

# Inhaltsverzeichnis

Einleitung: Paradigmenwechsel im Identity Matching .....	1
Teil I: Grundlagen des linguistischen Identity Matching .....	7
1 Grundkonzepte .....	9
1.1 Identity Matching und Name Matching .....	9
1.2 Datenprofile und Suchabfragen .....	10
1.3 True und False Positives, True und False Negatives .....	12
1.4 Trefferquote und Genauigkeit (Recall und Precision) .....	13
1.5 Linguistisches Identity Matching.....	13
2 Anwendungsfelder.....	17
2.1 Know Your Customer (KYC) und Enhanced Due Diligence (EDD).....	17
2.2 Bekämpfung von Geldwäsche (AML) und Terrorismusfinanzierung (CFT) .....	18
2.3 Customer Data Integration (CDI) und Daten-Deduplizierung .....	20
2.4 Customer Relationship Management (CRM) .....	22
2.5 Kriminalitätsbekämpfung und Strafverfolgung .....	23
2.6 Informationsdienstleistungen.....	24
2.7 Fazit.....	25
3 Grundlegendes zu Personennamen .....	27
3.1 Drei Merkmale von Personennamen: Unterscheidungskraft, Konstanz, Bekanntheitsgrad .....	27
3.2 Personennamensysteme in ihrer historischen Entwicklung .....	28
3.2.1 Rufnamen und Beinamen.....	29
3.2.2 Patronyme und Metronyme.....	30
3.2.3 Vornamen und Familiennamen.....	31
3.3 Personennamensysteme der Welt.....	32
3.3.1 Westliche Personennamen .....	32
3.3.2 Russische Personennamen .....	34
3.3.3 Arabische Personennamen.....	35
3.3.4 Chinesische Personennamen.....	37
3.4 Implikationen für das Name Matching.....	38

4	Transkription.....	41
4.1	Transkription, Transliteration und Translation .....	41
4.2	Romanisierung .....	44
4.3	Romanisierung kyrillischer Namen.....	46
4.3.1	Geschichte und Verbreitung des kyrillischen Alphabets.....	46
4.3.2	Variationsquellen.....	48
4.4	Romanisierung arabischer Namen .....	50
4.4.1	Geschichte und Verbreitung des arabischen Alphabets .....	50
4.4.2	Variationsquellen.....	51
4.5	Romanisierung chinesischer Namen .....	55
4.5.1	Geschichte und Verbreitung der chinesischen Schrift.....	55
4.5.2	Variationsquellen.....	56
4.6	Fazit: Transkription als die Achillesferse des Name Matching .....	58
5	Abgeleitete Namensformen .....	61
5.1	Verniedlichungsformen.....	61
5.2	Namen in Übersetzung .....	63
5.3	Abgeleitete und übersetzte Formen in Namen juristischer Personen ....	65
6	Phonetisches Matchen.....	67
6.1	Homophonie.....	67
6.2	Das Matchen von Homophonen .....	68
7	Tippfehler .....	71
7.1	Begriffliche Abgrenzung: Variationen, Schreibfehler, Tippfehler.....	71
7.2	Motorisch bedingte Tippfehler und die Rolle der Computertastatur.....	72
7.3	Optical Character Recognition (OCR) .....	73
7.4	Fazit: Tippfehler im Name Matching .....	74
	Teil II: Name-Matching-Verfahren .....	75
8	Name-Matching-Verfahren der 1. Generation .....	77
8.1	Einleitung .....	77
8.2	G1 String Comparison: Levenshtein Distance und n-gram .....	78
8.2.1	Ähnlichkeit und Editieroperationen .....	78
8.2.2	Brauchbarkeit der Levenshtein Distance im Name Matching ..	80
8.2.3	Vergleich von Substrings mit n-gram-Verfahren.....	81
8.2.4	Brauchbarkeit von n-gram-Verfahren im Name Matching .....	82

8.3	G1 Phonetic Encoding mit Soundex .....	83
8.3.1	Phonetische Similarity Keys.....	83
8.3.2	Brauchbarkeit von Soundex im Name Matching.....	85
8.4	G1-Suche mit Varianten: Thesauri.....	86
8.4.1	Ein Katalog von Namensvariationen.....	86
8.4.2	Brauchbarkeit von Thesauri im Name Matching.....	87
8.5	Brauchbarkeit der G1-Verfahren im Überblick.....	88
8.6	Warum G1-Verfahren heute noch verbreitet sind .....	91
8.6.1	Name Matching als Mitgift .....	91
8.6.2	Strukturprobleme auf Anbieterseite .....	92
8.6.3	Fehlende Expertise auf der Käuferseite.....	93
8.6.4	Fehlen eines normativen Standards.....	94
9	Name-Matching-Verfahren der 2. Generation .....	97
9.1	Einleitung .....	97
9.2	G2 String Comparison: Erweiterungen von Levenshtein und n-gram...97	
9.2.1	Erweiterungen.....	97
9.2.2	Brauchbarkeit von G2 String Comparison im Name Matching..98	
9.3	G2 Phonetic Encoding: Erweiterungen von Soundex .....	99
9.3.1	Erweiterungen.....	99
9.3.2	Brauchbarkeit von G2-Phonetic Encoding im Name Matching 100	
9.4	G2-Suche mit Varianten: Generative Algorithmen .....	101
9.4.1	Konzept.....	101
9.4.2	Anwendungsbeispiele .....	102
9.4.3	Brauchbarkeit generativer Algorithmen im Name Matching .105	
9.5	Brauchbarkeit der G2-Verfahren im Überblick.....	106
9.6	Fazit: Drei Jahrzehnte Name Matching.....	108
10	Name-Matching-Verfahren der 3. Generation .....	111
10.1	Einleitung .....	111
10.2	Grundanforderungen an G3-Verfahren.....	111
10.2.1	Allgemeine Grundanforderungen .....	112
10.2.2	Spezielle Grundanforderungen .....	113

10.3	Multilinguale Similarity Keys für das Matchen von Transkriptionsvarianten und Homophonen .....	116
10.3.1	Komplexität durch Sprachenvielfalt .....	116
10.3.2	Komplexität durch Suchgenauigkeitsstufen.....	117
10.3.3	Komplexität durch Berücksichtigung des Zeichenkontextes ..	118
10.4	Thesauri für Vornamensformen und Spezialfälle .....	119
10.5	Generative Algorithmen für Tippfehler.....	120
10.6	Integration der Verfahren .....	122
10.7	Fazit.....	126
11	Benchmarkstudie: Die Verfahren im Vergleich.....	129
11.1	Datengrundlage und Testnamen .....	129
11.2	Verfahren und Versuchsbedingungen .....	130
11.3	Vorgehen und Ergebnisse .....	132
11.3.1	G1-Verfahren.....	132
11.3.2	G2-Verfahren.....	133
11.3.3	G3-Verfahren.....	135
11.3.4	Limitationen .....	136
11.3.5	Schlussfolgerungen .....	137
Teil III: Bereit für den Paradigmenwechsel .....		139
12	G3 Name Matching und Identity Matching .....	141
12.1	Raumbezogene Identitätsattribute.....	141
12.1.1	Länderdaten: Nationalität, Geburtsland, Gründungsland .....	142
12.1.2	Oikonyme: Namen von Städten, Stadtteilen und Ortschaften ..	144
12.1.3	Adressen .....	145
12.2	Zeitbezogene Identitätsattribute .....	146
12.3	Klassifikatorische Identitätsattribute.....	148
12.4	Identifikationscodes.....	150
12.5	Integration der Einzelvergleiche .....	151
12.5.1	Das Filtermodell .....	151
12.5.2	Das Gewichtungsmodell .....	152
12.5.3	Kombinierte Modelle .....	153
12.6	Fazit.....	154

13	Tipps zur Tool-Evaluation .....	157
13.1	Einleitung .....	157
13.2	Erhebung der Anforderungen.....	159
13.3	Long List, Short List und Request for Information .....	160
13.4	Testgegenstand und Testdesign.....	162
13.5	Auswahl der Testdaten und der Test-Queries .....	163
13.6	Vorabstimmung mit dem Anbieter .....	166
13.7	Auswertung .....	168
13.7.1	Trefferquote und Präzision .....	168
13.7.2	Trefferbewertung.....	169
13.7.3	Konfiguration.....	169
13.8	Schlussbetrachtung .....	170
14	The Linguistic Search Standard .....	173
14.1	Die Notwendigkeit eines Suchstandards .....	173
14.2	Die Prinzipien .....	175
14.2.1	Prinzipien 1-6 (Match Level Precise) .....	176
14.2.2	Zusatzprinzipien 7-10 (Match Level Close) .....	177
14.2.3	Zusatzprinzipien 11-13 (Match Level Broad) .....	178
14.3	Der Linguistic Search Standard im Original-Wortlaut .....	178
	Literatur .....	185
	Sachwortverzeichnis .....	187