



Solution Paper

DATEN EFFIZIENT NUTZEN

Einsatzfelder für performante, fehlertolerante Suchalgorithmen
und ihre Mehrwerte für die Business Performance

Vorwort

Datengetriebene Prozesse brauchen saubere Daten. Sichere Entscheidungen brauchen vernetzte Daten. Erfolgreiche Digitalisierung braucht eine integrierte Sicht auf alle Daten.

Digitalisierung ist ein wesentlicher Erfolgsfaktor und betrifft alle Bereiche eines Unternehmens und der öffentlichen Verwaltung. Sie ist mehr als nur die Einführung neuer Technologien. Bereits die Aufrüstung und Stabilisierung bestehender IT-Systeme kann einen großen Beitrag zum Gelingen leisten. Dies gilt auch für vorhandene Datenbanken und datenführende Systeme. Die Implementierung einer fehlertoleranten Ähnlichkeitssuche macht Stammdaten fit für die Digitalisierung und zentral für alle Geschäftsprozesse nutzbar.

Die systemübergreifende Verfügbarkeit von Stammdaten – unabhängig von Datenqualität, Transliteration oder Format – ist hierbei eine zentrale Voraussetzung. Ein wichtiger Schritt auf dem Weg zur Digitalisierung: Mit einem System die Basis für die Vernetzung von Daten schaffen und einen zentralen Stammdaten-Hub für alle Geschäftsprozesse etablieren.

Mit der approximativen Indexierungstechnologie matchmaker leisten wir unseren Beitrag zur Digitalisierung. Gerne begleiten wir auch Sie in Ihre digitale Zukunft. Dieses Solution Paper soll Ihnen einen Überblick geben, wie die High Performance Matching Engine matchmaker arbeitet, welche Bandbreite an Prozessen Sie damit automatisieren können und von welchen Mehrwerten schließlich Ihre gesamte Organisation profitiert. Lernen Sie typische Anwendungsfälle aus großen und mittelständischen Unternehmen sowie der öffentlichen Verwaltung kennen und wie matchmaker dort den effizienten Umgang mit Daten beschleunigt.

Nutzen auch Sie das Potential Ihrer Daten für Ihre erfolgreiche Digitalisierung!

Eine interessante Lektüre wünscht Ihnen



Thilo Torkler

Mitglied der Geschäftsleitung der exorbyte GmbH
Leiter Markt & Vertrieb

Inhalt

Vorwort	2
Inhalt	3
1. Perspektivwechsel für die Nutzung von Stammdaten	4
2. High Performance im Umgang mit Daten	5
2.1 Daten blitzschnell finden (Search)	5
2.2 Daten sauber zuordnen (Match)	5
2.3 Daten sicher erkennen und extrahieren (Detect)	6
3. Der Unterschied: Fehlertoleranz auf allen Ebenen	6
3.1 Semantische Toleranz	6
3.2 Phonetische Toleranz	6
3.3 Toleranz in der Feldzuordnung	7
4. Mehrwerte für die gesamte Business Performance	7
5. 100 % Compliance: Know your Customer (KYC)	9
5.1 Geldwäsche und Terrorismusfinanzierung	9
5.2 PEP- und Sanktionslisten	10
5.3 Betrugsprävention	11
6. Öffentliche Sicherheit und Daseinsvorsorge	11
6.1 Grenzsicherheit	12
6.2 Gewährung staatlicher Leistungen	12
6.3 Registermodernisierung	13
7. Datenqualität im Griff	13
7.1 Dublettenerkennung und -prävention	14
7.2 Adressmanagement	14
7.3 Datenkonsolidierung für verschiedene Systemstrukturen	15
7.4 Neukundengewinnung mithilfe externer Daten	16
8. Dokumentenmanagement	17
8.1 Multi-Channel Input Management	17
8.2 Zusammenhänge zwischen Dokumenten	18
9. 360° Kundensicht	18
10. Produktsuche für Online-Shops: Kunden zielgerichtet führen	19
11. Technologie in Partnerschaft	20
12. Der Weg zu datengetriebener High Performance	21
13. exorbyte GmbH	22
14. Technischer Exkurs	23

1. Perspektivwechsel für die Nutzung von Stammdaten

High Performance Matching Engine

Daten, Daten, Daten – alles dreht sich um Daten. Sie sind die Ressource der Zukunft, der entscheidende Vorteil im Wettbewerb auf allen Märkten. Dabei geht es um viel mehr als nur um das Sammeln von Daten. Nur wer es versteht, aus Daten nachhaltig Werte zu schöpfen, profitiert.

Bei stetig wachsendem Datenvolumen sind Lösungen zum Datenabgleich, die die Datenmenge mithilfe von Filtern sukzessive eingrenzen, zunehmend fehleranfällig. Es gilt, jederzeit alle relevanten Daten effizient zu nutzen. Dafür braucht es den Perspektivwechsel – weg von der Segmentierung der Datenmasse hin zu einer umfassenden Datenbetrachtung. Für die effiziente Nutzung aller verfügbaren Stammdaten ohne

jegliche Vorfilterung hat exorbyte die Technologieplattform matchmaker entwickelt.

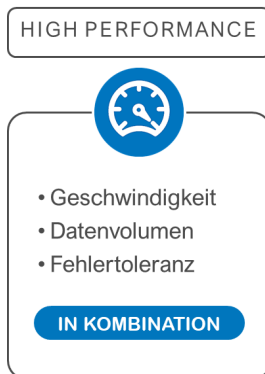


Abb. 1: Performance Parameter

gesucht werden müssen, kommt matchmaker zum Einsatz. Die Anbindung an die jeweiligen Prozesse und Anwendungen erfolgt im Back-End, sodass die datenführenden Systeme einer Organisation oder auch externe Datenquellen in kürzester Zeit vernetzt sind. Strukturierte und unstrukturierte Anfragen gleicht matchmaker gegen strukturierte Referenzdaten ab, welche aus beliebig vielen Datenquellen mit unterschiedlichen Formaten und Strukturen in einem [In-Memory-Index](#) zusammengeführt werden. Auch bei riesigen Datenmengen liefert matchmaker in höchster

Geschwindigkeit ein hochgradig fehler-tolerant ermitteltes, verlässliches Treffer-ergebnis.

Im Kern setzt die Technologie dabei auf die vielseitig nutzbare Implementierung des Levenshtein-Algorithmus und dessen Abwandlungen, z. B. des Damerau-Levenshtein. In Kombination mit weiteren leistungsstarken [Algorithmen](#) für Datumsangaben, Zahlenrelationen, Wortähnlichkeit, Klang-ähnlichkeit, semantische Bezüge u. v. m. versteht matchmaker Daten so dezidiert, wie kein anderes System. Seine außergewöhnliche Fehlertoleranz gewährleistet jederzeit einen umfassenden und unscharfen Abgleich mit allen Daten im Index. Durch die effiziente Ausnutzung von Standard-Hardware-Ressourcen können mehrere Hundert Millionen Datensätze in einem einzigen Server gespeichert und in Sekundenbruchteilen abgefragt werden.

Anwendungsspezifische Geschäftsregeln werden einfach implementiert, indem die für den Abgleich zu berücksichtigenden Datensatzattribute definiert und nach ihrer fallspezifischen Bedeutung für den Abgleich gewichtet werden. Auf Grundlage einer beliebigen Kombination von verfügbaren Attributen zu den einzelnen Entitätstypen – wie typischerweise Personen, Organisationen, Orte und Adressen – führt matchmaker eine [hochpräzise](#) Identity Resolution durch. Da keine Einschränkung der Daten erfolgt, ist auch keine Reihenfolge der Verarbeitung einzelner Teilinformationen spezifiziert. So wird jeder, auch fehlerhafter Eintrag in den Daten gefunden, der nach vorgegebener Spezifikation ähnlich ist. Ganz natürlich ermöglicht die Technologie hierbei einen mehrfach fehlertoleranten Datenzugriff über verschiedene Sprachräume und Schriftsysteme hinweg (z. B. Arabisch, Chinesisch, Kyrillisch etc.) und dies auf äußerst robuste, skalierbare und intuitive Weise.

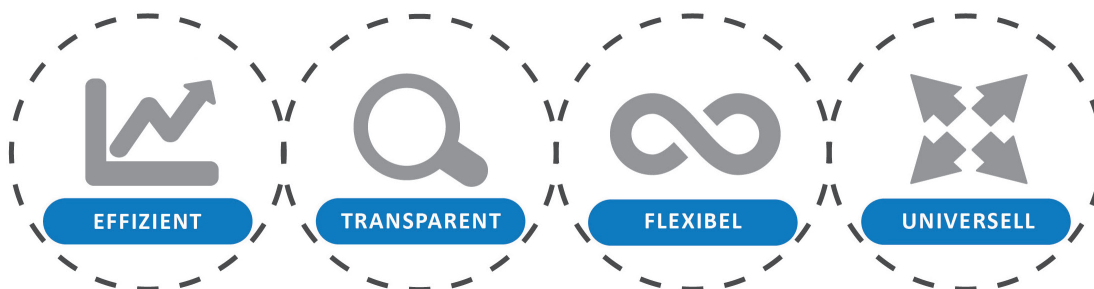


Abb. 2: Grundprinzipien der exorbyte Technologie

2. High Performance im Umgang mit Daten

Search, Match, Detect

Klassische Aufgabenstellungen bzw. Geschäftsprozesse, die mit matchmaker robuster, schneller und damit werthaltiger

werden, sind das Suchen, Zuordnen und Extrahieren von Daten oder eine Kombination dieser Basisprozesse im Umgang mit Daten.

2.1 Daten blitzschnell finden (Search)

Die Stammdatensuche in matchmaker erlaubt ein hochgradig fehlertolerantes Suchen in **Echtzeit** und das in Millionen von Datensätzen – auf Wunsch auch über ein einziges Suchfeld, wie man es von klassischen Internet-suchmaschinen kennt. Dabei kann der Anwender seine Suchworte oder Datenbruchteile ohne Einhaltung von Regeln oder Konventionen frei eingeben. Er trägt in das Suchfeld das ein, was er noch glaubt vom Datensatz zu



kennen, beispielsweise ein Vorname in Kombination mit den ersten beiden Ziffern der Postleitzahl. Sofort präsentiert ihm ein gruppiertes Suggest (individuell konfigurierbar) die relevantesten Datensätze, auch mit feldübergreifenden Kombinationen, aus dem matchmaker Index. Eine automatisch mit der Ergebnismenge generierte Klassifikation der Daten (Facetten) bietet dem Anwender weitere Möglichkeiten zur Eingrenzung der Treffermenge.

2.2 Daten sauber zuordnen (Match)

Datensätze oder auch nur Teile davon gegen Referenzdaten abzugleichen, wird mit matchmaker in gleicher Weise wie bei der Suche ohne eingrenzende Filterung gelöst. Egal, ob wenige oder viele Informationen für das Matching übergeben werden: matchmaker gleicht die Eingabe gegen den kompletten Datenstamm ab und errechnet in rasanter Geschwindigkeit den passenden Datensatz und ggf. auch



dazu ähnliche Datensätze – als Einzelabfrage oder im Batch-Modus.

Es spielt dabei auch keine Rolle, ob es sich um Daten gleicher Struktur, z. B. für die Dublettenerkennung handelt, oder um **Fragmente** davon, wie es bei der postalischen Adressvalidierung häufig der Fall ist. Das Ergebnis ist hochvalide und in wenigen Millisekunden verfügbar.

2.3 Daten sicher erkennen und extrahieren (Detect)

Die matchmaker Stammdatenerkennung extrahiert Identitäten aus Volltexten bzw. Ganzseiten (PDF-Dokumenten, Briefen oder Faxen etc.) und vergleicht diese mit Referenzdaten. Dabei identifiziert matchmaker das gesamte Dokument als Query und zerlegt dieses in Sub-Strings. Jeder einzelne Sub-String wird



hochperformant gegen den gesamten Datenstamm abgeglichen. Damit können z. B. im digitalen Posteingang Dokumente, effizient und sicher automatisiert zugeordnet werden – egal, ob Leseergebnisse der OCR oder Informationen in Datenbanken unerwartete Fehler enthalten.

3. Der Unterschied: Fehlertoleranz auf allen Ebenen

Die Wahrheit erkennen

Meist sind Daten in verschiedenen Systemen gespeichert und von verschiedenen Personen erfasst worden. Sie werden durch vielfältige Technologien übermittelt, sind flüchtig und manche in ihrer Art per se dynamisch. Dies sind nur einige Aspekte, die Daten in ihrer Gültigkeit beeinträchtigen. Die Technologie von exorbyte begegnet dieser

Fehlerhaftigkeit mehrdimensional. Die Fehlertoleranz beschränkt sich im String-Abgleich nicht auf die verschiedenen Arten formaler Fehler (z. B. Tippfehlertoleranz durch den Levenshtein-Algorithmus oder datentypspezifische Zugriffsmethoden), sondern wirkt in mehreren Dimensionen vollumfänglich: semantisch, phonetisch, feldunabhängig.

3.1 Semantische Toleranz

Unter semantischem Matching versteht man den Abgleich formal verschiedener Daten, die auf Bedeutungsebene in einer Identitäts- oder Ähnlichkeitsrelation stehen, d. h. Wörter oder Zeichenketten, die in der äußerlichen Form unähnlich sind, aber das Gleiche oder Ähnliches bedeuten. Synonymbeziehungen und Aliasnamen wie „Heidi“ zu „Adelheid“ oder „Heinz“ zu „Heinrich“ zählen hier dazu und werden von matchmaker äquivalent berücksichtigt. Anwender können dabei auf

umfassendes linguistisches Weltwissen zurückgreifen. Die in matchmaker integrierten Synonym- und Aliaslisten bündeln 20 Jahre Datenexpertise und sind jederzeit durch unternehmensspezifische Vorgaben erweiterbar. Im Rahmen der Analyse der Datenbasis können zudem weitere auf den individuellen Daten basierende Zusammenhänge als linguistische Verknüpfung erkannt werden, für die automatisch generierte semantische Relationen im System hinterlegt werden.

3.2 Phonetische Toleranz

Beim phonetischen Matching werden unterschiedlich verschriftlichte Daten, die im Klang bzw. im Sprachlaut in einer Identitäts- oder Ähnlichkeitsrelation stehen, gegeneinander abgeglichen. Die phonetische Toleranz beschreibt

die Fähigkeit, Wörter oder Namen, die gleich oder ähnlich ausgesprochen, aber anders geschrieben werden, gleichermaßen als ähnlich zu erkennen. Ähnliche Sprachlaute kommen in verschiedenen Sprachen formal sehr unter-

schiedlich zum Ausdruck. Daher ist die phonetische Fehlertoleranz beim Abgleich internationaler Datenquellen besonders relevant.

Das exorbyte Team hat die phonetische Fehlertoleranz zu einer einzigartigen **Vielfalt** ausgebaut. Für einen noch präziseren Abgleich setzt exorbyte neben gängigen phonetischen Transformationen (Kölner Phonetik, Soundex, Metaphone) und Transformationen zur Transliteration aus anderen Schriftsystemen (Devanagari, Hangul, Katakana, Mandarin u. a.) eine hauseigene Phonetik ein.

Herausragende Performance bietet matchmaker in dieser Hinsicht auch durch eine spezielle Kanonisierungsfunktionalität für arabische Namen, wodurch gängige Transkriptionsvarianten abstrahiert werden. So behandelt matchmaker beispielsweise die dem gleichen arabischen Zeichen „ع“ ent-

sprechenden lateinischen Buchstaben bzw. -sequenzen „o“, „u“, „w“, „oo“ und „ou“ als äquivalente Alternativen. „محمد“ wird beispielsweise im angloamerikanischen Sprachraum in der Regel als „Muhammad“ transkribiert, jedoch sind auch die Varianten Mohammed, Mohamed, Muhamed, Muhammed, Muhamet oder Mohammad gängig und in matchmaker abgebildet. Für „عمر“ sind die Varianten „Omar“, „Oumar“, „Umer“, „Umar“ oder „Ömer“ üblich. Die exorbyte Technologie reduziert diese Varianten auf eine normalisierte Schreibweise in lateinischer Schrift, deren Ähnlichkeit mit der buchstäblichen Ähnlichkeit zu einem Suchausdruck variabel verrechnet werden kann. So ist eine Zuordnung auch dann möglich, wenn z. B. bei der Transkription aus dem arabischen Alphabet oder bei der schriftlichen Aufnahme im Kundengespräch eine andere Schreibweise gewählt wurde, als die in den Daten vorhandene.

3.3 Toleranz in der Feldzuordnung

Durch die einzigartige Flexform-Technologie erlaubt matchmaker die variable Zuordnung von (Teil-)Eingaben auf bestimmten Feldern zu anderen Attributen. Sind beispielsweise Vor- und Nachname vertauscht oder die Hausnummer im Straßenfeld enthalten, ermittelt matchmaker trotzdem sicher und schnell den korrekten Suchtreffer.

Die herausragende Geschwindigkeit der **Suchalgorithmen** in matchmaker ermöglicht durch Kombination all dieser Approximationsebenen, sämtliche Datenfelder ohne Voreinschränkung zu durchsuchen. Somit garantiert matchmaker als einzige Software ihrer Art das beste Suchergebnis bei minimalem Zeitaufwand.

4. Mehrwerte für die gesamte Business Performance

Einfach, schnell, sicher

Mit seiner Schnelligkeit, Präzision und Fehlertoleranz steht matchmaker für das beste Ergebnis. Die exorbyte Technologie beschleunigt und verbessert die Automatisierung datengetriebener Prozesse erfahrungsgemäß um ein Vielfaches – ab dem ersten Tag der Inbetriebnahme. Beispielsweise wird die manuelle Nachbearbeitung

des digitalisierten Posteingangs deutlich reduziert, und damit werden wichtige Ressourcen freigesetzt für Aufgaben, die die fachspezifische Expertise der Unternehmen erfordern. Teilweise wurde der Return on Investment in den mit matchmaker neu aufgesetzten Abgleichprozessen bereits in wenigen Monaten erreicht.

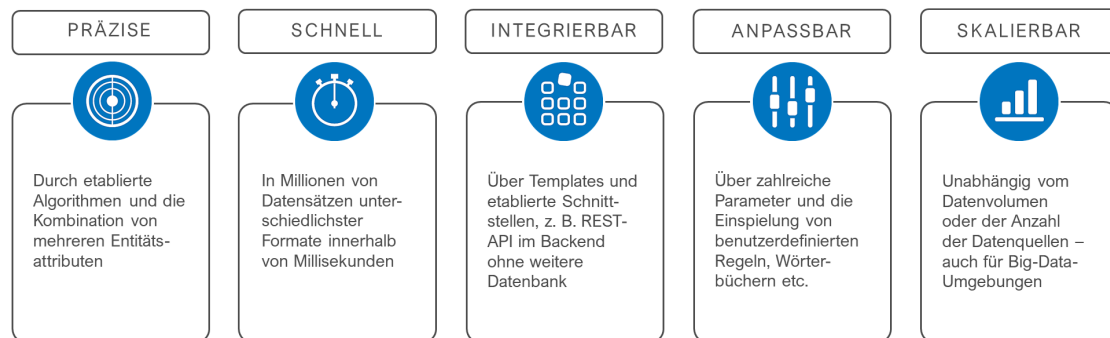


Abb. 3: Vorteile der Technologieplattform matchmaker

matchmaker beschleunigt (parallelisiert) Prozesse, setzt Ressourcen frei, senkt Kosten, minimiert Risiken und stützt so das Kerngeschäft Ihrer Organisation sowie die langfristige Beziehung zu Ihren Kunden.

Der Zeitvorsprung und der Qualitätsaufschwung, die **matchmaker** für die datenbasierten Such- und Abgleichprozesse realisiert, wirken sich positiv auf die gesamte Business Performance aus.

Nachhaltig wirkende Mehrwerte sind u. a.:

- **Schlanke IT**
Eine Technologieplattform versorgt alle datengestützten Geschäftsprozesse.
- **Systemstabilität**
Integration im Back-End der vorhandenen Systemlandschaft in kürzester Zeit
- **Datenverfügbarkeit**
Schneller und sicherer Datenzugriff über einen zentralen Daten-Hub
- **Datenqualität**
Die präventive Dublettenerkennung stärkt die Datenbasis zukunftsorientiert.
- **Datenanreicherung**
Automatisierte Anreicherung mit beliebigen Referenzdaten für noch mehr Geschäftspotenzial
- **Einfache Datenintegration**
Interfaces sorgen für minimale Integrationskosten und eine schnelle Zusammenführung von Daten.
- **Kontaktqualität**
Automatisch normierte und valide Adressen ebnen Kundenkontaktchancen.
- **Compliance-Sicherheit**
Einfache Einhaltung gesetzlicher Compliance-Anforderungen
- **Nutzerakzeptanz**
Einheitliche anwenderfreundliche Suchfunktion in allen Systemen
- **Risikoreduktion**
Minimierung der Fehlerrate (falsche Zuordnung) und daraus resultierender Mehraufwendungen oder Fehlentscheidungen

5. 100 % Compliance: Know your Customer (KYC)

Rechtskonform Geschäfte abschließen

Kennen Sie Ihre Geschäftspartner? Das ist nicht allein eine Frage der Kontaktpflege, sondern verbindliche Sorgfaltspflicht für wirtschaftliche, gesellschaftliche und politische Sicherheit. Geldwäsche kann jeden betreffen – Unternehmen wie Privatpersonen und dies weltweit. Täter zu ergreifen und zu bestrafen, ist Aufgabe des Staates. Um Geldwäsche präventiv zu erkennen, sind alle Beteiligten im globalen Wirtschaftskreislauf gefordert. Rechtssichere Geschäfte bedürfen dazu der Prüfung aller Geschäftspartner, Lieferanten, Kunden und auch Mitarbeiter gegen PEP- und Sanktionslisten. Darin

gelistet: politisch exponierte Personen, terrorverdächtige Personen sowie Personen und Organisationen, für welche weltweit oder länderspezifisch Embargos bestehen. Nicht zuletzt stellt der allgemeine Betrug ein nicht unerhebliches Risiko für Unternehmen dar. Die eindeutige Identifikation von Personen und Organisationen bildet die Basis für rechtssichere Geschäftsbeziehungen.

Inhomogene und dynamische Daten mit intelligenten Algorithmen verarbeiten

5.1 Geldwäsche und Terrorismusfinanzierung

Um rechtssichere Geschäfte abzuschließen – legale Transaktionen mit legal erwirtschaftetem Vermögen – müssen Geschäftspartner sorgfältig analysiert werden. Wer illegal erwirtschaftete Gelder in den legalen Wirtschaftskreislauf einschleust oder auch nur als unwissender oder wissender Mittelsmann Wege dafür eröffnet, macht sich nach deutschem, europäischem und internationalen Recht strafbar. Das Geldwäschegesetz (GWG) nimmt Unternehmen aus dem Finanzsektor besonders in die Pflicht und reguliert Vorsichtsmaßnahmen zur Verhinderung von Geldwäsche und die damit oft verbundene Terrorismusfinanzierung. Verantwortlichkeiten und Prozesse zur Sorgfalt und Aufsicht sind hier organisatorische Pflichtmaßnahmen. Für ein wirksames Risikomanagement ist die frühzeitige Erkennung und Meldung von Verdachtsfällen essentiell. Im Zeitalter der Digitalisierung erfordert dies einmal mehr umfassende, linguistische Datenexpertise sowie die schnelle und automatisierte Vernetzung aller relevanten Daten: öffentlich verfügbare Listen zu politisch exponierten Personen (PEP), Sanktionslisten, interne Blacklists, Transaktionsdaten etc.

Da Geldwäsche grenzüberschreitend wirkt und zum Teil sehr komplexe Transaktionswege nutzt, steht das [Risikomanagement](#) der verpflichteten Unternehmen vor vielschichtigen Herausforderungen. Allein die Mehrsprachigkeit der Datenbanken erweist sich für die Verpflichteten als schwierige Aufgabe. Um die Informationen aus den verschiedenen Datenbanken korrekt korrelieren zu können, sind insbesondere diese sprachlichen Hürden zu überwinden:

- Ungewissheit zur korrekten Schreibweise
- Unterschiedliche Transkriptionsvarianten
- Unterschiedliche Namenssysteme der internationalen Sprachräume
- Unklare Zuordnung von Identitätsattributen wie Titel, Vorname, Nachname

Hinzu kommen formale und prozessuale Schwierigkeiten wie z. B. die Harmonisierung abweichender Datenstrukturen und -formate oder die instantane Bearbeitung aller Datenquellen. Die erfolgreiche Früherkennung impliziert zudem die Installation eines regelmäßigen Bewertungsverfahrens für alle neuen und bestehenden Geschäftspartner.

Mit **matchmaker** prüfen Sie Ihre Anfragen höchst fehlertolerant gegen alle Einträge in Referenzdaten und erhalten Gewissheit über das Vorhandensein oder Nicht-Vorhandensein Ihrer Geschäftspartner. Auch der sprachübergreifende Datensatzabgleich wird mit gleichbleibend hoher Genauigkeit

und Geschwindigkeit durchgeführt – automatisiert, digital und im Back-End Ihrer Systeme. Das Matching kann zwischen verschiedenen Schriftsystemen z. B. Englisch und Arabisch erfolgen. Chinesische oder kyrillische Schriftzeichen versteht **matchmaker** gleichermaßen fehlertolerant.

5.2 PEP- und Sanktionslisten

Geschäftsbeziehungen mit politisch exponierten Personen (PEP) gelten als besonders riskant. Unter PEP werden Personen verstanden, die ein national oder international hochrangiges politisches Amt ausführen oder ausgeführt haben, Familienmitglieder sowie Personen im engen beruflichen Umfeld von hochrangigen Politikern. Aufgrund ihres Einflusses sind diese Personen häufiger als andere in Korruption und Geldwäsche involviert. Nach dem Geldwäschegesetz sind Geschäftspartner auf einen möglichen PEP-Status zu prüfen. Die PEP- und Sanktionslistenprüfung muss dabei insbesondere Probleme aus dem Bereich der Datenqualität lösen:

- **Variabilität**
Sanktionslisten unterliegen ständiger Änderungen.
- **Internationalität**
Es existieren verschiedene ländertypische Verzeichnisse in verschiedenen Schriftsystemen.
- **Umfang**
Sanktionslisten nennen meist nur rudimentäre Angaben zu den gelisteten Personen/Organisationen.
- **Dynamik**
Daten sind von Natur aus sehr dynamisch.
- **Datenstruktur**
Sanktionslisten liegen in unterschiedlichen Formaten vor.

Es reicht daher nicht aus, den potenziellen Geschäftspartner einmalig vor Vertragsabschluss zu überprüfen. Nur ein permanenter Abgleich von Kunden-, Lieferanten- und Partnerdaten, der mit dem ersten Kontakt zu den jeweiligen Personen oder Organisationen greift und kontinuierlich vollzogen wird, bietet Sicherheit. Dabei gilt es aber auch fehlerhafte Zuordnungen auszuschließen, um kein Umsatzpotenzial zu verlieren. Hier ist das Vertrauen in das Finden und Nicht-Finden gleichermaßen von Bedeutung.

matchmaker bietet nachhaltig Sicherheit. Erkennen Sie PEPs einfach und ganz automatisch – auch bei abweichenden Schreibweisen, Sprachen und verschiedenen Transkriptionen zwischen den Sprachen: ob in Echtzeit direkt bei der Dateneingabe, im Posteingang (Scanning) oder als regelmäßige Prüfung aller Kundendaten in der Batch-Variante. Sie identifizieren ungewünschte Geschäftspartner, bevor Sie Ihre Ressourcen in diese Geschäftsbeziehung investieren und Risiken entstehen. Da die Indexierungstechnologie auch feldübergreifend wirkt, erübrigt sich das aufwendige Umformatieren und strukturelle Konsolidieren Ihrer verteilten Daten vor dem Abgleich. Alle Daten werden in ihrer vorliegenden Struktur vollumfänglich ausgelesen und gegen alle relevanten Sperrlisten abgeglichen – bei der Neuanlage und nachfolgend in regelmäßigen Abständen.



5.3 Betrugsprävention

Laut „Zahlungsmoralbarometer 2019 – Deutschland“ des internationalen Kreditversicherers Atradius haben sich die Zahlungsausfälle bei deutschen Unternehmen innerhalb eines Jahres verdreifacht. Gezielter oder gar organisierter Betrug kann hier als eine Ursache vermutet werden. Was für Unternehmen eine existenzielle Notlage bedeuten kann, verursacht auf staatlicher Seite enorme Verluste (Antragsbetrug).

Die Herausforderung: vielfältige, veränderliche Betrugstaktiken schnell erkennen. Eine erfolgreiche Betrugsprävention erfordert klare Compliance-Vorgaben und in der Ausführung schließlich effiziente Prüfprozesse auf Echtzeit-Niveau sowie intelligente Technologien, die auch die raffiniertesten Taktiken aufdecken. Eine intelligente Identitätsprüfung signalisiert schließlich Bestellungen oder Anträge mit zweifelhaften Identitätsdaten (z. B. falsche Personen- oder Adressangaben, gestohlene Kreditkarten- oder Kontodaten oder negative Bonität). Ein möglicher Zahlungsausfall kann mit Hilfe anerkannter Auskunftsteilen evaluiert werden oder anhand interner Blacklists erfolgen, die säumige Zahler oder andere „Betrüger“ umfasst. Um nicht als solch „schlechter Akteur“ identifiziert zu werden, versuchen Betrüger z. B. durch ähnliche, aber nicht dieselben, Informationen erneut Güter oder Leistungen zu beziehen. Deshalb ist sowohl höchste Präzision (keine

oder wenige Fremdübereinstimmungen) als auch Fehlertoleranz beim Abgleich interner und externer Daten gefragt. Wer sich allein auf den Namensabgleich verlässt, lädt Betrüger mit offenen Türen ein. Wirksame Betrugserkennung erfordert die Einbeziehung weiterer [Entitätsattribute](#). Es geht hier auch um Personen, die namentlich zunächst nicht kritisch erscheinen, aber tatsächlich auf nicht so offensichtliche Weise miteinander verbunden sind. Auch in der Ähnlichkeit mit anderen kann ein Betrugsverdacht begründet sein. Ein paar Beispiele dazu: mehrere, ähnliche Versicherungsansprüche für Dienstleistungen desselben Anbieters; Asylanträge, die an dieselbe Telefonnummer gebunden sind; mehrere Rabattanträge von im Wesentlichen derselben Adresse.

Mit [matchmaker](#) gleichen Sie schnell und präzise Entitäten gegen Referenzdaten ab und etablieren in Ihren Geschäftsprozessen eine frühzeitige Betrugserkennung auf höchstem Niveau. Denn matchmaker zeigt Ihnen auch die Ähnlichkeiten zwischen Datensätzen jenseits von Personennamen, indem andere gemeinsame Attribute wie Adressen, Firmen- und Telefonnummern etc. als Beziehung erkannt werden. Verschleierte Betrugsversuche wehren Sie so sicher ab. Statt eines nachgelagerten Prozesses: Prävention in Echtzeit. Dieser Prozess kann vollautomatisch parallel zur Datenerfassung erfolgen.

6. Öffentliche Sicherheit und Daseinsvorsorge

Schützen und unterstützen

Öffentliche Sicherheit und Daseinsvorsorge sind staatliche Aufgaben, die einerseits das Wohl der Gemeinschaft und zugleich das Wohl des Einzelnen, z. B. durch direkte staatliche Leistungen für einen berechtigten Personenkreis, zum Ziel haben. Zur gewissenhaften Ausführung dieser Aufgaben brauchen Sicherheitsbehörden sowie öffentliche

Leistungsträger verlässliche Instrumente zur Klärung der Identität. Sie müssen zweifelsfrei und nachvollziehbar klären, mit wem sie es genau zu tun haben. Die Globalisierung stellt dabei das Anforderungsprofil für ein gutes Prüfsystem: Beherrschung von internationalen Schriftsätzen plus umfassende Datenexpertise.

6.1 Grenzsicherheit

Klarheit über alle Akteure, die die Grenze überschreiten, trägt wesentlich zur Sicherheit bei. Jedoch müssen Sicherheitsbehörden bei der Recherche von Personendaten in verschiedenen Systemen und Listen vielfältige Probleme bewältigen, nicht nur bei der Grenzkontrolle. Nicht selten erweist sich diese Aufgabe für die Beamten als mehrstufig, aufwendig und bisweilen auch als optimierungsbedürftig im Ergebnis.

Es liegt in der Natur der Sache, dass ein großer Prozentsatz der zu überprüfenden Namen international ist. Einige davon sind Transliterationen aus nicht-lateinischen Schriften. Daher erfordert die Suche nach Personennamen für die Grenzsicherheit einen hochpräzisen **Namensabgleich**, der eine Vielzahl von Phänomenen der Namensvarianten fehlertolerant bewältigen kann – und das in kürzester Zeit. Jeden Tag durchqueren Millionen von Reisenden Flughäfen, Seehäfen und andere Einreisehäfen. Die Grenzkontrolle muss so schnell wie möglich, aber auch zuverlässig erfolgen. Neben dem Namen sind typischerweise auch andere Attribute wie Geburtsdatum,

Geburtsort, Nationalität, Organisation und Adresse anhand von Beobachtungslisten zu überprüfen. Dieser Abgleich ist für die Sicherheitsbehörden ein zeitraubender Prozess, da die Datenstrukturen nicht übereinstimmen und die Entitätsattribute in den verschiedenen Kombinationsmöglichkeiten weitere Prüfkriterien ergeben.

Mit **matchmaker** führen Sie einen Namensabgleich mit mehreren Hundert Millionen Datensätzen und Reaktionszeiten in weniger als einer Sekunde durch. Die leistungsstarken Suchalgorithmen liefern ein sauberes **Echtzeit-Screening**. Ihre außergewöhnliche Geschwindigkeit erlaubt es, alle Daten unscharf abzugleichen. Im Ergebnis erhalten Sie Gewissheit, dass alle möglichen, auch sehr feine, Ähnlichkeiten im Namen (zeichentechnisch, phonetisch, semantisch) sowie in den weiteren Entitätsattributen im Suchergebnis aufgezeigt werden. Zusätzlich können den Attributen Gewichte zugewiesen werden, um die am besten geeigneten Geschäftsregeln für die Grenzsicherheit widerzuspiegeln.



6.2 Gewährung staatlicher Leistungen

Staatliche Leistungen sind in den meisten Fällen an Kriterien gebunden, die zum Leistungsbezug berechtigen. Dazu kann die Gewährung einer wiederkehrenden Prüfung unterliegen, die die Berechtigung zyklisch beurteilt. Schließlich handelt es sich um Geld,

das nach dem Solidarprinzip von der Gemeinschaft aufgebracht wird.

Mit Sicherheit
richtig entscheiden

In Arbeits- und Sozialämtern beispielsweise

ist zu klären, ob und welche Leistungen einem Antragsteller zustehen und wie lange. Damit muss der Leistungsträger auch **Antragsbetrug** ausschließen, etwa wenn der Antragsteller die Leistungen bereits über eine andere staatliche Stelle bezieht. Unterschiedliche IT-

Systeme innerhalb der öffentlichen Verwaltung erschweren diese Prüfung. Bei der Erfassung von Antragstellungen sind nicht nur verschiedene Systeme im Einsatz, auch die Erfassung der Daten ist fehleranfällig, sei es durch fehlende Kenntnis zur korrekten Schreibweise, Flüchtigkeitsfehler oder formale Übermittlungsfehler.

Abweichende Schreibweisen von Namen, Adressen oder anderer persönlicher Identitätsmerkmale sind u. a. Gründe für Mehrfachleistungen an ein und dieselbe Person. Nicht selten vermindern Personendubletten und Adressabweichungen die Datenqualität innerhalb des Fachverfahrens, das in einem Länderkreis zur Anwendung kommt. Da-

mit die Leistungsgewährung gerecht und verantwortungsvoll erfolgt, brauchen die zuständigen Ämter intelligente Technologien, die eine hochvalide Identitätsauflösung ermöglichen. Dies schließt effiziente Prozesse, also eine schnelle Fallbearbeitung, mit ein. Eine manuelle Prüfung der Identitätsmerkmale in den verschiedenen Systemen und Registern ist allein bezogen auf die Personalkosten zu kostspielig.

Mit [matchmaker](#) können Sie sich auf eine hochvalide Identity Resolution verlassen. In einem einzigen Suchfeld können Sie alle Ihnen bekannten Daten zu einer Person unstrukturiert

eingeben und sicher sein: matchmaker gleicht jede einzelne Angabe höchst fehlertolerant gegen alle relevanten Referenzdatenbanken ab. Insbesondere die über verschiedene Schriftsysteme anwendbare phonetische Fehlertoleranz macht matchmaker hier zu einem mehrfach wertvollen Tool. Denn neben dem sicheren Finden von Personen mit ausländischen Namen können mit den gleichen Algorithmen Cluster von potenziellen Personendoubletten erzeugt werden. Zusatzinformationen, z. B. zu Familienverbänden werden berücksichtigt und visualisiert, sodass bei der Anlage neuer Daten [Doubletten](#) erkannt und präventive Maßnahmen ergriffen werden können.

6.3 Registermodernisierung

„Moderne Register sind das Fundament besserer Verwaltungsleistungen für Bürger und Unternehmen“, nimmt der Normenkontrollrat der Bundesregierung in einem Gutachten zur Registermodernisierung Stellung zur Bedeutung der Registerlandschaft. Sie sind u. a. die Basis für staatliche Entscheidungen. Die vielfältige Registerlandschaft Deutschlands mit über 200 Registern begründet den mehrstufigen und komplexen Prozess des Registerabgleichs. Schwierigkeiten bereitet dabei auch die Mehrfacherhebung von Daten. Entsprechend kritisch wird die Datenqualität der Register und die Nutzbarkeit der Registerdaten beurteilt.

Mit [matchmaker](#) kann der Registerabgleich unabhängig von der Struktur der Daten und unabhängig von ihrer Qualität im Handumdrehen erfolgen. Durch die Indexierung im Back-End sind alle Systeme einfach integrierbar und alle Daten aus den Registern instantan gebündelt. Über ein einziges Suchfeld können alle verfügbaren Informationen zum Abgleich übermittelt werden. Das fehlertolerante Matching erfasst sprachunabhängig offensichtliche und nicht offensichtliche Verbindungen zwischen Datensätzen und stellt diese nach individueller Gewichtung der Entitätsattribute klar heraus.

7. Datenqualität im Griff

[Matching, Cleansing, Enrichment](#)

Eine gute Datenqualität ist maßgeblich für eine gute Informationsqualität als Basis für die Steuerung von Geschäftsprozessen und für fundierte Entscheidungen. Häufig entstehen „schlechte Daten“ schon bei der Anlage – und das über verteilte Systeme hinweg. In der datengetriebenen, digitalen Wirtschaft

kommen Unternehmen nicht umhin, prozessuale Schwächen in der Stammdatenpflege zu identifizieren, zu beheben und gleichzeitig eine verlässliche Datengüte in allen Quellsystemen herzustellen. Letzteres umfasst die Deduplizierung, Korrektur, Anreicherung, Klassifizierung und Konsolidierung von Stammdaten.

7.1 Dublettenerkennung und -prävention

Die Konsolidierung und Harmonisierung von Stammdaten dient in erster Linie dem effizienten Ablauf von Geschäftsprozessen. Deshalb ist das Denken in Geschäftsprozessen maßgeblich für das Aufsetzen eines Projekts zur Datenbereinigung. Mit einer eigenständigen Organisation und cleveren Software-Tools legen Sie die Voraussetzung für eine erfolgreiche Konsolidierung und Harmonisierung mehrerer Datenquellen.

Von Qualitätsmängeln kann immer dann gesprochen werden, wenn ein oder mehrere der Qualitätsmerkmale, die für einen reibungslosen Geschäftsablauf im Unternehmen gegeben sein müssen, nicht vorliegen. Die Ursachen dafür sind in Unternehmen oft vielschichtig und komplex.

Exzellente Prozessabläufe zeichnen sich durch die permanente Sicherstellung von sauberen Daten aus. Nach der Dublettenbereinigung gilt es, die Datenqualität zu halten. Deshalb ist die Dublettenprävention für **redundanzfreie** Daten das oberste Ziel. Datenerfassungssysteme, z. B. ein CRM, verfügen heute über Lösungen zur Dublettenvermeidung bei der Neuanlage eines

Datensatzes, die auf vorgegebene Parameter beschränkt sind. Weicht der neu anzulegende Datensatz aufgrund einer anderen Schreibweise vom Bestand ab, stößt die systeminterne Dublettenerkennung an ihre Grenzen.

matchmaker findet mit Sicherheit alle Dubletten, die im Laufe der Zeit über Systeme verteilt unbemerkt entstanden sind und führt diese zur Bereinigung einer Dialoganwendung zu. Mithilfe von feldübergreifenden, variablen Feldzuordnungen und feldbezogenen Fehlertoleranzparametern bleiben Ähnlichkeiten zwischen Datensätzen nicht verborgen – selbst bei Fehlschreibweisen, Feldvertauschern oder auch unterschiedlichen Formaten und Strukturen, sodass auch diese Datensätze eindeutig als Dublette identifiziert werden können. Mit seiner mehrdimensionalen Fehlertoleranz steht matchmaker für ein nachhaltig sauberes Ergebnis. Die Matching Engine prüft permanent gegen den Datenbestand aller indexierten Quellsysteme, erkennt potenzielle Dubletten und warnt entsprechend. So vermeiden Sie Dateninkonsistenz.



7.2 Adressmanagement

Fehlerhafte Adressen bedeuten direkte und indirekte Kosten, z. B. beim Mehrfachversand von Briefen an dieselbe Adresse oder durch lange Postlaufzeiten von Rechnungen. Die Fehler in den Adressdaten sind auf eine mangelhafte oder gar fehlende Adressprüfung zurückzuführen. In erster Linie und direkt messbar steigen dadurch die Kosten des Direktmarketings. In der Folge schaden unsaubere Adressdaten auch dem Image. Wer potenzielle Kunden z. B. namentlich falsch anspricht, gilt als unprofessionell.

Die meisten Fehler entstehen bei der Adresserfassung: Straße und Hausnummer werden in einem Feld erfasst; Straßenbezeichnungen

sind veraltet, unvollständig oder schlichtweg falsch; Postleitzahlen enthalten Zahlendreher oder andere Fehler; die Adressangaben sind allgemein nicht post-konform oder gelten als ungültig. Die dadurch entstehenden Kosten für Unternehmen sind hoch. Wer diese Fehler vermeidet, erspart sich zudem den Prozess der **Adressbereinigung**.

Adressdaten sind typischerweise sehr dynamisch. Laut Umzugsstudie 2018 der Deutschen Post ziehen in Deutschland jährlich mehr als acht Millionen Menschen um. Dadurch entstehen Unternehmen nicht nur unzählige Postrückläufer mit den entsprechenden Porto- und Bearbeitungskosten, son-

dern sie verlieren im Jahr viele ihrer Kundenkontakte. Valide Adressdaten sind in großem Maße eine Geschäftsgrundlage. Deshalb ist der Umzugsdatenabgleich bestenfalls als permanenter automatisierter Prozess etabliert.

Mit **matchmaker** sind Sie bzw. Ihre Daten immer auf dem aktuellen Stand. Ihre bisherige Datenbasis wird dadurch grundlegend bereinigt. Im Abgleich mit validen Referenzdaten werden Straßen- und Ortsnamen berichtigt, fehlende Postleitzahlen erzeugt, veraltete Straßennamen ersetzt, Adressen richtig strukturiert und Verstorbene oder Umzüge registriert. Über den Completion-Modus

steht Ihnen gleichzeitig ein Assistent zur Adresserfassung zur Verfügung. Fragmente einer Anschrift genügen zur Eingabe und matchmaker findet die passende Adresse und mögliche Alternativen auf Basis der unscharfen Algorithmen von exorbyte. Der Assistent kann auch in Online-Systeme für die Nutzung durch den Endkunden eingebunden werden. Dadurch erhöhen Sie die Nutzerakzeptanz, verringern die Abbruchraten, reduzieren die Anzahl der Fake-Adressen und verbessern nebenbei auch noch die Qualität der Adressdaten. Das Ergebnis, das auch bei Ihren Kunden ankommt: Ihre Post bzw. Ihre Botschaft.

7.3 Datenkonsolidierung für verschiedene Systemstrukturen

Die eigene Datenbasis aktuell und korrekt zu halten, ist eine permanente Aufgabe für Unternehmen. Heterogene Systemlandschaften, wie sie z. B. in der Versicherungswirtschaft durch strukturell sehr unterschiedliche Vertriebswege und z. T. sehr spezielle Produkte über viele Jahre gewachsen sind, können die Datenkonsolidierung erschweren. Die Datenstrukturen sind mannigfaltig, die Daten selbst ggf. nur dezentral verfügbar – Hindernisse auf dem Weg zur zukunftsorientierten, datengetriebenen Wertschöpfung? Ein zusammenführendes **Migrationsprojekt** bedeutet meist einen weitreichenden Veränderungsprozess, dessen Auswirkungen vorab nicht immer klar bemessen werden können. Erfahrung gemäß sind die Prozesse sehr zeitaufwendig und kostenintensiv, wobei unkalkulierbare Folgekosten eine große Rolle spielen. Dennoch geht die IT-Konsolidierung häufig einher mit dem Ziel der Datenkonsolidierung. Denn die Instandhaltung und Pflege der Systemvielfalt kann erhebliche Ressourcen binden. Zentrale Fragestellungen der Datenqualitätsmanager und IT-Verantwortlichen sind dabei: Wie kann eine gemeinsame Datenbasis geschaffen werden? Welche Systeme sollen weiter-

geführt, zusammengeführt oder gar ersetzt werden? Wie können abweichende Datenstrukturen harmonisiert werden? Wie wird der laufende Betrieb durch die Umgestaltung der Systeme beeinträchtigt?

Am Beispiel eines Zusammenschlusses zweier Unternehmen A und B können mögliche Ansätze zur Zusammenführung unterschiedlicher Datenwelten aufgezeigt werden, die alle von der exorbyte Technologie profitieren:

- **Systemwelt A und B bleiben erhalten.**

Hier gilt es, die Systemvielfalt effizient zu beherrschen durch kontinuierlichen Datenabgleich, automatisches Data Cleansing über alle Systeme sowie die Vernetzung aller Daten für eine 360° Kundensicht.

- **Systemwelt A und B gehen in einem neuen System auf.**

Für die Konsolidierung müssen die Datenstrukturen berücksichtigt werden, um dann durch Matching und Cleansing saubere Daten in ein neues System zu überführen, idealerweise mit automatischer Korrektur von fehlerhaften Einträgen.

■ **Systemwelt A migriert in B (vice versa).**

Die Datenstrukturen sind in ein einheitliches Datenformat zu transformieren. Matching, Cleansing und Enrichment der Datensätze in A um Zusatzinformationen aus B sorgen für eine saubere und sichere Zusammenführung.

Mit **matchmaker** können Datenqualität und Datenverfügbarkeit auch in einer dezentralen und heterogenen Systemlandschaft sichergestellt werden. Etablierte Legacy Systeme werden durch matchmaker für die digitale Zukunft aufgerüstet. Die Technologieplattform sammelt systemübergreifend ein und liefert blitzschnell aus: alle relevanten Daten, die ein Bearbeiter oder ein automatisierter Prozess in seiner täglichen Arbeit braucht.

Aufgrund der umfassenden Suche und der mehrdimensionalen Fehlertoleranz können Sie darauf vertrauen, dass matchmaker alles findet. Und was nicht gefunden wird, ist mit Sicherheit nicht da. Wenn Sie dagegen Ihre dezentral gespeicherten Daten in ein zentrales System überführen möchten, sichert Ihnen matchmaker einen präzisen Abgleich, die automatische Validierung, Korrektur und Bereinigung Ihrer Datenbasis sowie die Anreicherung der Datensätze mit aus allen Systemen zusammengefassten Informationen. Weitere Systeme, die später durch Akquisition oder interne Neustrukturierung dazukommen, können mit wenig Aufwand schnell angebunden werden – in eine gemeinsame konsolidierte Systemwelt oder in die vernetzte, dezentrale Systemwelt.

7.4 Neukundengewinnung mithilfe externer Daten

Zur Neukundengewinnung greifen Unternehmen regelmäßig auf öffentlich verfügbare und durch Adressdienstleister entgeltlich bereitgestellte Personen- und Adressdaten zurück. Ein präziser Abgleich der Daten mit den im Unternehmen vorhandenen Daten empfiehlt sich, um die Kosten für den Dateneinkauf zu minimieren, die **Datenqualität** nicht unnötig zu strapazieren und um nachgelagert Streuverluste bei Marketingaktionen zu vermeiden. Schließlich sollen Neukunden gewonnen und nicht bestehende Kunden oder bereits „verlorene Kontakte“ angesprochen werden. Die Zusammenführung der externen Kontaktdatenbanken und der im Unternehmen vorliegenden Kundendaten erweist sich meist als sehr aufwendig. Große Datenvolumina erschweren den Abgleichprozess, die Datenstrukturen sind in mehrstufigen Prozessen zu harmonisieren. Weitere, selektive Abgleiche nach zielgruppenspezifischen Kriterien (Clustering) bedeuten zusätzliche Investitionen in Zeit und Personalressourcen.

matchmaker gleicht für Sie externe Datenbestände mit Ihren internen Daten fehler tolerant ab, ist sprachsouverän und unabhängig von den einzelnen Feldstrukturen der Datenbanken. Über Gewichtungen für einzelne Attribute können Sie sehr präzise die Daten finden, in denen für Sie echtes Geschäftspotenzial steckt. Für Telekommunikations- oder Versorgungsunternehmen ist beispielsweise besonders interessant, über gleiche Attribute, z. B. eine Telefonnummer, Haushalte und Familienverbände zu erkennen. Damit grenzen sie das relevante Publikum genauer ein und sparen schließlich Marketingkosten. Das fehler-tolerante Matching erzielt zudem signifikant geringere Durchlaufzeiten für die Abgleichprozesse bei deutlich höherer Ergebnisqualität. Sie reduzieren die Kosten für die Neukundengewinnung nachhaltig und öffnen den Weg für die weitere ressourcensparende Automatisierung.

8. Dokumentenmanagement

Automatisch richtig erkennen und verarbeiten

Das papierlose Büro ist noch nicht ganz Realität geworden, doch die geschäftlichen Prozesse basieren heute zum Großteil auf einer elektronischen Dokumentenverwaltung. Die digitale Dokumentation der

Geschäftsvorgänge ist die Basis für eine schnelle Verfügbarkeit prozessrelevanter Daten und damit auch Voraussetzung für die weitere Automatisierung von daten-gestützten Standardprozessen.

8.1 Multi-Channel Input Management

Unternehmen stellen ihren Kunden heute vielfältige Kommunikationskanäle zur Verfügung: von der klassischen Briefpost über das schon altertümlich anmutende und doch in einigen Bereichen noch täglich genutzte Fax, das Telefon bis hin zur digitalen Post per E-Mail – sie alle liefern Daten, die professionell verarbeitet werden müssen. Die Umwandlung der Eingangspost in digitale Dokumente ist ein Weg zur einheitlichen **Datenhaltung**. Häufig werden eingehende Dokumente gescannt, zeitaufwendig klassifiziert, die relevanten Informationen manuell extrahiert und schließlich die Unterlagen der zugehörigen Bearbeitung zugeführt.

Vollwertige Dokumentenbetrachter oder integrierbare Viewer-Lösungen bieten den Anwendern eine einheitliche Sicht auf verschiedene Dokumenttypen und unterstützen die Arbeitsprozesse durch Notizfunktionen und die schnelle Navigation durch das Dokument. Für den Geschäftsprozess relevante Informationen stehen, sofern diese nicht als Meta-Daten vorliegen, nicht zur Verfügung. Risikoakteure, z. B. PEPs werden i. d. R. im Posteingangsprozess nicht automatisch erkannt oder haben sich zum Zeitpunkt der Bearbeitung wieder geändert.

Mit **matchmaker** können Sie per OCR-Technologie digitalisierte Dokumente sowie elektronisch übermittelte Daten automatisch gegen Referenzdaten validieren und die benötigten Personen- bzw. Vorgangsinformationen sicher erkennen, extrahieren und nach Ihren unternehmensspezifischen Anforderungen prüfen – auch bei verstreuten (unstrukturierten) und fehlerhaften Informationen (faktische Fehler oder OCR-Lesefehler). So erreichen Sie automatisierte Verarbeitungsraten von über 90 Prozent bei signifikant geringerer Durchlaufzeit. Selbst bei Dokumenten, die mit herkömmlichen Methoden nicht erschlossen werden können, ist **matchmaker** dank ausgefeilter Algorithmen in der Lage, valide Ergebnisse zu liefern. Dazu können falsche Leseergebnisse der OCR automatisch korrigiert werden, so dass Sie **valide** Daten speichern. Im selben Arbeitsschritt können Sie notwendige Abgleichprozesse, z. B. gegen Sanktions- oder PEP-Listen durchführen und die Dokumente optimal routen. Bereits bei der Dokumentöffnung können Sie den Hinweis zu einem Matching mit einer Sperrliste oder anderen Regularien erhalten.



8.2 Zusammenhänge zwischen Dokumenten

In der Versicherungswirtschaft sind mehrere Verträge mit ein und demselben Kunden über verschiedene Versicherungsleistungen eine gängige Situation. Wer einem Anbieter vertraut, schließt gern alle nötigen Versicherungen mit ihm ab. Die **Zuordnung** von Kunden zu Verträgen und ggf. zum zuständigen Bearbeiter muss entsprechend treffsicher sein. Zudem sind in Versicherungsfällen in der Regel mehrere Personen involviert – als Anspruchsberechtigte, Schuldner, Zeugen oder andere Beteiligte. Deshalb wird die Erkennung von dokumentenübergreifenden Beziehungen von Kundendaten oder auch die Verknüpfung von zusammenhängenden Dokumenten als eine wesentliche Arbeitserleichterung und Prozessbeschleunigung gewertet.

Mit **matchmaker** können Sie Inhalte des dokumentierten Sachverhaltes extrahieren und gegen alle verfügbaren Stammdaten und Dokumente abgleichen. Ihre Sachbearbeiter erhalten dazu eine Trefferliste mit Verknüpfungen zu anderen Dokumenten zur selben Fallnummer, anderer ähnlicher Attribute oder zu weiteren relevanten Informationen, wie laufende Verträge des Kunden, der Familienangehörigen oder weiterer abgeschlossener oder derzeit noch offener Vorgänge. Über Trefferwahrscheinlichkeiten qualifiziert matchmaker die Ergebnisse zusätzlich. So kann der Sachbearbeiter alle relevanten Daten und Zusammenhänge bei der Vorgangsbearbeitung berücksichtigen.

9. 360° Kundensicht

Vernetztes Wissen



Verteilte, unvollständige oder unbekannte Informationen erschweren einen ganzheitlichen Blick auf den Kunden. Um eine Kundenanfrage zu bearbeiten, durchlaufen Mitarbeiter lange und verzweigte Prozessketten. Sie durchsuchen verschiedene Systeme und tragen die Teilinformationen zusammen. In vielen Anwendungen wird die Suchfunktionalität des Unternehmens durch verschiedene Schreibweisen von Personen- und Unternehmensnamen aufgrund von Rechtschreibfehlern, Transliterationsvariationen, Spitznamen oder anderen Gründen beeinträchtigt. Die Prozesskosten stehen dabei in keinem akzeptablen Verhältnis zum Output, die **Suchzeit** ist unproduktive Zeit, im Suchergebnis steckt nur ein Teil der Wahrheit und die Wahrscheinlichkeit für Fehlentscheidungen ist schlicht viel zu hoch. Im telefonischen Kundenservice kann diese umständliche Vorgehensweise zum einen

als Inkompetenz interpretiert werden, zum anderen sorgt es beim Anrufer für Unzufriedenheit.

Nachhaltig gute Beziehungen zwischen Kunden und Unternehmen sind die Basis für stetiges Wachstum. Für ein kundenorientiertes Angebot ist es unerlässlich, zu wissen, wer der Kunde ist und welche Erwartungen er haben könnte. Ein professioneller Kundenservice trägt dazu bei, Kunden langfristig zu binden. Eine hohe Datenqualität und die schnelle Verfügbarkeit aller Daten sind dabei der Schlüssel für den kompetenten und zielführenden Kundendialog. Zudem ist die 360° Sicht auf die Gesamtheit der Daten einer Organisation unerlässlich für eine fundierte Entscheidungsfindung.

Von der Kür zur Pflicht: Durch die Europäische Datenschutzgrundverordnung ist eine schnelle Suchmaschine, die sämtliche Daten zu einer Person bereitstellt,

essenziell für alle Unternehmen. Laut Artikel 15 der EU-DSGVO haben Verbraucher das Recht, von Verantwortlichen zeitnah

eine Bestätigung zu erhalten, ob personenbezogene Daten verarbeitet werden. Wenn Daten verarbeitet werden, muss das jeweilige Unternehmen weitere Informationen in Bezug auf die Daten zur Verfügung stellen.

Mit **matchmaker** recherchieren Sie performant, fehlertolerant und interaktiv in Ihren

Daten. Die Suche lässt sich einfach in alle Anwendungen als Einfeld- oder Mehrfeldsuche integrieren. Für Datenschutzanfragen können Sie einfach einen zentralen **Suchindex** aufbauen, in dem die Daten aus den unternehmensweiten Quellen und Systemen zusammengeführt werden. In weniger als einer Sekunde finden Sie heraus, ob eine Person in Ihren Systemen geführt wird – und wenn ja, in welchen. So können Sie umgehend eine Ersteinschätzung an den Antragsteller geben und die Anfrage ohne aufwendige Einzelsuche in Ihren Systemen rasch bearbeiten. Für alle anderen Aufgaben im Kundensupport zählt sich diese einfache und schnelle Suche ebenso aus, weil Sie damit eine integrierte Sicht auf alle Kundendaten schaffen und so den Kunden individuell betreuen können.

ART. 15 EU-DSGVO

(1) Die betroffene Person hat das Recht, von dem Verantwortlichen eine Bestätigung darüber zu verlangen, ob sie betreffende personenbezogene Daten verarbeitet werden [...]

Abb. 4: Auskunftsrecht

10. Produktsuche für Online-Shops

Kunden zielgerichtet führen

Usability und User Experience sind nicht nur Modewörter, sondern zu Standardkriterien für erfolgreiche Websites geworden. Besucher einer Website wollen sich schnell zurechtfinden und zielgerichtet Informationen sammeln. Mit der wachsenden Bedeutung der **Plattformökonomie** steigen auch die Anzahl der Artikel und schließlich das gesamte Datenvolumen rasant an. Eine gute Suchfunktion kommt nicht ohne intelligente Algorithmen aus. Denn das eine Mal können wir uns nicht mehr genau an einen Artikelnamen oder einen Markennamen erinnern, ein anderes Mal wünschen wir gezielt Vorschläge zu einem Themenkomplex, oder wie so häufig geben wir einen Suchbegriff auf die Schnelle falsch ein. Unsere Erwartungshaltung als Nutzer von Websites ist in allen Fällen, dass die Website unsere Eingabe dennoch richtig versteht. Vielleicht versuchen wir noch eine zweite Eingabe, aber wer auch dann nichts findet, verlässt die Seite vermutlich sehr schnell.

matchmaker ist auch für Online-Shops die ideale Suchmaschine, die Kunden wie Shopbetreiber gleichermaßen zufrieden stellt. Auch bei Hunderten Millionen Produkten, falschen Eingaben und mehrsprachigem Nutzerkreis versteht **matchmaker** die Suchanfrage fehlertolerant und liefert in rasanter Geschwindigkeit die richtigen Treffer. Für Shopbetreiber ergeben sich direkt messbare wirtschaftliche Vorteile: Moderne Algorithmen korrigieren fehlerhafte Eingaben des Nutzers, finden suchrelevante Vorschläge und erlauben ein individuelles (marketinggetriebenes) Ranking der Suchergebnisse in einer einfachen oder gruppierten Trefferliste (Suggest). Individuell konfiguriert unterstützt die Suchmaschine mit verkaufsfördernden Funktionen nachweislich die zentralen Key Performance Indicators Ihrer E-Commerce-Plattform: Usability, Conversion, Umsatz.



11. Technologie in Partnerschaft

Technologien kombinieren und Wertschöpfung bereichern

Die Digitalisierung von Unternehmen führt Prozesse zusammen, transformiert Standardprozesse in automatisierte Abläufe und baut technologische Brücken zwischen den Teilprozessen für eine effiziente, lückenlose Wertschöpfung. Als universelle und flexible Indexierungstechnologie, die Daten systemübergreifend bündelt und unternehmensweit zur Verfügung stellt, ist matchmaker eine wertvolle Ergänzung für verschiedenste Technologien der Automatisierung, insbesondere in der Erfassung und Verarbeitung von Stammdaten. Über REST- und native Schnittstellen lässt sich die Technologieplattform auch für prozessspezifische Workflows sehr gut in andere Technologien integrieren – zum Vorteil aller Beteiligten. Der Anwender kann seine Wertschöpfung beschleunigen (integrierte Prozesse) und schont Ressourcen, die eine eigene Anbindung der Technologie erfordern würde. Die Technologiepartner erschließen neue Kundenmärkte durch ganzheitliche Prozesslösungen.

Insbesondere Unternehmen mit intensivem Endkundenkontakt in verschiedenen Kanälen profitieren von diesen integrierten Prozesslösungen, z. B. Versicherungen,

Banken, Energieversorger. Auch bei der Archivierung von großen Mengen papierbezogener Altakten ergeben sich aus diesen Partnerschaften entscheidende Vorteile für eine schnelle Digitalisierung bei gleichzeitiger Datenvernetzung mit den Informationen in den Bestandsystemen.

In diesen Technologien bereichert exorbyte die digitale Wertschöpfungskette in Partnerschaft mit international agierenden Lösungsanbietern:

- **Scan-Technologie**

Automatischer Abgleich, Validierung und Zuordnung digitalisierter/gescannter Papierbelege

- **OCR-Lösungen**

Validierung maschinell eingelesener Dokumente inkl. Anhänge (unstrukturierte Daten) gegen Referenzdatenbanken

- **Dokumentenviewer**

Anreicherung von Dokumenten durch relevante Daten und -zusammenhänge (Verknüpfung zu Dokumenten, Vorgängen, Personen)

12. Der Weg zu datengetriebener High Performance

Passgenaue Lösungen für individuelle Anforderungen



Abb. 5: Phasen der Implementierung von matchmaker

exorbyte bietet Ihnen nicht nur eine außergewöhnlich leistungsstarke Software. Wir sind Datenspezialisten und beraten Sie mit einem interdisziplinären Team aus IT-Experten, Linguisten und Projektmanagern zur Lösung Ihrer Herausforderungen. Wir verstehen uns als Entwicklungspartner für die Umsetzung Ihrer Datenstrategie und begleiten Sie in die Digitalisierung und darüber hinaus. Das volle Potenzial Ihrer datengetriebenen Prozesse lässt sich erfahrungsgemäß in fünf Phasen erschließen (Abb. 5).

Die Optimierung des Datenmanagements für eine effiziente Verfügbarkeit aller Stammdaten kann von Unternehmen zu Unternehmen stark variieren, denn Daten sind vielschichtig und die Rahmenbedingungen und Anforderungen der Unternehmen sehr individuell. Die Datenspezialisten von exorbyte verfügen über eine breite Erfahrung in Datenprojekten in Bezug auf Suche, Abgleich und Erkennung. Überzeugen Sie sich von der umfassenden Datenkompetenz in einem persönlichen Gespräch.

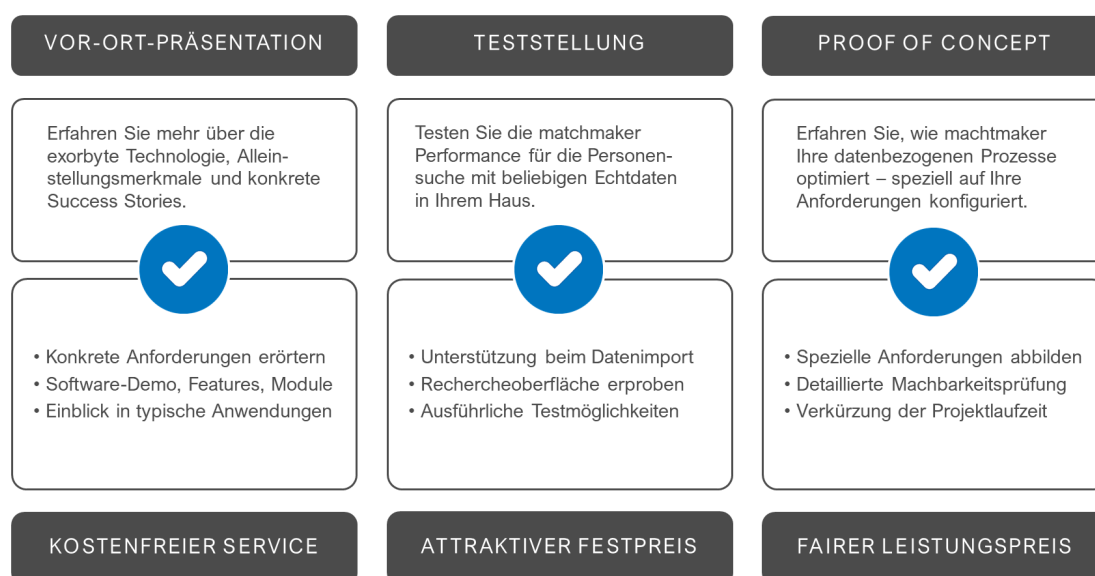


Abb. 6: Möglichkeiten, um matchmaker zu erleben

13. exorbyte GmbH

Partner für die effiziente Nutzung von Stammdaten

Für die hocheffiziente Nutzbarmachung Ihrer Stammdaten entwickelt die exorbyte GmbH mit Sitz in Konstanz am Bodensee die High Performance Matching Engine matchmaker. Die saubere und skalierbare Client-Server-Architektur sorgt für hochverfügbare, flexible und vielseitige Lösungen, die über alle Geschäftsprozesse eine integrierte Sicht auf Ihre Stammdaten bietet.

Die einzigartige Kombination von Geschwindigkeit und Fehlertoleranz von matchmaker ermöglicht die Verbindung von heterogenen Systemlandschaften – unabhängig von Schrift, Sprache, Format und Datenqualität. Unser Alleinstellungsmerkmal ist die besondere Fähigkeit, die zu durchsuchenden Stammdaten nicht beschneiden, einschränken oder vorfiltern zu müssen –

auch match prospecting, blocking oder pruning genannt. Damit ist matchmaker eine echte approximative Indexierungstechnologie, die ein zuverlässiges, direktes und schnelles Finden aller ähnlichen Entitäten zusichert. Zahlreiche Anwendungsfälle zwischen Datenverfügbarkeit, Datenqualität und Dateninput ergeben sich daraus für große und mittelständische Unternehmen.

Versicherungs- und Versorgungsunternehmen, Behörden und Dienstleister der öffentlichen Verwaltung, Finanzinstitutionen, Telekommunikationsdienstleister, Händler und Online-Plattformen setzen beim Umgang mit Stammdaten auf die individuellen Lösungen von exorbyte. Lassen Sie sich zu Ihren datenintensiven Prozessen beraten.

Ihr Ansprechpartner



Thilo Torkler

Lösungsberater

Mitglied der Geschäftsleitung

thilo.torkler@exorbyte.com

+49 176 13633906

14. Technischer Exkurs

Phonetik und Transliterationen

Als wichtiges Hilfsmittel bei der approximativen Suche sind phonetische Codes bekannt, die meistens als alternativer Suchschlüssel verwendet werden. Dies ist also keine Fehlertoleranz, sondern nur ein weiterer Key, wie z. B.:

■ Stemming

Der Wortstamm wird zum Suchen verwendet.

■ N-Gram-Techniken

Zwei- bis fünf-buchstabige Teile werden als Suchschlüssel verwendet, oder aus Anfangsbuchstaben mehrerer Felder zusammengesetzte Schlüssel.

Der Nachteil dieser Techniken ist, dass es sich meistens um eine Äquivalenzrelation handelt. Das heißt zwei Wörter sind „ähnlich“, wenn sie den gleichen Code besitzen. Sie sind „unähnlich“, wenn dies nicht der Fall ist. Das führt zu Ungenauigkeiten an den Rändern der Äquivalenzklassen: Ist ein Wort gerade außerhalb der Klasse, findet man es nicht, ist es gerade noch in der Klasse, ist es ein gleich guter Treffer wie alle anderen in der Klasse auch.

matchmaker nutzt auch **phonetische Codes** – jedoch nur, um die Wortähnlichkeit graduell zu unterstützen. Das heißt auch, der Code selbst wird fehlertolerant gefunden und bewertet; die Gesamtqualität wird durch ein Voting-Verfahren ermittelt. Dabei können bekannte, aber auch speziell von exorbyte entwickelte Kodierungsverfahren verwendet werden.

Der Unterschied zu wirklichen Klang- oder Lautschriften ist, dass ein phonetischer Code in der Regel den String vereinfacht – also abstrahiert, damit er nicht mehr so fehleranfällig ist. Bei einer Lautschrift wird der String mit mehr Informationen spezialisiert. Dies macht es noch schwerer den richtigen Treffer zu finden.

Soundex ist die bekannteste, aber auch eine sehr vereinfachende phonetische Kodierung. Es werden nur der erste Buchstabe und die darauffolgenden ersten drei Konsonanten – letztere nur eingeteilt in sechs Gruppen – verwendet. So wird z. B. „Müller“ zu „M460“. Berücksichtigt werden also nur „M“ und zwei Konsonanten (nach Deduplizierung). Die „0“ füllt den Code auf. Dieser Code ist kein guter approximativer Code, da er immer sehr stark reduziert. Noch deutlicher wird seine Ungenauigkeit an diesem Beispiel: Für die Nachnamen „Meier“, „Mayer“, „Mair“, „Meyr“ ist der Soundex-Code derselbe: „M600“. Allerdings ist dies auch der Soundex-Code für: „Mr.“, „Meer“, „Mara“, „Marr“, „Muro“, „Mirra“ usw.; alles deutsche Nachnamen. Dagegen werden „Meyers“, „Mejer“, „Majer“ und „Mejerl“ nicht in der gleichen Weise codiert. Weitere Vertauscher mit unterschiedlicher Kodierung: „Tapes“ und „Tieves“ haben denselben Code, „Dapes“ und „Dives“ hingegen nicht!

Metaphone ist ein längenabhängiger Code, kennt auch keine Vokale und ist weit verbreitet in mehrsprachigen Anwendungen. Er fängt einige Varianten des Nachnamens „Meier“ besser ab, ist aber auch sehr kurz und enthält Vokale nur in speziellen Verbindungen mit benachbarten Konsonanten. Reibelauten mit ‚h‘ bleiben erhalten, die bei Soundex fehlen.

Die **Kölner Phonetik** ist ähnlich, aber für deutsche Anwendungen optimiert. Auch ihr fehlen die Vokale.

Am häufigsten wird in matchmaker die haus-eigene Phonetik **exoPhone** verwendet, welche an die Art und Weise, wie matchmaker Phonetiken bearbeitet, angepasst ist, d. h. weniger generalisiert und damit auch noch aussagekräftig ist bei „unscharfen“ bzw. fehlerhaften Phonetiktreffern. Diese Kombination macht die phonetische Suche robust und

gewichtet Fehler, die sich phonetisch nicht auswirken, geringer als phonetisch relevante Fehler. Es werden auch Vokalgruppen berücksichtigt. Aus „Müller“ wird so „NOLEL“.

Über das Server-Side-Scripting lassen sich weitere, auch selbst programmierte phonetische Kodierungen anbinden. Durch eine geschickte Einbindung in die Algorithmen von matchmaker hat dies keine Auswirkungen auf die Laufzeit der Abfrage.

Eine spezielle Transformation, die keine echte phonetische Kodierung darstellt, sich aber sehr ähnlich verhält, ist die Kodierung der [Wortgestalt](#). Die Eigenentwicklung von exorbyte ist als SHAPE-Transformation in matchmaker auswählbar. Was die Phonetik für das gesprochene Wort ist, ist die Gestalt für das geschriebene Wort. Beispielsweise gibt in einer OCR-Anwendung die Form der Zeichen eine grobe Beschreibung der Zeichen, ohne die genaue Bedeutung zu kennen. Diese Kodierung verwendet wenige Zeichen, die nur Ober- und Unterlängen, Striche oder Rundungen darstellen, um die Gestalt eines Wortes näherungsweise darzustellen. So wird „Budapest“ zu „Errotoyoooot“.

Durch seinen modularen Aufbau ist matchmaker leicht um zusätzliche Transliterationen erweiterbar. Darüber hinaus unterstützt matchmaker eine Reihe an phonetischen Algorithmen und eine zusätzliche Funktionalität zum phonetischen Abgleich von Transliterationsvarianten aus dem Arabischen.

Standardmäßig verfügbare Transliterationen:

- [Devanagari \(ITRANS\)](#)
- [Hangul \(RR\)](#)
- [Katakana \(Hepburn\)](#)
- [Hiragana \(Hepburn\)](#)
- [Mandarin \(Pinyin\)](#)
- [Cantonesisch \(Pinyin\)](#)
- [Arabisch \(ALA-LC\)](#)
- [Kyrillisch \(ALA-LC\)](#)
- [Griechisch \(UN/ELOT\)](#)
- [Hebräisch \(UNGEGN\)](#)



Herausgeber
exorbyte GmbH
Turmstraße 5
78467 Konstanz

Stand
Februar 2020

Druck
winz.druck, Rielasingen

Gestaltung
Media Lab GmbH, Konstanz

Bildnachweis
istock/Ani Ka (Titelseite)

Copyright
exorbyte GmbH, Konstanz. Alle Rechte vorbehalten.