

HW-Soft GmbH

Die Spezialisten für Dialog- und Database-Marketing-Systeme

HWsufi library

Die Softwarekomponente zum Suchen und Finden von Adressen, Dubletten und Daten

In vielen Datenbank-Systemen für Auftragsabwicklung, Customer-Relationship-Management (CRM) oder Dialog- und Database Marketing existieren Anschriften, zum Beispiel von Kunden und Interessenten, mehrfach in unterschiedlichen Schreibweisen. Insbesondere, wenn keine Software zur fehlertoleranten Suche und Erkennung von Daten, Namen und Adressen eingesetzt wird, ist dies der Fall.

Häufig ist der gleiche Kunde nur mit unterschiedlichen Ansprechpartnern oder verschiedenen Abteilungsbezeichnungen im Datenbanksystem enthalten.

Bei der Suche von Adressen im Dialog werden diese aufgrund unterschiedlicher Schreibweisen häufig nicht gefunden und daher neu erfasst, obwohl die Adresse bereits in der Datenbank vorhanden ist.

Insbesondere beim Einsatz von Database Marketing Systemen ist es erforderlich, Kunden und Interessenten aus den unterschiedlichsten Quellen mit ihren Zusatzinformationen regelmäßig mit Batchläufen in die Marketing Datenbank zu übernehmen.

Bei all diesen Aufgabenstellungen ist es notwendig und sinnvoll, eine Software einzusetzen, die sowohl bei der Suche im Dialog als auch bei der Suche und Zuordnung im Batch Adressen, Dubletten und Daten effizient und fehlertolerant findet. Dabei ist es wichtig, dass diese Software flexibel einstellbar ist, um die unterschiedlichsten Anforderungen und Aufgabenstellungen erfüllen zu können.



Die Highlights:

- ✓ sichere Erkennung durch umfangreiche Ähnlichkeits- und Näherungsverfahren
- ✓ hohe Performance
- ✓ suchen über bis zu drei Hierarchiestufen
- ✓ umfangreiche Möglichkeiten zur Parametrierung
- ✓ effiziente Gewichtung durch Worterkennungslogik

Die Komponente im Detail

Die Produktfamilie Hwsufi dient zum fehlertoleranten Suchen und Finden von Adressen, von Dubletten und von Daten in Datenbanken und Dateien.

Zurzeit ist Hwsufi für den deutschen Sprachraum verfügbar (weitere Sprachräume auf Anfrage). Zusätzlich zur Hwsufi library steht auch der „Hwsufi manager“ als selbständig ablauffähiges Softwaresystem zur Dublettenerkennung und -bereinigung von Dateien im Batch zur Verfügung.

Für die Ersteinrichtung werden mit einem komfortablen Dialogprogramm die in der Datenbank zu suchenden Informationen definiert.

Mit einem weiteren Programm werden dann einmalig Suchcodes generiert und in einer eigens dafür vorgesehenen Datenbank-Tabelle gespeichert. Diese Suchcodes benötigt die Hwsufi library zum Suchen und Finden von Adressen und Daten.

Die Hwsufi library ist kein selbständig ablauffähiges Softwaresystem. Sie kann in Programme eingebunden werden, die in Programmiersprachen entwickelt wurden, die den Aufruf von C/C++ Funktionen unterstützen. Auch die Implementierung in Internet- und Intranet-Anwendungen ist problemlos möglich.

Über eine Schnittstelle werden der Hwsufi library die zu suchenden Adressen und Daten übergeben. Diese werden dann unter Verwendung von phonetischen Näherungs- und anderen Ähnlichkeitsverfahren mit Hilfe der Suchcodes geprüft. Wenn gleiche oder ähnliche Adressen und Daten gefunden wurden, werden diese an der Schnittstelle mit allen ausgewählten Informationen zurückgeliefert.

Die Referenzdaten

Hwsufi arbeitet mit Referenzdaten, die zur Identifikation von Wörtern dienen, damit diese bei der Erkennung unterschiedlich gewichtet werden können. Dadurch wird insbesondere die Qualität der Dublettenerkennung erheblich verbessert. Das in den Referenzdaten enthaltene umfangreiche Wissen dient zur Erkennung von Titeln, Adelsprädikaten, Vornamen, Nachnamen, Gesellschaftsformen, Branchenbezeichnungen, geografischen Begriffen usw.

Die Referenzdaten sind Bestandteil der Hwsufi library deutsch. Sie werden durch die HW-Soft GmbH laufend aktualisiert und in unregelmäßigem Rhythmus bereitgestellt.

Die Referenzdaten werden für die Nutzung durch Hwsufi von der HW-Soft GmbH speziell aufbereitet und in Form einer Datei ausgeliefert. Diese Datei wird in den Hauptspeicher des DV-Systems komprimiert geladen, auf dem Hwsufi eingesetzt wird. Dadurch kann Hwsufi die Referenzdaten ohne Plattenzugriffe nutzen.

Beispiele

In den folgenden Beispielen zur Dublettenerkennung sind die Suchergebnisse für unterschiedliche Parametereinstellungen dargestellt.

Firmen-Dubletten

CSI Computer + Software Ingenieur-GmbH	70771 Leinfelden	Karlsruher Str. 11/1
Csi Ing.Gesellsch. Computer&Software	70771 Leinfelden	Karlsruher Str. 11

Verlag Karl Hofmann GmbH & Co.	73614 Schorndorf	Steinwasenstr. 6-8
Karl Hoffmann Verlag GmbH & Co	73614 Schondorf	Steinwasenstr. 6-8

Alfons W. Gentner Verlag GmbH & Co. KG	70193 Stuttgart	Forststr. 13
Gentner, A.W. Verlag GmbH & Co. KG	70193 Stuttgart	Forststr. 131

Personen-Dubletten

Wolfgang Winkler	79194 Gundelfingen	Bundesstr. 100
Wolfgang Winter	79194 Guntelfingen	Bundesstr. 100

Lars Schur	73230 Kirchheim	Brunenstr. 39
Bettina Schur	73230 Kircheim	Brunenstr. 39
Gustav Schur	73230 Kirchheim	Brunenstr. 39

Diese Erkennung erfolgt bei der Parametrierung Familientrennung = nein

B. Schur	73230 Kirchenheim	Brunenstr. 39
Bettina Schur	73230 Kirchheim	Brunenstr. 39
Betina Schur	73230 Kirchheim	Brunenstr. 39

Diese Erkennung erfolgt bei der Parametrierung Familientrennung = ja

Hans Link	78076 Villingen	Lichtensteinstr. 89
Dr. Hans Link	78056 Villingen-Schw.	Lichtensteinstr. 83

Diese Erkennung erfolgt bei der Parametrierung Hausnummerntrennung = nein. Wenn die Parametrierung auf Hausnummerntrennung = ja geändert wird, werden diese beiden Adressen nicht als Dublette erkannt

Rudolf Wagner	07980 Berga	Rob.-Guezou-Str. 8
Wagner, Rudi	07980 Berga	Guezou-Str. 8

Firmen- und Personen-Dubletten

H. Hermann GmbH	70327 Stuttgart	Ulmer Str. 300
Heinr. Hermann	70327 Stuttgart	Ulmer Str. 300

Diese Erkennung erfolgt bei der Parametrierung Personen- und Firmen-Trennung = nein. Wenn die Parametrierung auf Personen- und Firmen-Trennung = ja geändert wird, werden diese beiden Adressen nicht als Dublette erkannt.