

Ihre sprachlichen Altdaten auf Vordermann gebracht!

Sie haben es bisher nicht gewagt, Ihre sprachlichen Altdaten auf den heutigen Stand sprachlicher Qualitätsstandards zu heben? Ihre Produktstammdaten und die dazugehörigen Texte sind durchgängig in Großbuchstaben verfasst, kennen weder Umlaute noch ß oder folgen der alten Rechtschreibung? Leerzeichen sind wahllos gesetzt, Wörter teils abgeschnitten oder willkürlich abgekürzt? Standards für spezielle Datentypen wie Maßangaben, Schraubenbezeichnungen oder Teilenummern werden nicht konsequent eingehalten - ganz zu schweigen von terminologischen oder redaktionellen Vorgaben?

Dann ist es höchste Zeit, Ihre sprachlichen Altdaten einer Verjüngungskur zu unterziehen! Mit unseren linguistischen Analyseverfahren wandeln wir Ihre Altdaten Schritt für Schritt in Texte um, die nicht nur den allgemeinen sprachlichen Qualitätsstandards genügen, sondern auch Ihre terminologischen und redaktionellen Vorgaben berücksichtigen. Ihre Texte sind damit fit für den Einsatz moderner Sprach- und Informationstechnologien wie etwa einem Translation Memory oder einem Authoring Memory.



Datenbereinigung auf linguistischer Grundlage

Leerzeichen
normalisieren

Datentypen
standardisieren

- Maßangaben
- Schraubenbezeichnungen
- Teilenummern u.a.

Strings regel-
basiert ersetzen

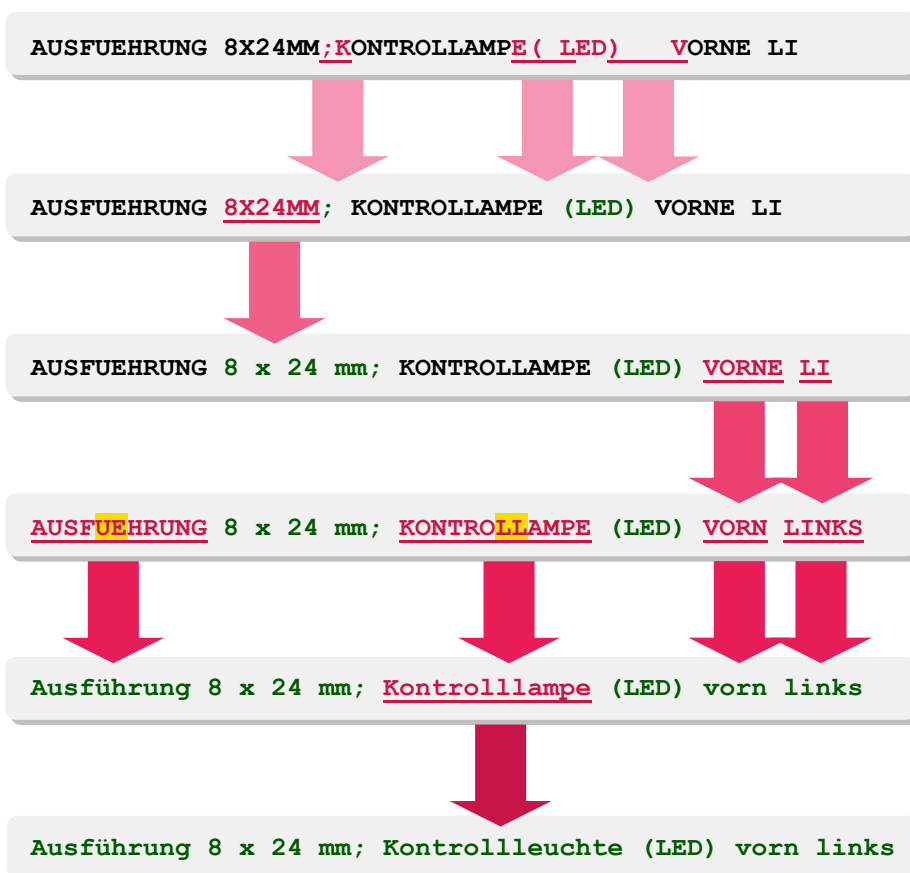
- Schreibvarianten
- Abkürzungen u.a.

Rechtschreibung
korrigieren

- Versalschreibung
- alte Rechtschreibung
- Falschreibungen u.a.

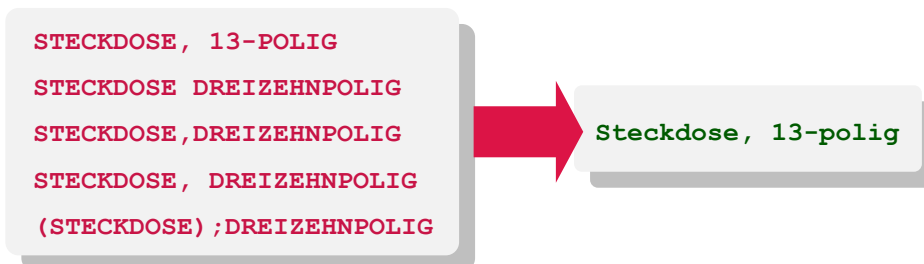
Terminologie
vereinheitlichen

- Negativterme
- Termvarianten

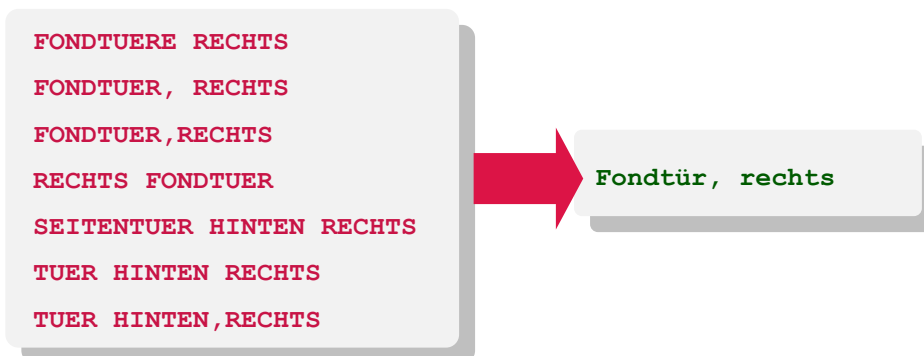


Von der Dublettenreduktion zur Variantenermittlung

Nach der Bereinigung Ihrer Altdaten werden sich bereits viele Texte als Dubletten erweisen. Darüber hinaus können Gruppen ähnlicher Texte ermittelt werden. Die Ähnlichkeitsberechnung orientiert sich dabei zunächst strikt an den Zeichenfolgen der Texte:



Ein erweitertes Verfahren der Ähnlichkeitsberechnung berücksichtigt die linguistischen Wortanalysen und lässt die Wortstellung außer Acht. Terminologische Konzepte können in die Ähnlichkeitsberechnung einbezogen und dabei stärker gewichtet werden als andere Inhaltswörter:



Indem Sie Textähnlichkeiten bereinigen, reduzieren Sie Schritt für Schritt das Volumen der übersetzungsrelevanten Texte. Sie sind damit bestens aufgestellt für den Einsatz eines Translation-Memory-Systems. Und Ihr konsolidierter Textbestand eignet sich hervorragend für die Erstbefüllung eines Authoring Memorys.

Weitere Informationen erwünscht?

Kontaktieren Sie uns mit Ihrem Anliegen zur Bereinigung und Konsolidierung Ihrer sprachlichen Altdaten: Produktstammdaten, Stücklisten, Teilekataloge usw. Wenn Sie spezielle Anforderungen an die Sprachdatenbereinigung haben, können wir Ihnen sicherlich weiterhelfen.

IAI Linguistic Content AG

Martin-Luther-Str. 14
66111 Saarbrücken
0681 38951-0

www.iailc.de
kontakt@iailc.de