewertung liegt zwischen 0 und 100.

Nur Datensätze oberhalb eines definierten Schwellenwerts werden den Nutzerinnen und Nutzern angezeigt.

10. Fallbeispiele

Fall A: Voerman und Vortmann

Eine klassische exakte Suche würde diese Namen als unterschiedlich betrachten.

Durch die Kombination von:

- Normalisierung,
- phonetischem Vergleich,
- übereinstimmendem Kontext

entsteht jedoch eine hohe Übereinstimmungsbewertung.

Solche Fälle gehören zu den wichtigsten Anwendungsbereichen der Methode.

Fall B: Velde und Felde

Der Wechsel zwischen V und F kommt in historischen Quellen regelmäßig vor.

Die phonetische Analyse erkennt, dass beide Schreibweisen wahrscheinlich dieselbe Namensform repräsentieren.

Dadurch werden potenzielle Verbindungen sichtbar, die bei manueller Recherche leicht übersehen werden könnten.

Fall C: Migrationsketten

Eine Person kann in mehreren Quellen erscheinen:

1. VOC-Register
2. Amsterdamer Aufgebot
3. Bürgerbuch
4. Lutherische Attestation

Jede Quelle liefert nur einen kleinen Ausschnitt des Lebensweges.

Die Kombination mehrerer Übereinstimmungen kann jedoch auf dieselbe historische Person hinweisen.

11. Grenzen der Methode

Die Methode besitzt wichtige Einschränkungen.

Keine automatische Identifikation

Eine hohe Übereinstimmungsbewertung bedeutet nicht automatisch, dass zwei Datensätze dieselbe Person beschreiben.

Alle Ergebnisse müssen anhand der Originalquellen überprüft werden.

Häufige Namen

Bei Namen wie:

- Meyer
- Janssen
- Schmidt

können mehrere plausible Kandidaten entstehen.

Hier bleibt die Interpretation durch Forschende unverzichtbar.

Fehlende Daten

Wenn wichtige Informationen fehlen, sinkt die Aussagekraft der Berechnung.

12. Nutzen für die historische Forschung

Der größte Mehrwert der Methode liegt nicht in der Automatisierung, sondern in der Unterstützung wissenschaftlicher Arbeit.

Forschende können schneller:

- Migrationsmuster erkennen,
- Lebensläufe rekonstruieren,
- Familienzusammenhänge untersuchen,
- Beziehungen zwischen Quellen sichtbar machen.

Dadurch verlagert sich der Fokus von einzelnen Archivdokumenten hin zu historischen Netzwerken von Personen und Ereignissen.

13. Zukünftige Entwicklungen

Mögliche Erweiterungen der Methode umfassen:

- geografische Distanzberechnungen,
- Netzwerkvisualisierungen,
- Familiengruppierungen,
- maschinelles Lernen als zusätzliche Bewertungsebene,
- Validierung anhand bekannter historischer Fallstudien.

Mit dem weiteren Wachstum der Datenbank wird auch die Qualität der vorgeschlagenen Übereinstimmungen zunehmen.

14. Schlussfolgerung

Die Identifikation historischer Personen stellt insbesondere in Datensätzen mit variabler Schreibweise, Mehrsprachigkeit und unvollständigen Informationen eine erhebliche Herausforderung dar.

Durch die Kombination von Textvergleich, Namensnormalisierung, Metaphone, Kölner Phonetik und einem gewichteten Bewertungssystem entsteht eine praktikable Methode zur Erkennung potenzieller Zusammenhänge zwischen historischen Personenvermerken.

Die Methode ersetzt die historische Forschung nicht. Sie unterstützt vielmehr den Forschungsprozess, indem sie Verbindungen sichtbar macht, die sonst verborgen geblieben wären.

Für die Erforschung der Artländer Migration zwischen 1550 und 1850 eröffnet dieser Ansatz neue Möglichkeiten, einzelne Quellenbelege zu umfassenderen Lebensgeschichten und Migrationsnetzwerken zusammenzuführen.