

Die ADITO Dublettenkontrolle

AID 060 DE



© 2017 ADITO Software GmbH

Diese Unterlagen wurden mit größtmöglicher Sorgfalt hergestellt. Dennoch kann für Fehler in den Beschreibungen und Erklärungen keine Haftung übernommen werden. Wir sind für Feedback zu den Themen, Inhalten, aber auch noch vorhandenen Fehlern dankbar und würden uns freuen, Ihre Meinung zu hören. Die in diesen Unterlagen enthaltenen Daten und Angaben, einschließlich URLs und anderer Verweise können ohne vorherige Ankündigung geändert werden. Alle in diesen Unterlagen aufgeführten Produkt- und Firmennamen sind unter Umständen Marken oder geschützte Zeichen der einzelnen Firmen. Ohne ausdrückliche schriftliche Einverständniserklärung der ADITO Software GmbH darf kein Teil dieses Dokumentes vervielfältigt oder in einer Datenverarbeitungsanlage gespeichert oder in diese eingelesen werden. Diese Einschränkung gilt unabhängig von Art und Weise der Datenerfassung.

Autor: FA, MN, MW, KN. Version 10.3. Zuletzt geändert 05.09.2017

Version	Bemerkung
10.3	Anpassung der Formatierungen
10.2	Anpassen der Bilder
10.1	Hinweis zu MergeOrg und MergePers hinzugefügt
10.0	Anpassung an ADITO4.6, Genauigkeiten erneuert
3.1	Screenshots und Tabelleninhalte aktualisiert
3.0	Praxiseinsatz der Dublettenkontrolle in neuem Kapitel beschrieben technische Umsetzung detaillierter beschrieben

Inhaltsverzeichnis

1.	Einleitung	4
2.	Dublettencheck im ADITO-Basissystem (LUCENE).....	5
2.1.	Einsatz der Dublettenkontrolle	5
2.1.1.	Bei Neuanlage eines Datensatzes	5
2.1.2.	Zur Nachverfolgung bei Datensatz.....	6
2.1.3.	Im Frame Dubletten.....	7
3.	Technische Umsetzung	8
3.1.	Verwendeter Dubletten-Algorithmus: Apache Lucene	8
3.1.1.	Beispiel im Frame Dubletten	8
3.1.2.	Suchgenauigkeit.....	8
3.2.	Suchmuster	8
3.3.	Einsatz.....	9
3.3.1.	Neuanlage von Datensätzen	9
3.4.	Tabellen und Funktionen	10
3.4.1.	Tabelle DUPLICATE.....	10
3.4.2.	Tabelle NODUPLICATE	10
3.4.3.	lib_duplicate	10
3.4.4.	sp_getOrgDuplicate	13
3.4.5.	sp_getPersDuplicate	13
3.4.6.	DUPLICATIONS	13
3.5.	Verwendung der Dublettenlogik in angepassten Systemen	13
4.	Alte Dublettenkontrolle in AO Referenz bis 2012 (SOUNDEX)	14
4.1.	Verwendeter Dubletten-Algorithmus Soundex	14
4.2.	Einsatz.....	14
4.2.1.	Neuanlage von Datensätzen	14
4.2.2.	Dublettenkontrolle bei bestehenden Datensätzen	15
4.3.	Konfiguration der Soundex-Genauigkeit	15
4.4.	Technologie	16
4.4.1.	lib_soundex.....	16
4.4.2.	lib_duplicate	16

1. Einleitung

Jede ADITO-Installation enthält standardmäßig Funktionsbibliotheken, die eine Dublettenkontrolle ermöglichen. Diese Funktionsbibliotheken können so angepasst werden, dass auch Eigenentwicklungen auf diese Funktionen zurückgreifen können.

Standardmäßig wird hier aktuell Apache Lucene verwendet. Diese bietet Suchfunktionalitäten mit ähnlichen Begriffen an, und auf Basis dieser Suchlogik wird nach ähnlich lautenden Firmen oder Personen gesucht.

In diesem Dokument finden Sie in [Kapitel 2](#) eine Beschreibung der fachlichen Umsetzung, in [Kapitel 3](#) die technische Umsetzung.

[Kapitel 4](#) fasst die Berechnung mit Hilfe der SOUNDEX-Logik zusammen, wie sie in alten ADITO-Basissystemen genutzt wurde.

2. Dublettencheck im ADITO-Basisystem (LUCENE)

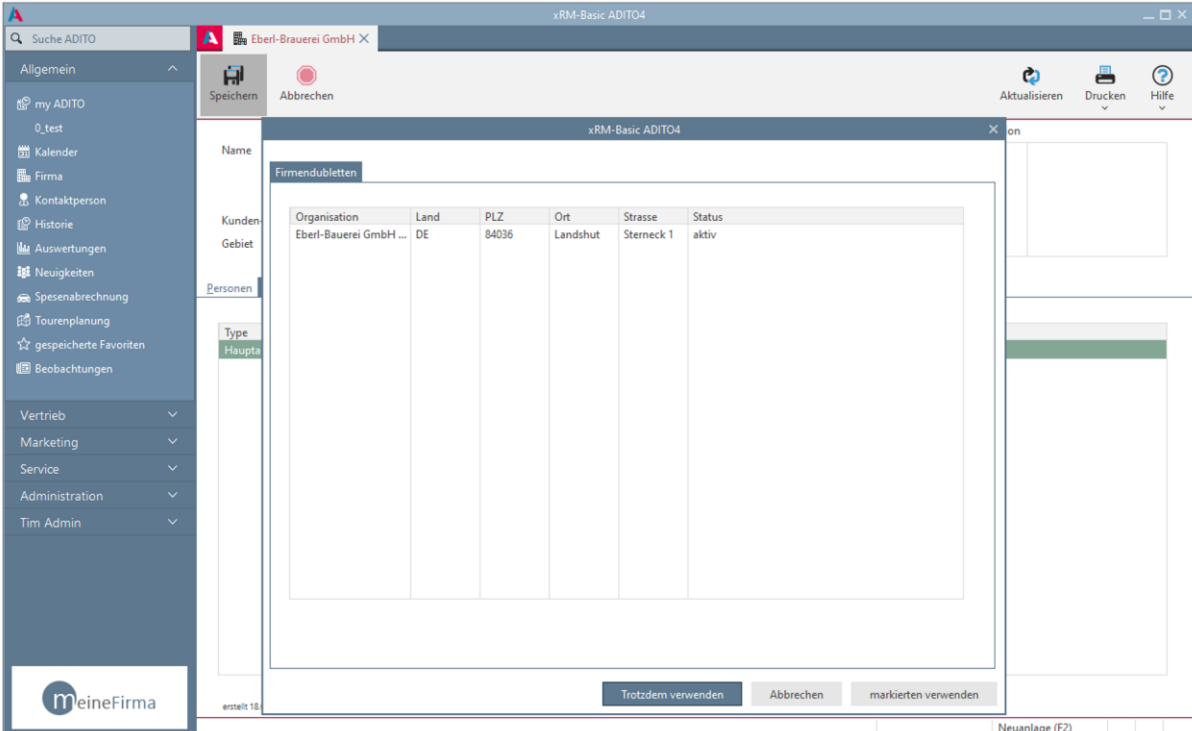
Die folgenden Absätze beziehen sich auf die Dublettenkontrolle, wie sie im Referenzsystem ab AOREferenz 2013* eingesetzt wurde. Gegebenenfalls kann die Logik in Ihrem System abweichen, deaktiviert oder ausgeblendet sein. Fragen Sie im Zweifel Ihren Systemadministrator oder kontaktieren Sie den ADITO-Support.

* AOREferenz 2013, AOREferenz2014, xRM2015, xRM-Basic

2.1. Einsatz der Dublettenkontrolle

2.1.1. Bei Neuanlage eines Datensatzes

Beim Anlegen einer **Person** oder **Firma** wird vor dem Speichern geprüft, ob mögliche Dubletten vorliegen. Greift die Dublettenerkennung hier ein, stehen dem Benutzer mehrere Möglichkeiten zur Verfügung, wie mit dem Datensatz weiter verfahren wird.

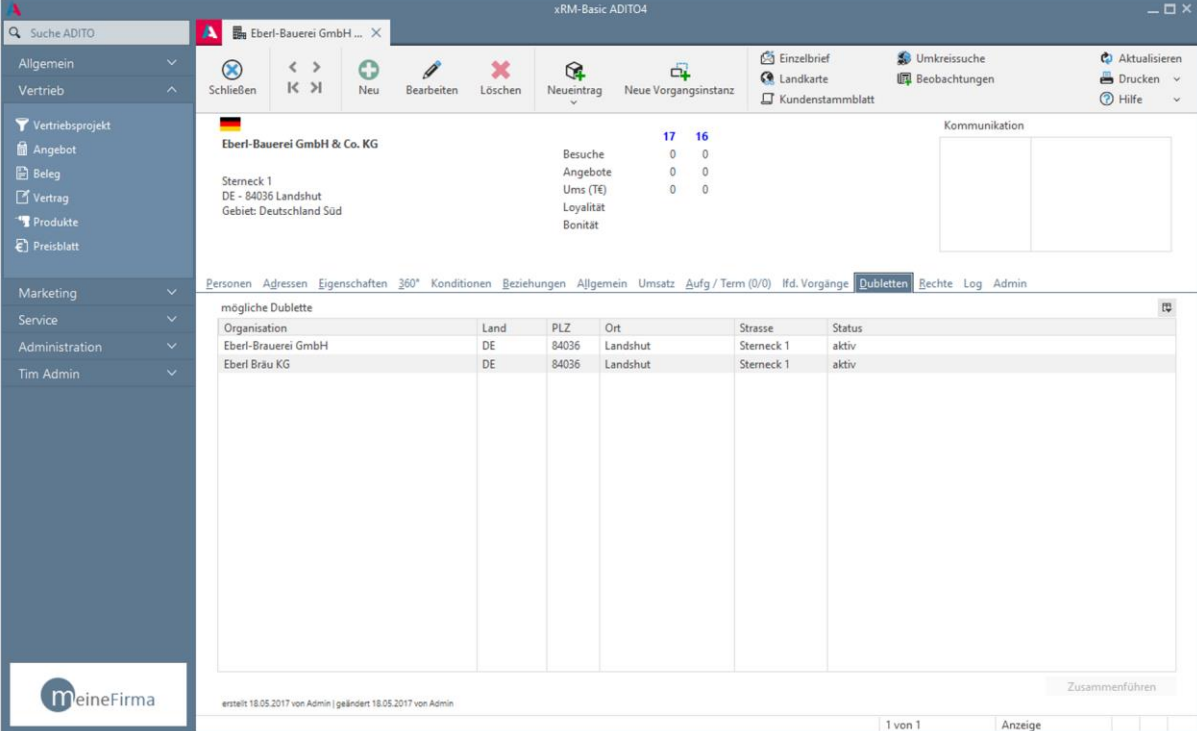


Organisation	Land	PLZ	Ort	Strasse	Status
Eberl-Brauerei GmbH ...	DE	84036	Landshut	Sterneck 1	aktiv

- **Trotzdem übernehmen:** Übernimmt den Datensatz ungeachtet der Dublettenwarnung. Diese Möglichkeit sollte dann genutzt werden, wenn ein ähnlich klingender, aber völlig anderer Datensatz angelegt wird (z.B. "Müller GmbH"). Bei "Trotzdem übernehmen" werden die beiden als Dubletten identifizierten Datensätze auch nicht mehr als Dubletten angezeigt.
- **Markierten verwenden:** Wechselt zur Anzeige der Dublette und verwirft die Neuanlage.
- **Abbrechen:** Wechselt zurück zum Neuanlagemodus.

2.1.2. Zur Nachverfolgung bei Datensatz

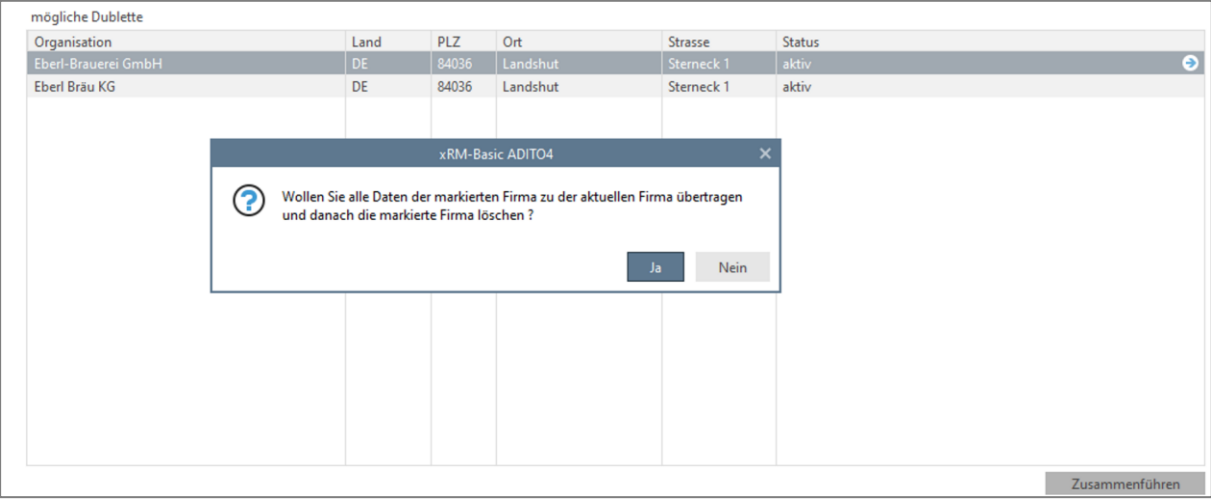
Wird ein Datensatz geöffnet, bei dem vermutlich eine Dublette vorliegt, so zeigen die Masken **Kontaktperson** oder **Firma** einen Dubletten-Reiter an, in welchem die möglichen Dubletten angezeigt werden.



The screenshot shows the ADITO software interface for a company profile. The main area displays the company name 'Eberl-Brauerei GmbH & Co. KG' and various statistics like 'Besuche' (17) and 'Angebote' (16). A 'Dubletten' (Duplicates) tab is active, showing a table of potential duplicate entries.

Organisation	Land	PLZ	Ort	Strasse	Status
Eberl-Brauerei GmbH	DE	84036	Landshut	Sterneck 1	aktiv
Eberl Bräu KG	DE	84036	Landshut	Sterneck 1	aktiv

Um die Dubletten aufzulösen, klicken Sie auf **Bearbeiten**, markieren den Datensatz der aufgelöst und mit der aktuellen Firma zusammengefügt werden soll und klicken **Zusammenführen**. Es erscheint ein Warnhinweis.



The screenshot shows the same duplicate table as above, but with a warning dialog box overlaid. The dialog box contains the following text:

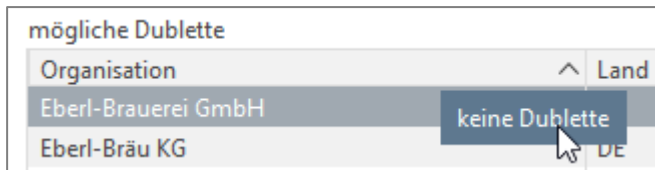
Wollen Sie alle Daten der markierten Firma zu der aktuellen Firma übertragen und danach die markierte Firma löschen?

Buttons: Ja, Nein



Die markierte Firma wird tatsächlich gelöscht und ist danach vom System aus nicht mehr wiederherstellbar.

Alternativ können Sie auch die Dublette in die Liste der nicht als Dubletten vorhandenen Datensätze übernehmen. Klicken Sie dazu mit der rechten Maustaste auf die Firma oder Person, die keine Dublette ist, und wählen `Keine Dublette`.



Der Datensatz wird dann dauerhaft in der Liste der nicht vorhandenen Dubletten gespeichert.

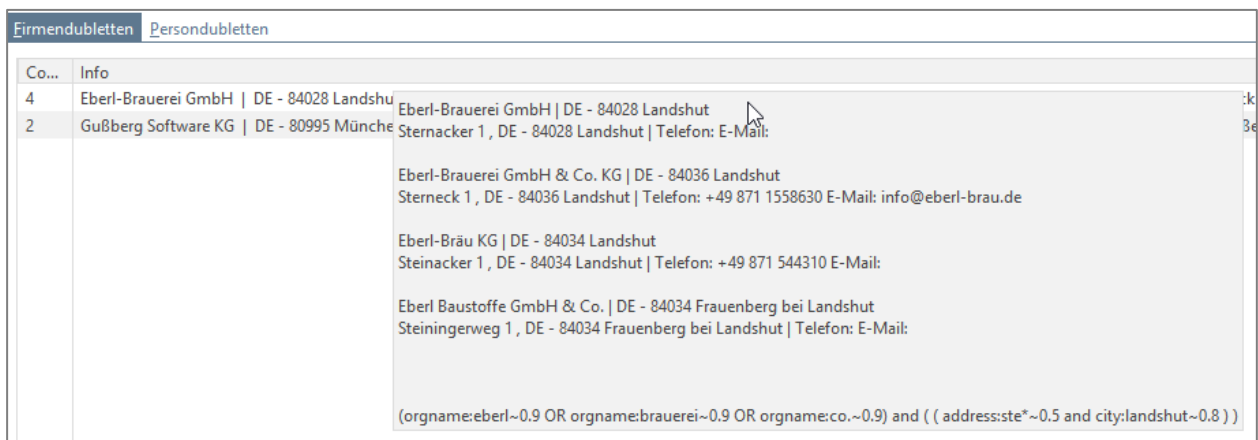
2.1.3. Im Frame Dubletten

Für Fachadministratoren steht ein gesonderter Dubletten-Frame zur Verfügung. In diesem können alle Firmen- und Personen-Dubletten angezeigt werden.

Die Besonderheit hierbei ist, dass der Dublettenvergleich hier **global** stattfindet.

Heißt: Es werden alle Firmen / Personen im System durchsucht und anhand der Gemeinsamkeiten des ersten gefundenen Kopfdatensatzes werden die Dubletten zueinander gruppiert.

Das ergibt eine etwas unschärfere Sichtweise auf den Datenbestand, weshalb hier möglicherweise Dubletten erscheinen, die in der Anzeige des Dubletten-Reiters im Frame `Firma` nicht erscheinen.



Diese Dubletten werden nicht automatisch berechnet. Um eine Überprüfung der Gesamtdubletten anzustoßen, müssen die Schaltflächen `Firmen-dubletten` aktualisieren und `Personen-Dubletten` aktualisieren angeklickt werden.

3. Technische Umsetzung

3.1. Verwendeter Dubletten-Algorithmus: Apache Lucene

Der Index, der mit Hilfe von Apache Lucene aufgebaut wird, dient in der Regel zur schnellen und leistungsfähigen Suche innerhalb von ADITO, um Firmen, Personen und dergleichen mehr zu finden.



Details zur Indexsuche, wie diese aufgebaut ist und im Kontext von ADITO4 funktioniert, finden Sie im Designer-Handbuch.

Für die Dublettenkontrolle wird die unscharfe Suche von Apache Lucene verwendet.

3.1.1. Beispiel im Frame Dubletten

Das heißt, es ist möglich, die Dublettenprüfung wie folgt z.B. für Firmen durchzuführen:

Die Prüfung auf `(orgname: gußberg~1 OR orgname: software~1) AND ((adress: am *~1 AND city: münchen~1))` führt zur Ermittlung von 2 Dubletten:

Firmendubletten		Persondubletten	
Co...	Info		
4	Eberl-Brauerei GmbH DE - 84028 Landshut Sternacker 1, DE - 84028 Landshut Telefon: E-Mail: Eberl-Brauerei GmbH & C		
2	Gußberg Software KG DE - 80995 München Am Obstberg 32, DE - 80995 München Telefon: E-Mail: Huber Software D		

3.1.2. Suchgenauigkeit

Die Suchabfrage mit "Tilde" (~) gibt die Genauigkeit der Suche an. Diese Suche kann in Werten zwischen 0 und 2 erfolgen – wobei 0 nur mit sehr hoher Ähnlichkeit sucht und 2 schon bei sehr niedriger Ähnlichkeit Ergebnisse anzeigt.

3.2. Suchmuster

Für die Suche nach den Dubletten müssen die Schlüsselwörter `DuplicateOrgPattern` für Firmen und `DuplicatePersPattern` für Kontaktpersonen definiert sein.

Die Spaltenzuweisungen dienen dabei den folgenden Zwecken:

Spalte	Definition
Aktiv	Gibt an, ob das untergeordnete Schlüsselwort bei der Dublettenprüfung berücksichtigt wird.
Keyname1	Feld, das bei Dublettenprüfung berücksichtigt werden soll (aus den Datenbanktabellen <code>ORG</code> , <code>PERS</code> , <code>RELATION</code> , <code>ADDRESS</code> und <code>COMM</code>).
Keyname2	Angabe der Suchgenauigkeit in der Form <code>~x</code> . Ist hier kein Wert eingetragen, entspricht das der Suchgenauigkeit 1.

Spalte	Definition
Details	Ausdrücke, die bei der Dublettenprüfung bei dem Feld ausgeschlossen werden.
Beschreibung	Wenn leer, wird das Feld mit AND verknüpft. Wenn die Verknüpfung OR angegeben ist, wird das Feld mit OR verknüpft.

Mit Hilfe dieser Schlüsselwörter wird bei der eigentlichen Dublettenprüfung dann das sogenannte Suchmuster ("Suchpattern") aufgebaut.

3.3. Einsatz

3.3.1. Neuanlage von Datensätzen

Standardmäßig wird die Dublettenkontrolle in den Frames „Person“ (**PERS**) und „Firma“ (**ORG**) durchgeführt.

In **PERS** (Kontaktpersonen) sind

- Vorname (Genauigkeit ~1)
- Nachname (Genauigkeit ~1)
- Telefonnummer
- Faxnummer
- E-Mail-Adresse
- Internetadresse
- Anrede

die geprüften Kriterien, in **ORG** (Firma)

- Firma (Genauigkeit ~1)
- Straße (Genauigkeit ~2)
- Ort (Genauigkeit ~1)
- Telefonnummer
- Faxnummer
- E-Mail-Adresse
- Internetadresse.

3.4. Tabellen und Funktionen

3.4.1. Tabelle DUPLICATE

Diese Tabelle wird zum Speichern der Dubletten benötigt. Sie besteht aus folgenden Spalten:

Spalte	Datentyp	NULL erlaubt?
DATE_NEW	TIMESTAMP	Ja
INFO	CLOB	Ja
NUMBEROF	INTEGER	Nein
OBJECT_ID	INTEGER	Nein
ROW_IDS	VARCHAR(2560)	Nein
USER_NEW	VARCHAR(63)	Ja

3.4.2. Tabelle NODUPLICATE

In dieser Tabelle werden diejenigen Datensätze gespeichert, die keine Dubletten sind.

Spalte	Datentyp	NULL erlaubt?
OBJECT_ID	INTEVER	Ja
DATE_EDIT	TIMESTAMP	Ja
DATE_NEW	TIMESTAMP	Ja
NODUPLICATEID	CHAR(36)	Nein
ROW1_ID	CHAR(36)	Ja
ROW2_ID	CHAR(36)	Ja
USER_EDIT	VARCHAR(30)	Ja
USER_NEW	VARCHAR(30)	Ja

3.4.3. lib_duplicate

Kern der Dublettenkontrolle ist diese Funktionsbibliothek. Hier sind alle Funktionen vereint, die prüfen, ob eine Person oder Firma schon Dubletten hat, oder ob solche Firmen als „keine Dubletten“ markiert sind.

3.4.3.1. escapepattern

Formatiert das Suchpattern in Kleinbuchstaben.

- **pString** (String): übergibt den Suchpattern

Rückgabewert ist das Suchpattern in Kleinbuchstaben formatiert.

3.4.3.2. getDuplicates

Ermittelt Details für die Frames auf Basis der ermittelten Dubletten.

- **pFrame** (String): übergibt den Namen des Frames
- **pDupIDs** (Array): enthält die Liste der möglichen Dubletten als Array

Rückgabewert sind ausführlichere Informationen zu den Dubletten, die in Frame-spezifischen Funktionen definiert werden können.

3.4.3.3. getDuplicatesOrg

Ermittelt die Details für die Dubletten einer Firma.

- **pDupIDs** (Array): enthält die Liste der möglichen Dubletten als Array

Rückgabewert sind ausführlichere Informationen zu den Dubletten, die in Frame-spezifischen Funktionen definiert werden können.

3.4.3.4. getDuplicatesPers

Ermittelt die Details für die Dubletten einer Kontaktperson.

- **pDupIDs** (Array): enthält die Liste der möglichen Dubletten als Array

Rückgabewert sind ausführlichere Informationen zu den Dubletten, die in Frame-spezifischen Funktionen definiert werden können.

3.4.3.5. getOrgFramePattern

Ermittelt das Pattern für die Indexsuche für den Frame `ORG`.

Rückgabewert ist das Suchpattern.

3.4.3.6. getOrgQuickPattern

Ermittelt das Dublettenkriterium für den Frame `QUICKINSERT`.

- **pCommData** (Array): enthält eine Liste der Kommunikationsdaten

Rückgabewert ist das Suchpattern.

3.4.3.7. getPattern

Erzeugt das Pattern für die Indexsuche einer `ORG`.

- **pData** (Array): Daten der zu prüfenden Firma
- **pConfig** (Array): Konfiguration aus den duplicate-Schlüsselwörtern

Rückgabewert ist das Suchpattern.

3.4.3.8. getPersFramePattern

Ermittelt das Pattern für die Indexsuche für den Frame `PERS`.

Rückgabewert ist das Suchpattern.

3.4.3.9. getPersQuickPattern

Ermittelt das Dublettenkriterium für den Frame `QUICKINSERTPERS`.

- **pCommData** (Array): enthält eine Liste der Kommunikationsdaten

Rückgabewert ist das Suchpattern.

3.4.3.10. hasOrgDuplicates

Diese Methode stößt die Dublettenkontrolle für Firmen an.

- **pRelationID** (String): `RelationID`
- **pPattern** (String): enthält das Suchpattern für die Ermittlung der Dubletten
- **pObjectID** (Ganzzahl als String): `ObjectID / FrameID`

Rückgabewert ist ein Array mit allen möglichen Dubletten.

3.4.3.11. hasPersDuplicates

Diese Methode stößt die Dublettenkontrolle für Kontaktpersonen an

- **pRelationID** (String): `RelationID`
- **pPersID** (String): `PersID`
- **pPattern** (String): enthält das Suchpattern für die Ermittlung der Dubletten
- **pObjectID** (Ganzzahl als String): `ObjectID / FrameID`

Rückgabewert ist ein Array mit allen möglichen Dubletten.

3.4.3.12. noduplicate

Fügt in die Tabelle `NODUPLICATES` Dubletten ein, die somit als ‚keine Dublette‘ gekennzeichnet werden.

- **pRowID1** (String): ID des ersten Datensatzes (`RelationID` bei Personen und Firmen, `PropertyID` bei Objekten)
- **pRowID2** (String): ID des zweiten Datensatzes (`RelationID` bei Personen und Firmen, `PropertyID` bei Objekten)
- **pObjectID** (String): `FrameID`, die den Typ des Datensatzes beschreibt (z.B. 2 für `PERS`)

`noduplicate` enthält keinen Rückgabewert.

3.4.3.13. mergeOrg

Diese Funktion führt zwei Organisationen zusammen. Unter Umständen muss die zusammengeführte Firma noch bearbeitet werden, da doppelte Werbeeinschränkungen oder Segmente vorhanden sein können.

- **pOldRelID** (String): ID der alten Relation

- **pNewRelID** (String): ID der neuen Relation

`mergePers` enthält keinen Rückgabewert.

3.4.3.14. `mergePers`

Diese Funktion führt zwei Kontaktpersonen zusammen. Unter Umständen muss die zusammengeführte Firma noch bearbeitet werden, da doppelte Werbeeinschränkungen vorhanden sein können.

- **pOldRelID** (String): ID der alten Relation
- **pNewRelID** (String): ID der neuen Relation

`mergeOrg` enthält keinen Rückgabewert.

3.4.3.15. `setStatements`

Diese Funktion ist eine Hilfsfunktion, die von `mergeOrg` heraus aufgerufen wird und baut ein SQL-Statement. Sie liefert das erstellte Statement als Rückgabewert.

3.4.3.16. `setStatements4Participants`

Erzeugt die Statements zur Zusammenführung von Tabellen für Kampagnenteilnehmer.

- **pTable** (String): Tabellenname
- **pColumn** (String): Spaltenname
- **poldValue** (String): Alter Wert
- **pnewValue** (String): Neuer Wert
- **pCondition** (String): Bedingung

3.4.4. `sp_getOrgDuplicate`

Dieser Serverprozess aktualisiert die Firmen-Dubletten.

3.4.5. `sp_getPersDuplicate`

Dieser Serverprozess aktualisiert die Personen-Dubletten.

3.4.6. DUPLICATIONS

Dieser Frame zeigt Firmen- und Personendubletten an.

3.5. Verwendung der Dublettenlogik in angepassten Systemen

Wollen Sie die Dublettenlogik des ADITO-Referenzsystems in Ihr System übernehmen, können Sie die in [Kapitel 3.4](#) vorhandenen Tabellen und Funktionsbibliotheken übernehmen. Beachten Sie aber, dass gegebenenfalls zu beachtende weitere Tabellen oder Besonderheiten in Ihrer Datenstruktur mit angepasst werden müssen – werden diese Datenstrukturen nicht berücksichtigt, kann es vorkommen, dass Dubletten falsch angezeigt werden.

4. Alte Dublettenkontrolle in AO Referenz bis 2012 (SOUNDEX)

Die folgenden Absätze beziehen sich auf die Dublettenkontrolle, wie sie im Referenzsystem bis einschließlich AOReferenz 2012 eingesetzt wurde. Gegebenenfalls kann die Logik in Ihrem System abweichen, deaktiviert oder ausgeblendet sein. Fragen Sie im Zweifel Ihren Systemadministrator oder kontaktieren Sie den ADITO-Support.

4.1. Verwendeter Dubletten-Algorithmus Soundex

Hierbei handelt es sich um einen Algorithmus, der Wörter und Phrasen nach ihrem Klang indizieren kann. Auf Basis der Klangfarbe eines Wortes wird ein Code entwickelt. Vergleicht man nun diese Codes, kann ermittelt werden, ob Wörter ähnlich klingen oder nicht.

Der Soundex-Code besteht aus seinem ersten Buchstaben, gefolgt von drei (oder mehr) Ziffern, ermittelt aus den Konsonanten des Wortes. Aufgebaut ist dieser Code also immer nach seinem Anfangsbuchstaben und dann den Codes für die Konsonanten, die aus einer Tabelle heraus ermittelt werden können (ADITO ist hier 3300). Hat ein Wort nicht mehr als drei Konsonanten, wird der Code mit „0“ aufgefüllt.

Über diese Methoden kann man nun Wörter miteinander vergleichen, die nicht dieselbe Schreibweise besitzen:

Designer: 3225

Desaigner: 3225

Deisinger: 3252

Der Soundex wird im ADITO-Referenzsystem bis einschließlich dem Referenzsystem 2012 verwendet.

4.2. Einsatz

4.2.1. Neuanlage von Datensätzen

Standardmäßig wird die Dublettenkontrolle in den Frames `Person (PERS)` und `Firma (ORG)` durchgeführt.

In **PERS** sind

- Vorname und
- Nachname

die geprüften Kriterien, in **ORG**

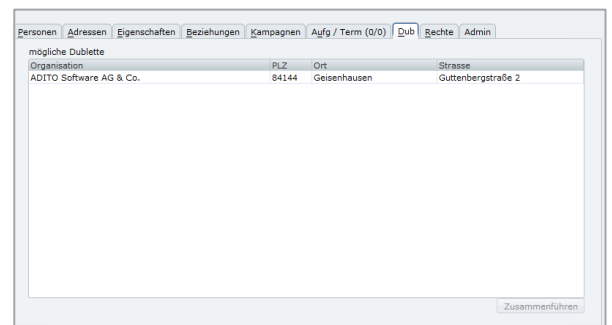
- Firma
- Straße
- Ort.

Wird eine mögliche Dublette ermittelt, so erscheint bei Klick auf **Speichern** eine Dublettenwarnung mit einer Liste der möglichen Dubletten. Hier kann ein Datensatz ausgewählt und die Entscheidung getroffen werden, was mit dem eben angelegten Datensatz geschehen soll:

- **Trotzdem übernehmen** legt den Datensatz ohne weitere Nachfrage in der Datenbank an, mit dem Risiko, dass eine Dublette angelegt wurde.
- **Markierten verwenden** bricht die Eingabe der bestehenden Daten ab. Sollte der alte Datensatz bei einer anderen Firma angelegt worden sein als die mögliche Dublette, so wird die Person auch bei der neuen Organisation eingetragen.
- **Abbrechen** wechselt zurück in den Edit-Modus, so dass die Eingaben noch einmal geprüft werden können.

4.2.2. Dublettenkontrolle bei bestehenden Datensätzen

Existiert zu einem bestehenden Datensatz eine mögliche Dublette, so wird in den entsprechenden Bildschirmmasken ein Reiter **Dub** angezeigt. In diesem Reiter sehen Sie eine Übersicht aller möglichen Dubletten zu diesem Datensatz (im nebenstehenden Beispiel: Firma).



Organisation	PLZ	Ort	Strasse
ADITO Software AG & Co.	84144	Geisenhausen	Guttenbergstraße 2

Sie können diese Datensätze markieren und per Doppelklick öffnen.

Wechseln Sie in den Edit-Modus, so steht Ihnen nach dem Markieren der Firma die Schaltfläche „Zusammenführen“ unterhalb der Tabelle zur Verfügung. Klicken Sie auf diese Schaltfläche, werden nach Sicherheitsabfrage die beiden Datensätze zusammengeführt.

Nach erfolgreichem Zusammenführen wird beim übrig gebliebenen Datensatz ein Historieneintrag angelegt.

4.3. Konfiguration der Soundex-Genauigkeit

Wie scharf oder unscharf der Soundex konfiguriert wurde, können Sie im Frame Systemverwaltung festlegen. Dort gibt es den Reiter **SoundexConfig**.

Firma Soundex		
Spalte	Länge	1.Buchstabe
Firma	8	<input type="checkbox"/>
Ort	4	<input type="checkbox"/>
Strasse	4	<input type="checkbox"/>

Person		
Spalte	Länge	1.Buchstabe
Nachname	4	<input type="checkbox"/>
Vorname	4	<input type="checkbox"/>

Firma eindeutig	
Spalte	
Strasse	
Firma	
Ort	
PLZ	

Person eindeutig	
Spalte	
Nachname	
Vorname	

Soundex-Konfiguration eintragen

Hier können Sie für Firma und Person angeben, welche Spalten für die Soundex-Generierung verwendet werden sollen, wie lange der Soundex für die jeweiligen Spalten ist und welche Spalten eindeutig sein müssen.

Die Angabe „1.Buchstabe“ gibt an, ob sich der erste Buchstabe des jeweiligen Begriffs unterscheiden kann oder nicht (so sind bei deaktivierter Funktion „Tesigner und Designer“ Dubletten, bei aktivierter Funktion nicht).

4.4. Technologie

4.4.1. lib_soundex

Kern der Dublettenkontrolle ist die Funktionsbibliothek `lib_soundex`, die eine JavaScript-Implementierung des Soundex-Algorithmus darstellt. Diese Funktionsbibliothek enthält nur eine Funktion:

4.4.1.1. SoundEx

- **WordString** ist das zu codierende Wort als Text.
- **LengthOption** die Länge des Soundex-Codes.
- **CensusOption** (`true / false`) beschränkt die Länge des Soundex-Codes auf vier Zeichen.

Der Rückgabewert ist der Soundex-Code des übergebenen Wortes.

4.4.2. lib_duplicate

Verarbeitet wird die Dublettenkontrolle aber in der Funktion `lib_duplicate`. In dieser Funktionsbibliothek stehen JDito-Funktionen zur Verfügung, welche die oben genannte Funktion der Dublettenkontrolle in Firma und Kontaktperson ermöglichen.

4.4.2.1. hasDuplicates

Diese Methode stößt die Dublettenkontrolle an.

- **pFrame** (Text): übergibt den Namen des Frames
- **pActID** (Text, optional): enthält eine Relation-ID, die ausgeschlossen werden kann

Rückgabewert ist das Ergebnis von `getDuplicateIDs`.

4.4.2.2. getDuplicateIDs

Hier werden die IDs der Datensätze ermittelt, zu denen eine mögliche Dublette angelegt wurde.

- **pFrame** (Text): übergibt den Namen des Frames
- **pValue** (Text): enthält den Soundex-Code, der verglichen werden muss
- **pActID** (Text, optional): enthält eine Relation-ID, die ausgeschlossen werden kann

Rückgabewert ist eine Liste der IDs, die als Dubletten in Frage kommen.

4.4.2.3. getDubCriterion

Diese Funktion dient dazu, die Kriterien zur Dublettenkontrolle festzulegen. Durch Hinzunehmen weiterer Felder können Sie die Kontrolle genauer einstellen.

- **pFrame** (Text): übergibt den Namen des Frames

Möchten Sie die Kontrolle um eine weitere Komponente erweitern, so nehmen Sie sie einfach in die Liste der zu prüfenden Komponenten auf - im Fall von „Stadt“ bei der Kontaktperson:

```
case "PERS":  
pValues = new Array(vars.getString("$comp.lastname")  
,vars.getString("$comp.firstname"), vars.getString("$comp.city"));  
break;
```

Rückgabewert ist der Soundex-Code aus den übergebenen Werten.

4.4.2.4. getDupFields

Analog zur Funktion `getDubCriterion` werden hier die Datenbankspalten angegeben, die aus den Komponenten der vorgenannten Funktion ausgelesen werden.

- **pFrame** (Text): übergibt den Namen des Frames
- **unique** (`true / false`): bei `true` wird für die Kontaktpersonen zusätzlich die Adresse als Kriterium hinzugefügt, bei `false` nicht

Rückgabewert ist die Liste der Datenbankspalten.

4.4.2.5. getSoundex

Diese Funktion lässt den Soundex-Code von der Funktionsbibliothek `SoundEx` ermitteln.

- **pValues** (Text): enthält die Werte, von denen der Code ermittelt werden soll

Rückgabewert ist der Soundex-Code.

4.4.2.6. getDuplicates

Ermittelt Details für die Frames auf Basis der ermittelten Dubletten.

- **pFrame** (Text): übergibt den Namen des Frames
- **pDupIDs** (Array): enthält die Liste der möglichen Dubletten als Array

Rückgabewert sind ausführlichere Informationen zu den Dubletten, die in Frame-spezifischen Funktionen definiert werden können.

4.4.2.7. getDuplicates[X]

Für jeden Frame, der in der Dublettenkontrolle angegeben ist, existiert eine Hilfsfunktion, in der die Detailinformationen der Dublettenliste ermittelt werden.

- **pDupIDs** (Array): enthält die Liste der möglichen Dubletten als Array

Rückgabewert sind ausführliche Informationen, die in der Funktion definiert wurden.

4.4.2.8. setTableSoundex

Schreibt den Soundex-Code eines Datensatzes in eine Tabelle.

- **pTable** (String): Datenbanktabelle, in die der Soundex-Code geschrieben werden soll
- **pTableID** (String): ID des Datensatzes, zu dem der Soundex angelegt werden soll
- **pSelectStr** (String): SQL für die Ermittlung der zu aktualisierenden Daten
- **pSetField** (String): Datenbankspalte, die gesetzt werden soll

Rückgabewert ist die Anzahl der aktualisierten Einträge.

4.4.2.9. set[X]Soundex

In diesen Hilfsfunktionen wird `setTableSoundex` aufgerufen, um Frame-spezifisch den Soundex-Code setzen zu können.

4.4.2.10. mergeOrg

Diese Funktion führt zwei Organisationen zusammen.

- **poldRelID** (String): ID der alten Relation
- **pnewRelID** (String): ID der neuen Relation

`mergeOrg` enthält keinen Rückgabewert.

4.4.2.11. setStatements

Diese Funktion ist eine Hilfsfunktion, die von `mergeOrg` heraus aufgerufen wird und baut ein SQL-Statement. Sie liefert das erstellte Statement als Rückgabewert.