

TokenBatch – Framework für die maschinelle regelbasierte Verarbeitung und Aufarbeitung von Daten im Batch!

Große Datenmengen lassen sich manuell weder kontrollieren, bewerten noch korrigieren. Egal ob Adressen, Artikel oder sonstige Daten, große Datenmengen erfordern leistungsstarke Funktionen wie unter anderem

- Strukturieren und Formatieren von Daten
- Fehlertoleranter Datenabgleich mit Referenzdaten, Wörterbüchern, etc.
- Verarbeitung von Terminologie- und Synonym-Datenbanken
- Automatisches Prüfen und Korrigieren auf Basis von Regeln
- Dublettenprüfung und –bereinigung

Im Kern befindet ein **TokenAnalyser** (Erkennung) und der **RuleProzessor** (Verarbeitung).

Beispiele, wie es NICHT aussehen sollte:

Usrula, Brititte, J³rgen, Bethovenstraße, Koeln, GmbH, ohG, Ard, Ziegelfuß, 3X80G, GRau, GR.L, Rohr-Verschraubung 12CM, Fittings-NG MUFFE-LOSE 15MM;

Typische (beispielhafte) Regeln:

Groß-/Kleinschriftumsetzung; Leerzeichenoptimierung; Standardisieren von Abkürzungen und technischen Einheiten wie kg, cm, etc.; Zerlegen von Ausdrücken in Begriffen; Änderung von Reihenfolgen von Begriffen; Zeichensatzkonvertierungen; Generieren von Tokenlisten; Dublettenprüfung und –bereinigung; Abgleich mit Referenzen aus Wirtschaft, Technik und weiteren Sachgebieten.

Groß-/Kleinschriftumsetzung; Leerzeichenoptimierung; Standardisieren von Abkürzungen und technischen Einheiten wie kg, cm, etc.; Zerlegen von Ausdrücken in Begriffen; Änderung von Reihenfolgen von Begriffen; Zeichensatzkonvertierungen; Generieren von Tokenlisten; Dublettenprüfung und –bereinigung; Abgleich mit Referenzen aus Wirtschaft, Technik und weiteren Sachgebieten.