

5. TOLERANT Whitepaper

Wie man Dubletten im Datenbestand findet und vermeidet

Best Practices und Fallstudien

Kapitel 5-6 aus dem Buch mit dem Titel "Wie man Dubletten im Datenbestand findet und vermeidet" von Klaus-Dieter Sedlacek. ISBN 978-3-384-20706-7 (Abdruck mit freundlicher Genehmigung des Autors)



TOLERANT Software

1. Best Practices und Richtlinien

1.1 Strategien zur laufenden Pflege und Aktualisierung von Personen-Suchfeldern

Die laufende Pflege und Aktualisierung von Personen-Suchfeldern ist ein unverzichtbarer Prozess für Organisationen, um die Integrität, Zuverlässigkeit und Nutzbarkeit ihrer Datenbestände zu gewährleisten. Angesichts der dynamischen Natur personenbezogener Informationen – Menschen ändern Namen, Adressen, Telefonnummern und andere relevante Daten – erfordert die Datenpflege eine kontinuierliche Aufmerksamkeit und die Anwendung bewährter Verfahren. Hier sind einige Best Practices und Richtlinien, die Organisationen dabei unterstützen können, ihre Personen-Suchfelder effektiv zu verwalten:

1. **Regelmäßige Datenüberprüfungen:**
Führen Sie in regelmäßigen Abständen systematische Überprüfungen der Datenbestände durch, um Ungenauigkeiten, Veraltungen und mögliche Duplikate zu identifizieren. Dies kann halbjährlich oder jährlich geschehen, abhängig vom Umfang der Daten und der Häufigkeit von Änderungen.
2. **Automatisierung der Datenaktualisierung:**
Nutzen Sie automatisierte Tools und Softwarelösungen, die die Erkennung von Änderungen in Personen-Suchfeldern unterstützen. Beispielsweise können Datenabgleich-Tools dabei helfen, neue Informationen zu identifizieren und bestehende Datensätze entsprechend zu aktualisieren.
3. **Feedback-Schleifen mit Nutzern:**
Ermöglichen Sie Nutzern oder Kunden, ihre eigenen Daten zu überprüfen und Korrekturen oder Aktualisierungen vorzuschlagen. Dies kann über Online-Portale oder durch regelmäßige Aufforderungen zur Überprüfung der Daten erfolgen.
4. **Schulung der Mitarbeitenden:**
Stellen Sie sicher, dass Ihre Mitarbeitenden in den korrekten Verfahren zur Dateneingabe und -aktualisierung geschult sind. Dies umfasst Richtlinien zur Formatierung, zum Umgang mit Sonderfällen und zur Vermeidung von Fehlern bei der Datenerfassung.

5. **Einführung von Datenqualitätsstandards:**
Definieren Sie klare Standards und Richtlinien für die Qualität von Personen-Suchfeldern. Dies beinhaltet Vorgaben zur Formatierung, zur Vollständigkeit der Daten und zur Nutzung von Standardcodes (z.B. Ländercodes).
6. **Nutzung von Datenbereinigungsdiensten:**
Ziehen Sie die Verwendung externer Datenbereinigungs- und Anreicherungsdienste in Betracht, um die Aktualität und Genauigkeit der Personen-Suchfelder zu verbessern. Solche Dienste können helfen, Adressen zu validieren, Telefonnummern zu aktualisieren und E-Mail-Adressen auf Gültigkeit zu prüfen.
7. **Datenschutz und Compliance:**
Achten Sie darauf, dass alle Maßnahmen zur Datenpflege und -aktualisierung im Einklang mit den geltenden Datenschutzgesetzen und Bestimmungen stehen. Dies schließt die sichere Handhabung personenbezogener Daten und die Gewährleistung der Transparenz und Kontrolle für die betroffenen Personen ein.

Die Einhaltung dieser Best Practices und Richtlinien ermöglicht es Organisationen, ihre Personen-Suchfelder effizient und verantwortungsvoll zu verwalten. Die kontinuierliche Pflege und Aktualisierung dieser Daten trägt nicht nur zur Verbesserung der Datenqualität bei, sondern unterstützt auch eine datengestützte Entscheidungsfindung, optimiert Geschäftsprozesse und fördert das Vertrauen der Nutzer in die Datenverarbeitungspraktiken der Organisation.



1.2 Qualitätskontrolle und Testing von Präparationsmethoden

Die Qualitätskontrolle und das Testing von Präparationsmethoden sind wesentliche Bestandteile im Prozess der Datenbereinigung, die sicherstellen, dass die angewandten Verfahren effektiv zur Verbesserung der Datenqualität beitragen. Diese Schritte dienen dazu, die Konsistenz, Genauigkeit und Zuverlässigkeit der bearbeiteten Daten zu überprüfen und zu validieren, bevor sie für weitere Analysen, Berichte oder operative Prozesse genutzt werden. Hier sind zentrale Aspekte und Richtlinien für die Durchführung einer effektiven Qualitätskontrolle und das Testing von Präparationsmethoden:

1. **Entwicklung von Qualitätsstandards:**
Definieren Sie klare und messbare Qualitätsstandards für die bearbeiteten Daten. Diese Standards sollten sich auf Genauigkeit, Vollständigkeit, Konsistenz und Aktualität der Daten konzentrieren.
2. **Implementierung von Validierungsregeln:**
Setzen Sie automatisierte Validierungsregeln ein, um die Einhaltung der festgelegten Qualitätsstandards zu überprüfen. Dies kann die Überprüfung von Formatierungen, die Einhaltung von Wertebereichen und die Konsistenz von Schlüsselinformationen umfassen.
3. **Testdatensätze:**
Erstellen Sie repräsentative Testdatensätze, die eine breite Palette von Szenarien und potenziellen Datenfehlern abdecken. Diese Datensätze sollten verwendet werden, um die Effektivität der Präparationsmethoden zu testen und zu bewerten.
4. **Durchführung von Testläufen:**
Führen Sie Testläufe mit den Präparationsmethoden durch, um ihre Funktionalität und Effektivität unter verschiedenen Bedingungen zu überprüfen. Analysieren Sie die Ergebnisse, um Bereiche für Verbesserungen zu identifizieren.
5. **Benchmarking und Leistungsvergleiche:**
Vergleichen Sie die Performance der Präparationsmethoden gegenüber festgelegten Benchmarks oder früheren Versionen der Datenbearbeitung. Dies hilft, Fortschritte in der Datenqualität quantifizierbar zu machen.

6. **Feedback-Integration:**
Sammeln und integrieren Sie Feedback von Nutzern und Stakeholdern, die mit den bearbeiteten Daten arbeiten. Nutzerfeedback kann wertvolle Einsichten in die Praxistauglichkeit der Präparationsmethoden bieten.
7. **Dokumentation und Berichterstattung:**
Dokumentieren Sie die angewandten Präparationsmethoden, Testverfahren und Ergebnisse der Qualitätskontrolle. Eine transparente Berichterstattung unterstützt die Nachvollziehbarkeit und das Vertrauen in die Datenbearbeitung.
8. **Kontinuierliche Verbesserung:**
Nutzen Sie die Erkenntnisse aus der Qualitätskontrolle und dem Testing, um die Präparationsmethoden kontinuierlich zu verbessern. Ein iterativer Ansatz ermöglicht es, auf Veränderungen in den Daten oder den Anforderungen flexibel zu reagieren.

Die systematische Qualitätskontrolle und das gründliche Testing von Präparationsmethoden sind entscheidend, um die Integrität und den Wert der Datenbestände zu sichern. Durch die Anwendung dieser Best Practices können Organisationen sicherstellen, dass ihre Datenbereinigungs- und -aufbereitungsprozesse zuverlässige und hochwertige Daten liefern, die als solide Grundlage für Entscheidungsfindungen und operative Prozesse dienen.



1.3 Definition und Relevanz von Personen-Suchfeldern

Der sorgfältige Umgang mit personenbezogenen Daten, insbesondere im Kontext ihrer Bearbeitung und Bereinigung, ist nicht nur eine Frage der ethischen Verantwortung, sondern auch eine rechtliche Notwendigkeit. Datenschutzaspekte spielen eine zentrale Rolle, um die Privatsphäre von Individuen zu schützen und die Einhaltung gesetzlicher Bestimmungen sicherzustellen. Folgende Richtlinien und Best Practices sollten berücksichtigt werden, um Datenschutzaspekte bei der Bearbeitung von personenbezogenen Daten angemessen zu adressieren:

1. **Kenntnis der Datenschutzgesetze:**
Eine gründliche Kenntnis der anwendbaren Datenschutzgesetze und -richtlinien, wie die Europäische Datenschutz-Grundverordnung (DSGVO) oder ähnliche Regelungen in anderen Jurisdiktionen, ist unerlässlich. Diese Gesetze definieren die Rahmenbedingungen für die Sammlung, Verarbeitung und Speicherung von personenbezogenen Daten.
2. **Datenschutz durch Technikgestaltung (Privacy by Design):**
Integrieren Sie Datenschutzprinzipien bereits in der Entwicklungsphase von Datenbearbeitungs- und -bereinigungssystemen. Dies umfasst Maßnahmen zur Datensparsamkeit, Verschlüsselung und Anonymisierung, um die Sicherheit und den Schutz der Daten zu gewährleisten.
3. **Einholung von Einwilligungen:**
Stellen Sie sicher, dass eine gültige Einwilligung von den betroffenen Personen für die Verarbeitung ihrer Daten vorliegt, sofern diese erforderlich ist. Die Einwilligung muss spezifisch, informiert und freiwillig sein.
4. **Transparenz und Informationspflichten:**
Informieren Sie die betroffenen Personen transparent über die Art, den Umfang und den Zweck der Datenverarbeitung sowie über ihre Rechte im Hinblick auf ihre Daten. Dies schließt auch Informationen über Datenbereinigungs- und Aufbereitungsprozesse ein.
5. **Zugriffs- und Berichtigungsrechte:**
Gewähren Sie den betroffenen Personen Zugriff auf ihre gespeicherten personenbezogenen Daten und ermöglichen Sie die Korrektur falscher oder veralteter Informationen.

6. **Datensicherheit:**
Implementieren Sie angemessene technische und organisatorische Sicherheitsmaßnahmen, um personenbezogene Daten vor unbefugtem Zugriff, Verlust oder Beschädigung zu schützen. Dies beinhaltet auch Sicherheitsprotokolle für die Datenbereinigung und -aufbereitung.
7. **Datenminimierung:**
Beschränken Sie die Verarbeitung personenbezogener Daten auf das für den Verarbeitungszweck notwendige Maß. Nicht mehr benötigte Daten sollten sicher gelöscht oder anonymisiert werden.
8. **Dokumentation und Nachweisführung:**
Führen Sie eine sorgfältige Dokumentation über Datenverarbeitungsaktivitäten, einschließlich der Maßnahmen zur Datenbereinigung und -aufbereitung, um die Einhaltung der Datenschutzgesetze nachzuweisen.

Die Beachtung dieser Datenschutzaspekte ist entscheidend, um das Vertrauen der Personen in die Verarbeitung ihrer Daten zu stärken und rechtliche Risiken für die Organisation zu minimieren. Ein proaktiver und verantwortungsvoller Umgang mit personenbezogenen Daten bildet die Grundlage für datenschutzkonforme Datenbearbeitungsprozesse.



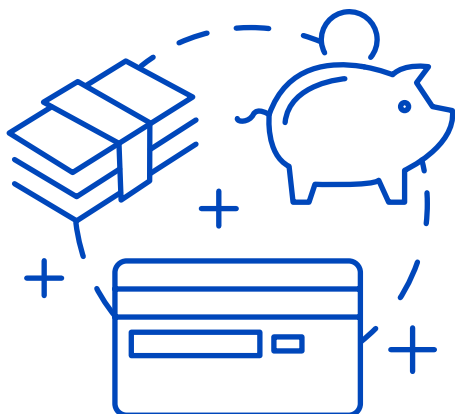
2. Fallstudien und Anwendungsbeispiele

2.1 Erfolgreiche Projekte zur Dublettenvermeidung durch geeignete Präparation

In einer Welt, die von Daten angetrieben wird, ist die Qualität der Informationen, mit denen Unternehmen arbeiten, entscheidend für ihren Erfolg. Ein zentrales Problem, das viele Unternehmen betrifft, ist die Präsenz von Dubletten in ihren Datenbanken. Dubletten können zu ineffizienten Geschäftsprozessen, ungenauen Analysen und letztendlich zu finanziellen Verlusten führen. Daher ist die Vermeidung von Dubletten durch geeignete Präparation von Personen-Suchfeldern ein kritischer Schritt in der Datenverwaltung. Hier beleuchten wir zwei Fallstudien, die die erfolgreiche Umsetzung solcher Projekte demonstrieren.

Fallstudie 1: Ein globales Finanzinstitut

Ein führendes globales Finanzinstitut stand vor der Herausforderung, die Integrität seiner Kundendatenbank zu wahren. Mit Kunden auf der ganzen Welt war die Datenbank des Instituts mit Millionen von Einträgen gefüllt, die zahlreiche Dubletten enthielten. Das Projektteam implementierte eine umfassende Strategie zur Datenpräparation, die eine Normalisierung von Vor- und Nachnamen, die Standardisierung von Adressformaten und die Normalisierung von Kontaktdaten umfasste. Durch den Einsatz von phonetischen Algorithmen zur Dublettenvermeidung konnte das Institut die Anzahl der Dubletten signifikant reduzieren. Das Ergebnis war eine saubere, effizient organisierte Datenbank, die präzisere Kundensegmentierungen und gezieltere Marketingkampagnen ermöglichte.



Fallstudie 2: Ein E-Commerce-Riese

Ein führendes Unternehmen im E-Commerce-Bereich sah sich mit dem Problem konfrontiert, dass seine umfangreiche Kunden- und Produkt-Datenbank von Dubletten durchsetzt war. Dies führte zu Problemen bei der Kundenkommunikation, Lagerhaltung und letztendlich bei der Kundenzufriedenheit. Das Unternehmen setzte auf eine Kombination aus Softwarelösungen und spezifischen Präparationsmethoden für seine Daten. Insbesondere wurde Wert auf die Bereinigung und Konsolidierung von Adresszusätzen und die einheitliche Formatierung von Telefonnummern und E-Mail-Adressen gelegt. Das Projekt führte zu einer deutlichen Verringerung der Dubletten und optimierte die Prozesse der Lagerhaltung und des Kundenservices. Darüber hinaus verbesserte sich die Zielgenauigkeit der Marketingkampagnen, was zu erhöhten Konversionsraten führte.

Diese Fallstudien illustrieren, dass durch die geeignete Präparation von Datenbeständen nicht nur Dubletten vermieden, sondern auch die Effizienz von Geschäftsprozessen gesteigert werden kann. Die Schlüssel zum Erfolg lagen in der sorgfältigen Analyse der vorhandenen Daten, der Auswahl der richtigen Werkzeuge und Methoden für die Datenpräparation sowie in der kontinuierlichen Überwachung und Anpassung der Datenqualitätsstrategien. Durch die Fokussierung auf diese Aspekte können Unternehmen aller Größen und Branchen die Qualität ihrer Datenbestände signifikant verbessern und so einen Wettbewerbsvorteil erzielen.



2.2 Analyse von Problemfällen und deren Lösungsansätze

Die Analyse von Problemfällen im Umgang mit Dubletten in Datenbeständen und die Entwicklung von effektiven Lösungsansätzen sind entscheidende Schritte für Unternehmen, die ihre Datenqualität verbessern möchten. In der Praxis treten verschiedene Szenarien auf, die spezifische Herausforderungen darstellen und individuelle Lösungsstrategien erfordern. Hier beleuchten wir beispielhaft zwei solche Problemfälle und skizzieren die jeweils angewandten Lösungsansätze.

Problemfall 1: Inkonsistente Namenseingaben

Ein häufiges Problem, das die Entstehung von Dubletten begünstigt, sind inkonsistente Namenseingaben. Diese können durch Tippfehler, unterschiedliche Schreibweisen oder die Verwendung von Spitznamen entstehen. Ein Telekommunikationsunternehmen identifizierte dieses Problem, als es feststellte, dass Kunden in seiner Datenbank unter verschiedenen Namen mehrfach geführt wurden, was die Kundenkommunikation und -betreuung erschwerte.

Lösungsansatz:

Zur Lösung dieses Problems führte das Unternehmen eine Reihe von Maßnahmen ein, darunter:

- > Die Implementierung der Namensfeld-Normalisierung, um eine konsistente Schreibweise von Namen zu gewährleisten.
- > Die Einführung einer automatisierten Prüfung auf Tippfehler und häufige Varianten von Namen.
- > Die Anwendung phonetischer Algorithmen, die es ermöglichen, ähnlich klingende Namen zu erkennen und zu konsolidieren.

Problemfall 2: Adressvariationen

Ein weiteres Problem, das zu Dubletten führt, sind Variationen in der Adressangabe. Unterschiedliche Formate, fehlerhafte Postleitzahlen und variierende Schreibweisen von Straßennamen können dazu führen, dass dieselbe Adresse mehrfach in einem Datenbestand vorhanden ist. Ein Online-Händler stieß auf dieses Problem, was zu Schwierigkeiten bei der Auslieferung von Bestellungen führte.

Lösungsansatz:

Um die Herausforderungen im Zusammenhang mit Adressvariationen zu bewältigen, implementierte der Händler folgende Strategien:

- > Standardisierung der Adressformate basierend auf nationalen und internationalen Normen.
- > Einsatz von Adressvalidierungstools, die Eingaben in Echtzeit überprüfen und korrigieren.
- > Integration eines Systems zur Adressbereinigung, das regelmäßig den gesamten Datenbestand durchläuft und Dubletten sowie fehlerhafte Einträge identifiziert und korrigiert.

Diese Lösungsansätze demonstrieren, wie durch die Kombination von technologischen Werkzeugen und klaren Datenmanagement-Strategien Probleme, die zur Entstehung von Dubletten führen, effektiv angegangen werden können. Entscheidend ist dabei nicht nur die Auswahl der richtigen Tools, sondern auch die kontinuierliche Pflege und Überwachung der Datenqualität, um langfristig eine hohe Datenintegrität zu gewährleisten.

Verzeichnis der TOLERANT Softwareprodukte

Die TOLERANT Software GmbH bietet eine Reihe von hochentwickelten Tools an, die darauf ausgelegt sind, Unternehmen in verschiedenen Aspekten des Datenmanagements zu unterstützen. Diese Produkte ermöglichen eine verbesserte Datenqualität, Compliance mit regulatorischen Anforderungen und effizientere Geschäftsprozesse. Im Folgenden wird eine Übersicht über die Kernprodukte von TOLERANT und ihre spezifischen Funktionen gegeben.

1. **TOLERANT Match**

Funktion: Dieses Tool dient der Identifikation und Zusammenführung von Duplikaten innerhalb von Datenbeständen. Es ermöglicht eine effiziente und präzise Datenbereinigung, was die Grundlage für zuverlässige Analysen und Geschäftsentscheidungen bildet.

2. **TOLERANT Post**

Funktion: TOLERANT Post gewährleistet die Korrektheit und Vollständigkeit von Adressdaten. Es unterstützt die Standardisierung und Bereinigung von Adressinformationen, was nicht nur die Kundenkommunikation optimiert, sondern auch die Zustellbarkeit von postalischen Sendungen verbessert.

3. **TOLERANT Bank**

Funktion: TOLERANT Bank prüft Bankdaten wie Kontonummer und Bankleitzahl und korrigiert diese gegebenenfalls. BIC und IBAN werden automatisch ermittelt.

4. **TOLERANT MPM (Marketing Permission Management)**

Funktion: TOLERANT MPM verwaltet die Einwilligungen und Berechtigungen für Marketingaktivitäten und stellt sicher, dass alle Kundenkommunikationen den gesetzlichen Vorgaben und den Wünschen der Kunden entsprechen.

5. **TOLERANT PEP**

Funktion: Ein spezialisiertes Tool zur Identifizierung politisch exponierter Personen (PEPs) in Kunden- und Geschäftspartnerdaten. Es hilft Unternehmen, Compliance-Anforderungen wie die Anti-Geldwäsche-Richtlinien (AML) zu erfüllen.

6. **TOLERANT Name**

Funktion: TOLERANT Name strukturiert Namensbestandteile und generiert daraus korrekte Briefanreden, so dass Kunden stets richtig angesprochen werden.

7. **TOLERANT Sanction**

Funktion: Dieses Tool überprüft Personen und Unternehmen gegen nationale und internationale Sanktionslisten. Es ist entscheidend für die Einhaltung von Compliance-Vorschriften und verhindert Geschäftsbeziehungen mit sanktionierten Einheiten.

8. **TOLERANT Move**

Funktion: TOLERANT Move ermittelt Umzugsdaten von unbekannt verzogenen Empfängern und erkennt verstorbene Personen im Adressbestand.

Diese Produktreihe von TOLERANT deckt ein breites Spektrum an Funktionen ab, von der Datenbereinigung über Compliance-Überwachung bis hin zum Management von Kundenberechtigungen, und bietet damit umfassende Unterstützung für das Datenmanagement in modernen Unternehmen.

