

Software Solution Dixendris AG

Jetzt neue Version verfügbar!
Mit der neuen 64 Bit-Version, können in
weniger als 1 Stunde bis zu 6,5 Millionen
Adressdatensätze bereinigt werden.
Voraussetzungen: Quad-Core Prozessor, 8 GB
RAM Arbeitsspeicher, 64-Bit Betriebssystem

Dateneduplizierungssystem



Problematik

Die meisten Datenquellen enthalten ca. 3-5% doppelte Einträge.

Doppelte Adressen verschlechtern die Kundenbeziehung und verursachen unnötige Kosten.

Lösung

Dixendris d-doop ist ein Hochgeschwindigkeitssystem für das Auffinden von Dubletten in Datenbeständen.

Unabhängig von der Datenart (Adressen, Materialstammdaten, etc.).

Verarbeitung von sehr grossen Datenbeständen (siehe Tabelle auf der Rückseite).

Kontakt

Nehmen Sie Kontakt mit uns auf für eine kostenlose Beratung:

Dixendris AG
Tel. +41 61 272 25 15

www.dixendris.com
info@dixendris.com

Unter www.d-doop.com finden Sie unseren Dublettenrechner!

Dixendris d-doop ist eine performante Lösung für das Auffinden von Dubletten in unterschiedlichsten Datenbeständen, die Ihnen bei der Beseitigung von doppelten Datensätzen zur Seite steht.

Ausgangssituation

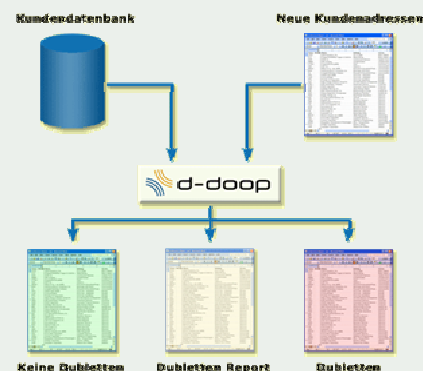
Die meisten Datenquellen enthalten ca. 3-5% oder sogar mehr doppelte Einträge. Doppelte oder auch fehlerhafte Adressen z.B. verschlechtern die Kundenbeziehung und verursachen meist unnötige Kosten, insbesondere im Zusammenhang mit einem allfälligen Briefversand. In grossen Datenmengen derartige Dubletten zu finden und zu entfernen, kann nur mit einer geeigneten Softwarelösung ausgeführt werden.

Dixendris **d-doop** ist eine hoch performante Lösung für das Auffinden von Dubletten in unterschiedlichsten Datenquellen und erlaubt Ihnen das Deduplizieren von Millionen von Datensätzen in kürzester Zeit.

Einsatzfelder

Dixendris d-doop ist ein System, welches sich durch Konfiguration leicht

an Ihre Bedürfnisse anpassen lässt. Es kann als standalone Lösung eingesetzt (Grafik 1) oder einfach in Ihre Systemumgebung integriert werden, um z.B. einen bestehenden Workflow mit der Deduplizierung zu erweitern.



Grafik 1: Einsatzbeispiel
Abgleich neuer Adressdaten mit best. Datenbank



Deduplizierungsablauf

Der Deduplizierungsprozess ist so ausgelegt, dass möglichst wenig Speicher belegt und die Ausführungsgeschwindigkeit sehr schnell ist. **d-doop** unterstützt die gleichzeitige Anbindung unterschiedlicher Datenquellen, so z.B. das Lesen von Daten aus einer Datei mit separierten Feldern oder das Auslesen von Daten aus einer Datenbank mittels JDBC Anbindung. Dabei können auch Datensätze aus verschiedenen Datenquellen zu einem Datenpool zusammengefügt werden. Für jeden einzelnen Datensatz wird ein "digitaler Fingerabdruck" erstellt, der vom FAME-System (Fingerprint Accelerated Matching Engine) für die internen Vergleiche verwendet wird.

Die Datensätze werden hierauf in grobe Cluster aufgeteilt. Clustering ist nicht zwingend notwendig, verkürzt aber die Ausführungszeit dramatisch. Sehr grosse Datenmengen sind ohne geeignetes Clustering kaum zu bewältigen. Nun werden alle potentiellen Duplikat-Kandidaten ermittelt. Mittels einer gewichteten und konfigurierbaren Distanzberechnung wird der Ähnlichkeitswert der Duplikat-Kandidaten bestimmt. Dabei bedeutet ein Wert von 1.0 eine im Sinne der Konfiguration 100%-Übereinstimmung und z.B. 0.75 eine 75% Übereinstimmung. Die ermittelten Duplikate sowie auch die Unikate können dann gesondert verarbeitet werden.

Technologie

FAME (Fingerprint Accelerated Matching Engine) Technologie

Plattformen

Windows
Unix
Mac OS X

Datenbanken

Alle relationalen Datenbanken (RDBMS) werden unterstützt, z.B.: MS SQL, Oracle, MySQL

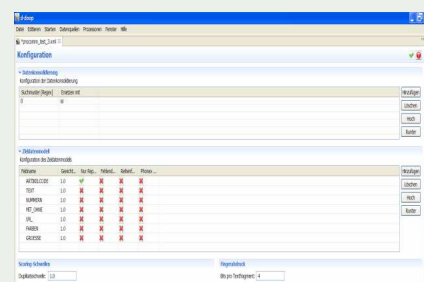
Schnittstellen

MS Excel, MS Access, AVANTI, SAP, EpsiPol, Lotus Notes, Oracle, DB2, Sybase, MS SQL, MySQL

Weitere Produkte: Dixendris AMS®



Approach Management System
CRM
Cockpit
Infodesk
Planning
Marketplace
E2E Bridge



Grafik 2: Benutzeroberfläche **d-doop**

Lösungsbeschreibung

Es gibt zwei Arten von Dubletten, die vollständig identischen und solche die einander bis zu einem gewissen Grad ähnlich sind. Die zweite Art von Dubletten entstehen in der Regel durch Schreibfehler wie z.B. das Verdrehen oder Weglassen von Buchstaben in einzelnen Wörtern. Eine weitere Fehler-

quelle entsteht durch den Zukauf von Adressmaterial von Drittanbietern. Mit Dixendris **d-doop** werden sowohl völlig identische Datensätze wie auch ähnliche Datensätze gefunden. Es beinhaltet also eine unscharfe Suche (Fuzzy Search). Da Ähnlichkeiten einen sehr komplexen Sachverhalt darstellen, ist das Identifizieren von ähnlichen Datensätzen im Gegensatz zum Finden von identischen Datensätzen eine sehr anspruchsvolle Aufgabe. Beim Finden eines solchen Duplikats kann nicht auf eine einfache ja/nein Logik zurückgegriffen werden, sondern es muss ein möglichst aussagekräftiger Ähnlichkeitswert ermittelt werden. Dieser zeigt zu wieviel Prozent z.B. zwei Datensätze einander ähnlich sind. Das Deduplizieren von Datensätzen, bei welchen eine solche Ähnlichkeitssuche angewandt wird, zeigt inhärent eine mit der Anzahl an Datensätzen exponentiell wachsende Ausführungszeit. Deshalb ist eine effiziente Bearbeitung der Datensätze durch clevere, auf Geschwindigkeit optimierte Algorithmen von absoluter Dringlichkeit um eine solche Aufgabe in nützlicher Zeit auszuführen. Dixendris d-doop ist ein Hochgeschwindigkeits-Deduplizierungssystem, das dank seiner einzigartigen FAME (Fingerprint Accelerated Matching Engine) Technologie auch eine grosse Anzahl von Datensätzen innerhalb kürzester Zeit verarbeiten kann. Dank FAME lassen sich pro Sekunde über 85'000'000 Duplikat-Vergleiche durchführen.

Dixendris AG
Binningerstrasse 15
CH - 4051 Basel
Switzerland
T: +41 61 272 25 15
F: +41 61 272 25 16
info@dixendris.com
www.dixendris.com