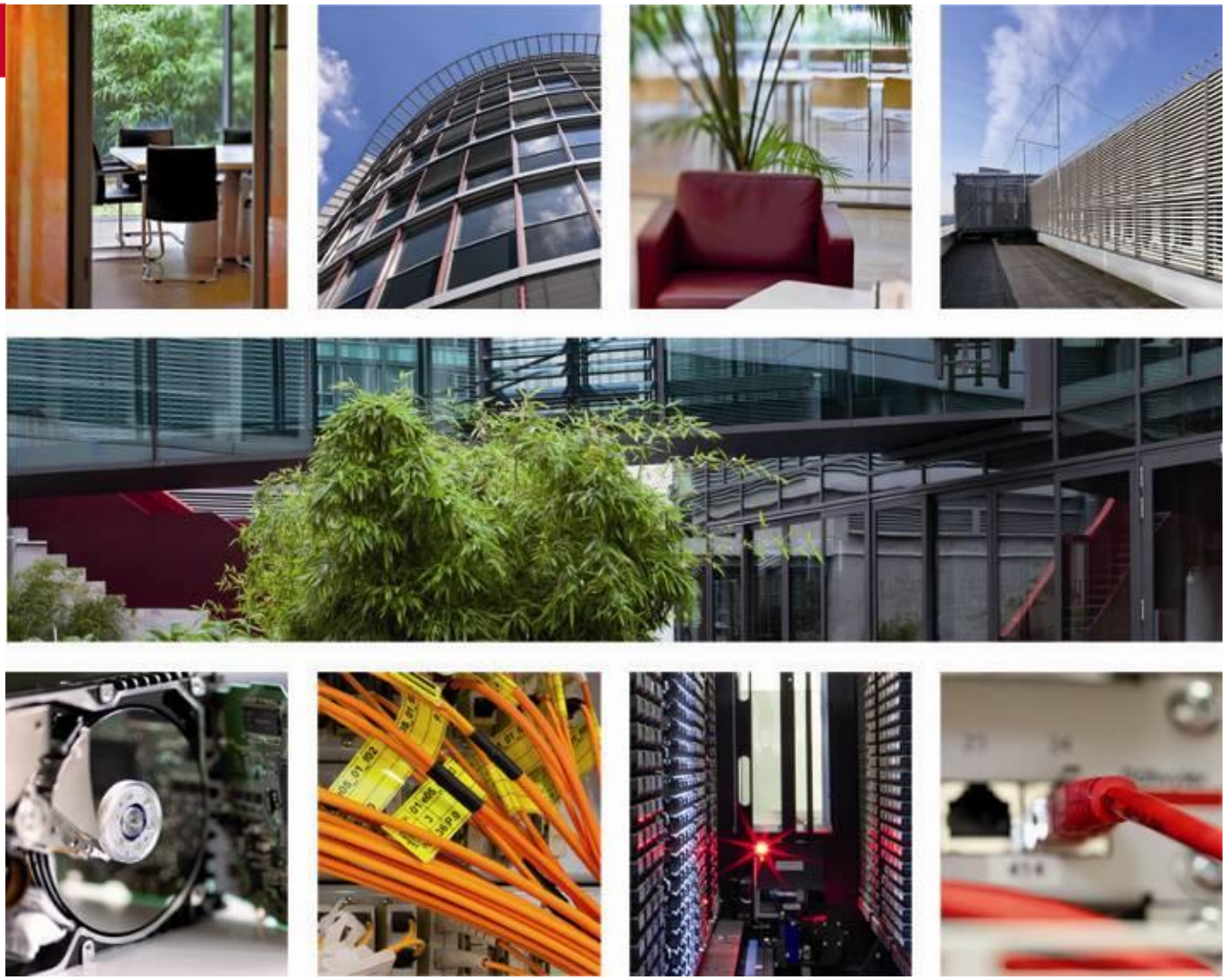
HZD
Hessische Zentrale für Datenverarbeitung

HESSEN

Ergänzung zur Schnittstellenbeschreibung OASIS WS:

Erweiterung der Suchfunktionalität bei Statusabfragen in OASIS durch Transliteration (Umschlüsselung) und Permutation

Ab OASIS Release 7.4



Inhaltsverzeichnis

1 Transliteration 3

1.1 Abbildung der Daten auf einen Grundzeichensatz: Transliteration 3
1.1.1 Ligaturen entfernen 4
1.1.2 Diakritika entfernen 4
1.1.3 Wandlung in Großbuchstaben 4
1.1.4 Ersetzen von Zeichen 4
1.1.5 Entfernen und Ersetzen von Zeichen und Zeichenfolgen 4
1.1.6 Zu entfernende Wörter und Ziffern im Ort und Geburtsort 5
1.1.7 Ersetzungen im Vornamen, Nachnamen, Geburtsnamen 6
1.1.8 Behandlung des Geburtsdatums 7
1.1.9 Ersetzungen in den Feldern Straße, Hausnummer, Postleitzahl 7
1.1.10 Ersetzungen im Feld Adresszusatz 9

2 Permutationen und Suche nach Worten 9

2.1 Zusammenfassung und Permutation von Vor- und Nachnamen 9

2.2 Erweiterte Suche 10

2.3 Beispiele mit mehreren Vornamen / mehreren Nachnamen 10

3 Umsetzung 11

3.1 Abbildung der Daten auf einen Grundzeichensatz und die Transliterationen 12

3.2 Stufe 1: Allgemeine Normierung (Textfelder) 14

3.3 Stufe 2: Feldspezifische Normierung 15

3.4 Stufe 3: Entfernen von Füllworten und Zeichenketten 15

1 Transliteration

Die Treffsicherheit bei der Statusabfrage wird durch Umschlüsselung der übergebenen Daten einer Spielerstatusabfrage auf einen gemeinsamen Grundzeichensatz verbessert. Neben der reinen Transliteration werden Regeln angewandt, die sich in der bisherigen Praxis als sinnvoll herausgestellt haben.

Beim Vorgehen soll sich an deutsche und internationale Standards gehalten werden. Im Bedarfsfall werden diese um zusätzliche Regeln erweitert. Grundsätzlich sollten bei einer manuellen Eintragung von Daten stets die identischen Zeichen des Ausweisdokumentes übernommen werden. Bei abweichender Eintragung von Daten, kann ein Treffer bei der Spielerstatusabfrage nicht mehr gewährleistet werden.

1.1 Abbildung der Daten auf einen Grundzeichensatz: Transliteration

Die Abbildung der Daten auf einen gemeinsamen Grundzeichensatz erfolgt in mehreren Schritten, die in den folgenden Kapiteln erläutert werden. Die Umschlüsselung erfolgt basierend auf dem durch DIN SPEC 91379 herausgegebenen Standard „Normative Zeichen“ (Zeichen der Gruppen, die als normativ bezeichnet werden). Dieser Zeichensatz beinhaltet die lateinischen Zeichen aus Unicode einschließlich diakritischer Zeichen und Ligaturen.

Zur Veranschaulichung der einzelnen Schritte wird als Beispiel die fiktive Person Dr. Nœl Isaać Schäfer-Strauß, Friedrich-Ebert-Allee 12e, 65185 Wiesbaden, geboren am 01.12.1970, geboren in Wehrda jetzt Marburg, herangezogen:

Abbildung 1: Fiktive ID-Card von Herrn Dr. Schäfer-Strauß

1.1.1 Ligaturen entfernen

Ligaturen – zusammengesetzte Zeichen in bestimmten Sprachen und Kulturkreisen – werden ersetzt.

Beispiel: Nœl → Noel

1.1.2 Diakritika entfernen

Diakritika sind spezielle Zeichen, die im Zeichenvorrat einer Sprache oder eines Kulturkreises vorkommen und den lateinischen Grundzeichensatz erweitern. Beispiele hierfür sind z.B. ü, ö, ä. In unserem Beispiel wird aus den Vornamen Isaa**ć** → Isaa**c**, der Nachname Sch**ä**fer zu Sch**ae**fer und Strau**ß** zu Strau**ss**.

1.1.3 Wandlung in Großbuchstaben

Dieser Schritt wird in allen Standards durchgeführt und hat den Vorteil, dass später in den Anwendungen häufige Zeichenoperationen, die z.B. Klein- in Großbuchstaben umwandeln, entfallen können.

1.1.4 Ersetzen von Zeichen

Die aus dem durch DIN SPEC 91379 Standard herausgegebenen Sonderzeichen

!"#%&()*,-./:;?@[\]_{}¡«·»¿$+<=>^|~¢£¤¥¦§©¬®°±¶×÷€¸¨¯

werden durch Leerzeichen ersetzt (MUELLER-MEIER → MUELLER MEIER).

Eine Sonderstellung hat das Zeichen „-“ (Minus bzw. Bindestrich):

- Wenn „-“ zusammen mit anderen Zeichen in einem Feld vorkommt:

  Wenn das „-“ zusammen mit anderen Zeichen vorkommt, wird es wie oben beschrieben durch ein Leerzeichen ersetzt (z.B. im Tag der Geburt oder Monat der Geburt können zwei „—“vorkommen: in diesem Fall werden beide Minuszeichen durch *jeweils* ein Leerzeichen ersetzt).

- Wenn „-“ als einziges Zeichen in einem Feld vorkommt:

  Wenn nur das „-“ als Feldinhalt vorhanden ist (also keine anderen Zeichen in diesem Feld vorhanden sind), bedeutet „-“, dass der Wert dieses Feldes nicht vorhanden ist. Zum Beispiel bedeutet ein „-“ im Feld „Wohnort“, dass die Angabe nicht existiert (evtl. obdachlos) oder ein „-“ in Feld „Vorname“ bedeutet, dass es in diesem Kulturkreis keinen Wert für „Vornamen“ gibt.

  Der Feldinhalt „-“ führt dazu, dass dieses Feld von der Suche ausgenommen wird, also behandelt wird wie ein Platzhalter. Ausnahmen bilden hier die Felder „Jahr der Geburt“ und „Nachname“: Diese beiden Pflichtfelder dürfen nicht nur ein „-“ als Inhalt haben!

  (Für Tag der Geburt und Monat der Geburt sind jeweils zwei „—“ möglich.)

1.1.5 Entfernen und Ersetzen von Zeichen und Zeichenfolgen

Folgende Zeichen oder Zeichenfolgen werden entfernt:

1. Alle Vorkommen von „DR“ und „PROF“ in den Feldern Vor- und Nachname sowie Geburtsname werden entfernt. Diese Regel wird nur angewandt, wenn die Zeichen alleine stehen.

2. alle Apostrophe in Namen „`“, „‘“, „´“(dies betrifft auch die Fehleingaben von Hochkommata wie z.B. „D`Artagnon“ und D´Artagnon“).

   Beispiel: D’ARTAGNON → DARTAGNON.

   Anwendung: in allen Feldern außer Geburtsdatum.

3. Während der Umwandlung evtl. auftretende mehrfache Leerzeichen werden durch ein Leerzeichen ersetzt.

   Anwendung auf alle Felder.

4. Führende oder nachfolgende Leerzeichen bei einem Attribut werden in allen Feldern entfernt.

1.1.6 Zu entfernende Wörter und Ziffern im Ort und Geburtsort

Der Geburtsort wurde zuvor schon von den bisherigen genannten Regeln von Sonderzeichen bereinigt.

1. Aus Orten und Geburtsorten entfernt werden:
   - JETZT
   - BAD
   - VOR
   - DER
   - BEI
   - AUF
   - IM
   - IN
   - AN
   - AM
   - OB
   - Alle Zahlen
2. Auch Ortsnamen, die aus weniger als zwei Zeichen bestehen, sind erlaubt

Beispiele aus OASIS HSpielhG:

die Suche nach einem „Bad Homburg v.d. Höhe“ → würde umgewandelt in die Suche nach einem: HOMBURG V D HOEHE

die Suche nach einem BAD HOMBURG V D HÖHE → würde umgewandelt in die Suche nach einem: HOMBURG V D HOEHE

die Suche nach einem Bad Homburg v d Höhe → würde umgewandelt in die Suche nach einem: HOMBURG V D HOEHE

die Suche nach einem Bad Homburg v. d. Höhe → würde umgewandelt in die Suche nach einem: HOMBURG V D HOEHE

die Suche nach einem Bad Homburg v.d. Höhe → würde umgewandelt in die Suche nach einem: HOMBURG V D HOEHE

die Suche nach einem Bad Homburg V.D. Höhe → würde umgewandelt in die Suche nach einem: HOMBURG V D HOEHE

die Suche nach einem BAD HOMBURG2 VD HÖHE → würde umgewandelt in die Suche nach einem: HOMBURG VD HOEHE

die Suche nach einem Bad Homburg vor der Höhe → würde umgewandelt in die Suche nach einem: HOMBURG HOEHE

3. Wenn ein Feld aus mehreren Wörtern besteht, spielt die Eingabereihenfolge bei der Suche keine Rolle.

   (Achtung: die oben aufgeführten Präpositionen werden im Ort/Geburtsort ganz entfernt)

4. Alle Wörter der Eingabe müssen exakt (nach der Transliteration, Normalisierung,..) so in der Suchtabelle vorkommen. Der Eintrag in der Suchtabelle kann mehr Wörter enthalten als die Suchanfrage, aber nicht weniger!

Beispiele:

Anlegen einer Sperre mit Ort	Nutzer sucht mit	Treffer [ja|nein]
Bad Homburg	Homburg	Ja
Bad Homburg v.d. Höhe	Bad Homburg	Ja
Bad Homburg v.d. Höhe	Homburg Bad	Ja
Bad Homburg v.d. Höhe	Höhe	Ja
Homburg	Bad Homburg	Ja
Frankfurt	Frankfurt Main	Nein
Bad Homburg	Bad	Ja
Bad Homburg	Bad Homburg v.d. Höhe	Nein
Bad Homburg v.d. Höhe	Bad Testburg v.d. Höhe	Nein
Homburg	Burg	Nein

1.1.7 Ersetzungen im Vornamen, Nachnamen, Geburtsnamen

Vorname, Nachname: Es wird der vollständige Name gemäß Antragsdatensatz, das heißt einschließlich Namensbestandteile wie „Freiherr von und zu“, aber ohne Geburtsname, übernommen.

Für **alle Namensfelder** (Vorname, Nachname und Geburtsname) gilt: Alle Ziffern werden ausgeschnitten.

Vorname:

Im Feld „Vorname“ werden ab dem ersten Auftreten des Wortes „geb“ (Groß- / Kleinschreibung wird ignoriert) alle Zeichen (inkl. „geb“) abgeschnitten.

Ein alleine stehendes „geb“ (sonst keine weiteren Zeichen in diesem Feld) wird nicht ausgeschnitten.

Zum Beispiel wird bei der transliterierten Suche aus „Müller geb. Fischer“ nun „Müller“

Nachname:

Im Feld „Nachname“ werden ab dem ersten Auftreten des Wortes „geb“ (Groß- / Kleinschreibung wird ignoriert) alle Zeichen (inkl. „geb“) abgeschnitten.

Ein alleine stehendes „geb“ (sonst keine weiteren Zeichen in diesem Feld) wird nicht ausgeschnitten.

Zum Beispiel wird bei der transliterierten Suche aus „Müller geb. Fischer“ nun „Müller“.

Die alleinige Eingabe eines „-“ (sonst keine weiteren Zeichen in diesem Feld) ist nicht erlaubt und wird abgewiesen!

Geburtsname:

Anders als in „Vor-“ oder „Nachname“ werden im Feld „Geburtsname“ bis zum ersten Auftreten des Wortes „geb“ (Groß- / Kleinschreibung wird ignoriert) alle Zeichen (inkl. “geb“) abgeschnitten.

Zum Beispiel wird aus „Müller geb. Fischer“ nun „Fischer“.

Ein alleine stehendes „geb“ (sonst keine weiteren Zeichen in diesem Feld) wird nicht ausgeschnitten.

1.1.8 Behandlung des Geburtsdatums

Bei der Erfassung und Suche ist das Geburtsdatum bereits nach dem Muster JJJJ-MM-TT formatiert, wobei MM und TT auch zwei Minus-Zeichen enthalten können.

Die drei Datumsbestandteile Jahr, Monat Tag („Jahr der Geburt“, „Monat der Geburt“, „Tag der Geburt“) werden getrennt. Anschließend werden eventuell vorhandene Minus-Zeichen entfernt. Die Suche erfolgt nach Jahr, Monat bzw. Tag der Geburt getrennt.

Ein „-“ für das Feld „Jahr der Geburt“ ist nicht gültig und wird abgelehnt.

Beispiele (hier zur Illustration nur für verschiedene Kombinationen von Monat und Tag und nicht vollständig):

Abfrage Geburtsdatum	Sperreintrag	Jahr transliteriert	Monat translite-riert	Tag translite-riert	Treffer
1974-04-30	1974-04-30	1974	04	30	ja
1974-09-30	1974-04-30	1974	09	30	nein
1974-07-30	1974----30	1974	07	30	ja
1974-07-25	1974----30	1974	07	25	nein
1974-04-13	1974-04---	1974	04	13	ja
1974-10-13	1974-04---	1974	10	13	nein
1974-11-29	1974------	1974	11	29	ja
1974-11-29	1980------	1974	11	29	nein
1974----30	1974-01-30	1974		30	ja
1974----13	1974-01-30	1974		13	nein
1974-04---	1974-04-15	1974	04		ja
1974-04---	1974-06-30	1974	04		nein
1974------	1974-04-30	1974			ja
1974------	1986-04-30	1974			nein

1.1.9 Ersetzungen in den Feldern Straße, Hausnummer, Postleitzahl

Straße:

Die Daten aus dem Feld „Straße“ werden entsprechend den zuvor aufgestellten Regeln transliteriert. Zusätzlich werden alle Zahlen und Leerzeichen entfernt.

Ein abschließendes „STR“ oder „STRASSE“ wird entfernt. Die verbliebenen Zeichen werden vollständig verglichen.

Beispiele:

Straße	Transliterierte Straße
Am Holzacker	AMHOLZACKER
5 Rue du Mail	RUEDUMAIL
Hauptstr.	HAUPT
Straße der Republik	STRASSEDERREPUBLIK

Hausnummer:

Mehrere Leerzeichen werden durch ein Leerzeichen ersetzt. Nicht numerische Zeichen in der Hausnummer werden durch Leerzeichen ersetzt. Nach der Ersetzung befindliche Leerzeichen am Anfang und Ende werden entfernt.

Aus 12a wird 12, aus 28-30 wird 28 30.

Führende Nullen am Anfang der bereits transliterierten Hausnummer werden zusätzlich entfernt. Somit wird die Hausnummer 0 zu einer leeren Eingabe!

Die Prüfung der Hausnummer ergibt auch einen Treffer, wenn eins der Worte gefunden wird, wobei Worte hier die durch genau ein Leerzeichen getrennten numerischen Zeichenketten sind.

Beispiel: „28-30“ wird nach der Transliteration zu „28 30“ – bestehend aus den zwei Worten „28“ und „30“. Die Hausnummer wird als übereinstimmend gewertet, wenn bei der Suche „28 bis 30“ oder nur „28“ oder nur „30“ eingegeben wird. Bei einer Suche nach Hausnummer „29“ wird kein Treffer gemeldet. Die Suche erfolgt in beliebiger Reihenfolge.

Beispiele

Abfrage Hausnummer	Sperr-Eintrag	Transliterierte Hausnummer	Treffer
28-30	28-30	28 30	ja
28 bis 30	28-30	28 30	ja
28 - 30	28-30	28 30	ja
30	28-30	28 30	ja
29	28-30	28 30	nein
30 - 28	28-30	28 30	ja

Postleitzahl:

Die Daten aus dem Feld „PLZ“ werden entsprechend den zuvor aufgestellten Regeln transliteriert.

Danach werden alle Buchstaben entfernt.

Beispiel: 12345a wird zu 12345, abcd wird zu „" und 12asdaf45 wird zu 1245.

1.1.10 Ersetzungen im Feld Adresszusatz

Der Adresszusatz wurde zuvor schon von bisherigen genannten den Regeln von Sonderzeichen bereinigt. Innerhalb des bereits transliterierten Adresszusatzes wird nun mit einem Teilstring Parameter gesucht. Dieser meldet einen Treffer in diesem Feld, sobald der gesuchte Teil mit einem Teil des in der Datenbank befindlichen Teils übereinstimmt. Dies können auch einzelne Buchstaben sein.

Im Adresszusatz erfolgt keine Suche in beliebiger Reihenfolge.

Beispiele

Abfrage Adresszusatz	Sperr-Eintrag	Transliterierter Adresszusatzsuche	Treffer
c/o	c/o Kroneburger	C O	ja
Kroneburg	c/o Kroneburger	KRONEBURG	ja
R	c/o Kroneburger	R	ja
c/o Herr Kroneburger	c/o Kroneburger	C O HERR KRONEBURGER	Nein

2 Permutationen und Suche nach Worten

Alle im Folgenden genannten Operationen werden *nach* der Transliteration durchgeführt.

2.1 Zusammenfassung und Permutation von Vor- und Nachnamen

Es handelt sich um die Vertauschung von Vor- und Nachnamen. Hierbei wird eines der Hauptprobleme insbesondere bei ausländischen Ausweisdokumenten behandelt, bei denen nicht eindeutig zugeordnet werden kann, was Vor- und Nachname ist. Erschwerend kommt hinzu, dass solche Spielerinnen und Spieler oftmals unterschiedliche Ausweisdokumente mit unterschiedlichen Zuordnungen zu Vor- und Nachname vorlegen.

Beispiel: Wu Hu Fu.

Auf einem Dokument ist Fu dem Nachnamen zugeordnet und Wu Hu dem Vornamen. Auf einem zweiten Dokument genau umgekehrt und auf einem dritten Wu dem Vornamen und Hu Fu dem Nachnamen.

Ziel der Zusammenfassung und Permutation ist, auch solche Personen zu finden, sofern sie gesperrt sind, unabhängig davon, welches Ausweisdokument sie vorlegen.

Hierzu werden beim Eintragen bzw. Ändern von Sperren sowie bei Statusabfragen nach der Transliteration und Normalisierung die einzelnen Worte aus Vor- und Nachname zusammengefasst und alphabetisch sortiert.

Im Ergebnis handelt es sich also nicht nur um eine reine Vertauschung, sondern um eine Mischung von Vor- und Nachname.

Dadurch lassen sich solche Personen eindeutig identifizieren.

2.2 Erweiterte Suche

Um die Trefferwahrscheinlichkeit zu erhöhen, wurde mit R 6.4 die sogenannte erweiterte Suche implementiert.

Dies adressiert das Problem, dass Sperren mit Statusabfragen, die mit mehr Vornamen getätigt werden als in der Datenbank gespeichert sind, nicht gefunden werden.

Erfolgt die Statusabfrage mit mindestens 3 Namen (nach der Zusammenfassung und Sortierung von Vor- und Nachname), so werden alle Sperren mit ansonsten gleichen Attributen und allen Teilmengen der Namen der Suchanfrage, die genau einen Namen weniger enthalten als die Suchanfrage, gefunden.

2.3 Beispiele mit mehreren Vornamen / mehreren Nachnamen

Wenn ein Sperreintrag mit mehreren Namensbestandteilen im Feld „Nachname“ oder „Vorname“ eingetragen wurde, wird die Abfrage mit nur einem Namensbestandteil als Treffer gewertet werden.

Wenn die Abfrage einen Namenbestandsteil mehr ausweist als in dem Sperreintrag gespeichert ist, wird dies als Treffer gewertet. Bei mehr als einem zusätzlichen Namenbestandsteil ist dies nicht der Fall.

Beispiele

Abfragen		**Sperreinträge**		
Vorname	**Nachname**	**Vorname**	**Nachname**	**Treffer [ja|nein]**
Julio	Gustavo	Gustavo	Julio	ja
Hans	Lotz	Hans Juergen Wilhelm	Lotz	ja
Lotz	Hans	Hans Juergen Wilhelm	Lotz	ja
Susanne Evylin	Schmidt	Susanne	Schmidt	ja
Schmidt	Susanne Evylin	Susanne	Schmidt	ja
Susanne Evylin Marie	Schmidt	Susanne	Schmidt	nein; mehr als ein zusätzlicher Namenbestandsteil
Karl-Heinz	Werner	Karlheinz	Werner	nein; das Minus erzeugt zwei Worte „Karl“ und „Heinz“
Werner	Karl-Heinz	Karlheinz	Werner	nein; das Minus erzeugt zwei Worte „Karl“ und „Heinz“
Fassbinder	Rudolf	Fassbinder	Herbert	nein

Hans-Jürgen	Schleicher-Schüll	Schüll-Schleicher	Hans-Jürgen	ja
Hans	Schüll	Schüll-Schleicher	Hans-Jürgen	ja
Jürgen	Schleicher	Schüll-Schleicher	Hans-Jürgen	ja
Hans-Jürgen Dieter	Schüll-Schleicher	Schüll-Schleicher	Hans-Jürgen	ja; erweiterte Suche: Suchanfrage mit genau einem Wort mehr, vorhandene Sperre {HANS,JUERGEN,SCHLEICHER,SCHUELL} ist Teilmenge der Suchanfrage
Hans Dieter	Schüll-Schleicher	Schüll-Schleicher	Hans-Jürgen	nein; Suchanfrage und vorhandene Sperre haben die gleiche Anzahl Namen. Keine Teilmengenbeziehung
Hans-Jürgen Dieter-Gustav	Schüll-Schleicher	Schüll-Schleicher	Hans-Jürgen	nein; Kriterium für erweiterte Suche nicht erfüllt: zu viele Namen in der Suchanfrage

3 Umsetzung

Erklärungen von Begriffen, die in den folgenden Abschnitten benutzt werden:

Begriff	Bedeutung
Buchstaben	Die in dt. Sprache bekannten Groß- und Kleinbuchstaben A – Z, a – z, sowie das ß („Eszett“)
Diakritische Zeichen und Ligaturen	Die in **DIN SPEC 91379** definierten normativen Zeichen, die nicht unter Buchstaben, wie oben definiert, zu rechnen sind.
Zeichen	unter Zeichen sind hier alle Zeichen gemeint, die der Zeichensatz **DIN SPEC 91379** definiert

Großbuchstaben	Die in dt. Sprache bekannten Buchstaben A – Z, aber nicht a – z. Auch das ß („Eszett“) ist nicht Bestandteil der Großbuchstaben
Wort	Wie im umgangssprachlichen Sinne zu verstehen. Worte werden durch Leerzeichen voneinander getrennt. Ein Wort besteht aus mind. einem alphanumerischen Zeichen. Auch „28“ ist in diesem Sinn ein Wort.
Leerzeichen	Das Leerzeichen ist das Blank-Zeichen (ASCII 32), welches Worte voneinander trennt. Andere Nicht-sichtbare-Zeichen / Whitespaces sind keine Leerzeichen
entfernen	bedeutet „ausschneiden“ (nicht ersetzen), d.h. wenn ein Zeichen aus einer Zeichenkette ausgeschnitten wird, ist das Zeichen verschwunden und die Zeichenkette um die Anzahl der ausgeschnittenen Zeichen kürzer.
Ziffer	0 – 9
Suchform	Ein oder zwei Zeichen, durch die die diakritischen Zeichen und Ligaturen entsprechend **DIN SPEC 91379/ICAO** ersetzt werden Beispiele: - Suchform für Ä ist AE nach **DIN SPEC 91379** und AE oder A nach **ICAO** - Suchform für Ê ist E

3.1 Abbildung der Daten auf einen Grundzeichensatz und die Transliterationen

Diaktrische Zeichen und Ligaturen werden wie in der folgenden Tabelle nach DIN SPEC 91379 und ICAO ersetzt.

Zeichen	Translitera-tion gemäß DIN SPEC 91379	Zusätzliche Transliteration gemäß ICAO (Seventh Edi-tion, 2015)
AA̋aa̋ÀÁÂÃàáâãĀāĂăĄąǍǎȧẠẠ̈ạạ̈ẢảẤấẦầẨẩẪẫẬậẮắẰằẲẳẴẵẶặ	A	
ÆæǞǟǢǣǼǽ	AE	
Ää	AE	A
Ǻǻ	AA	
Åå	AA	A
BbḂḃḆḇ	B	
CC̀C̄C̆C̈C̕C̣C̦Çcc̀c̄c̆c̈c̕c̣c̦çÇÇ̆çç̆ĆćĈĉĊċČČ̕Č̣čč̕č̣Ƈƈ	C	
DD̂dd̂ĐðĎďĐđḊḋḌḍḎḏḐḑ	D	
EeÈÉÊËèéêëĒēĔĕĖėĘęĚěƏȨȩəḜḝẸẹẺẻẼẽẾếỀềỂểỄễỆệ	E	
FF̀F̄ƒƒ̀ƒ̄Ḟḟ	F	

GG̀gg̀ĜĝĞğĠġĢģǤǥǦǧǴǵḠḡ	G	
HH̄ḤH̱hh̄ḥĤĥĦħȞȟḢḣḤḥḦḧḨḩḪḫẖ	H	
IiÌÍÎÏìíîïĨĩĪḮīíĬĭĮįİıƗǏǐɨḯỈỉỊị	I	
Ĳĳ	IJ	
JJ́J̌jj́Ĵĵǰ	J	
KK̀K̂K̄K̇K̕ƘĶkk̀k̂k̄k̇k̕ƙķĶķĸǨǩḰḱḲḲ̄ḳḳ̄Ḵḵ	K	
ḴHḴhḵh	KH	
LL̂L̥̄L̥ĻlL̂L̥̄L̥ļĹĺĻļĽľĿŀŁłḶḷḺḻ	L	
MM̀M̂M̆M̐mm̀m̂m̆m̐ṀṁṂṃ	M	
NN̂N̄N̆Ņnn̂n̄n̆ņÑñŃńŅņŇňŉŊŋǸǹṄṅṆṇṈṉ	N	
OoÒÓÔÕòóôõŌōŎŏŐőƠơǑǒǪǫǬǭȬȭȮȯȰȱṒṓỌỌ̈ọọ̈ỎỏỐốỒồỔổỖỗỘộỚớỜờỞởỠỡỢợ	O	
ØøŒœǾǿȪȫ	OE	
Öö	OE	O
PP̀P̄P̕P̣pp̀p̄p̕p̣ṔṕṖṗ	P	
Qq	Q	
RŘR̥̄R̥rřr̥̄r̥ŔŕŖŗŘřȒȓṘṙṚṛṞṟ	R	
SS̀S̄S̕S̱ss̀s̄s̕s̱ŚśŜŝŞşŠšȘșṠṡṢṢ̄ṣṣ̄	S	
ßẞ	SS	
TT̀T̄T̈T̕Ƭtt̀t̄t̕ƭŢţŤťŦŧȚțṪṫṬṬ̄ṭṭ̄Ṯṯẗ	T	
Þþ	TH	
UU̇uu̇ÙÚÛÛ̄ùúûû̄ŨũŪūŬŭŮůŰűŲųƯưǓǔỤỤ̄Ụ̈ụụ̄ụ̈ỦủỨứỪừỬửỮữỰự	U	
ǕǖǗǘǙǚǛǜ	UE	
Üü	UE	U
Vv	V	
WwŴŵẀẁẂẃẄẅẆẇ	W	
XxẌẍ	X	
YyÝýÿÿ́ŶŷŸȲȳẎẏỲỳỴỵỶỷỸỹ	Y	
ZZ̀Z̄Z̆Z̈Zȥz̀z̄z̆z̈zȤ́ȥ́ŻżŽȤ̌Žȥ̌ȥ̌žƷǮǯʒẐẑẒẓẔẕ	Z	

Grundlage für die Ersetzung der Buchstaben ist der durch DIN SPEC 91379 herausgegebene Standard „Normative Zeichen“ (Zeichen der Gruppen, die als normativ bezeichnet werden).

Die zusätzliche, transliterierte Suchform nach ICAO wird nur auf die Felder „VORNAME“ und „NACHNAME“ angewandt.

Die Codierung NXX und UXX kann lt. Internetrecherche in spanischen Ausweisdokumenten verwendet werden. Sie wird in der Praxis aber wohl nicht angewendet (das „Ü“ ist kein spanischer Buchstabe). Die Codierungen „UXX“ und „NXX“ wurden daher nicht umgesetzt.

Die Transliteration wird immer nach dem gleichen Verfahren durchgeführt werden.

a) „Ä->AE“, „Å ->AA“, „Ö->OE“, „Ü-UE“ oder

b) „Ä->A“, „Å ->A“, „Ö->O“, „Ü->U“

Beispiele

Die Suchform(en) für die Namensfelder eines Sperr-Eintrags wird durch die Zusammenführung, die Transliteration und eine alphabetische Sortierung des Vornamen und des Nachname des Sperr-Eintrags gebildet. Die Suchformen bestehen ausschließlich aus Großbuchstaben.

Weitere Bearbeitung dieser Felder, die in diesem Kapitel und in folgenden Kapiteln beschreiben sind, werden angewandt.

Sperr-Eintrag VORNAME	Sperr-Eintrag NACHNAME	Transliterierte Suchform nach DIN SPEC	zusätzliche transliterierte Suchform nach ICAO
Rüdiger	Häberle	HAEBERLE RUEDIGER	HABERLE RUDIGER
Peter	Müller	MUELLER PETER	MULLER PETER
Rüdiger-Jörg	Müller	JOERG MUELLER RUEDIGER	JORG MULLER RUDIGER
Hannah	von Müller	HANNAH MUELLER	HANNAH MULLER
Ånne	Aron	AANNE ARON	ANNE ARON

Die Sonderzeichen aus dem durch DIN SPEC 91379 herausgegebenen Standard werden nicht durch ihre Suchform ersetzt, sondern durch genau ein Leerzeichen:

`Genau 1 Leerzeichen: !"#%&()*,-./:;?@[\]_{}¡«·»¿$+<=>^|~¢£¤¥¦§©¬®°±¶×÷€¸¨¯`

Die nicht in der Zeile Sonderzeichen enthaltenen 4 Formen des Apostrophs werden in Abschnitt 3.2 gesondert erwähnt.

3.2 Stufe 1: Allgemeine Normierung (Textfelder)

Schritt	Feld	Beschreibung
1	Alle außer Hausnummer, Postleitzahl und Geburtsdatum	Ligaturen und Diakritika auf *Suchform* abbilden
2	Alle außer Hausnummer, Postleitzahl und Geburtsdatum	` ‘ ´ ' (Apostrophe) *entfernen*
3	Alle außer Hausnummer, Postleitzahl und Geburtsdatum	alle Sonderzeichen aus Abschnitt 1.1.4 durch *Leerzeichen* ersetzen (Sonderbehandlung des Zeichens „Minus“ siehe Abschnitt 1.1.4)

4	Alle außer Hausnummer, Postleitzahl und Geburtsdatum	Mehrfach aufeinander folgende *Leerzeichen* durch ein einziges *Leerzeichen* ersetzen und entfernen von führenden und anhängigen *Leerzeichen*

Umsetzung in Großbuchstaben entfällt, da die Suchformen ausschließlich aus Großbuchstaben bestehen.

3.3 Stufe 2: Feldspezifische Normierung

Schritt	Feld	Beschreibung
5	Vorname, Nachname, Geburtsname Wohnort, Geburtsort, Straße	Alle *Ziffern entfernen*
6	Vorname, Nachname Geburtsname	*Wörter* "DR", "PROF" *entfernen,* Entfernung von Namensteilen bei Auftreten des Wortes „geb“
7	Wohnort, Geburtsort	*Entfernen von Füllwörtern und Zeichenketten*
8	Straße	Alle Leerzeichen (auch innerhalb) *entfernen*
9	Straße	Abschließendes „STR“ und „STRASSE“ *entfernen*
10	Hausnummer	Alle Zeichen, die keine *Ziffern* sind, durch *Leerzeichen ersetzen*
11	Hausnummer	In allen Worten führende Nullen *entfernen*
12	Hausnummer	Mehrfach aufeinander folgende *Leerzeichen* durch ein einziges *Leerzeichen ersetzen*, führende und anhängige *Leerzeichen entfernen*.
13	Postleitzahl	Alle Zeichen, die keine *Ziffern* sind, *entfernen*
14	Geburtsdatum	In Bestandteile Jahr, Monat, Tag zerlegen. Anschließend in jedem Bestandteil Minus entfernen.

3.4 Stufe 3: Entfernen von Füllworten und Zeichenketten

Es wird einen Katalog von Zeichenketten geben, in dem die zu entfernenden Namens- bzw. Ortsbestandteile aufgeführt werden. Aktuell enthält der Katalog folgende Zeichenketten

Zeichenkette	wird entfernt in
PROF	Vorname, Nachname, Geburtsname
DR	Vorname, Nachname, Geburtsname
JETZT	Wohnort, Geburtsort
BAD	Wohnort, Geburtsort
VOR	Wohnort, Geburtsort

AUF	Wohnort, Geburtsort
DER	Wohnort, Geburtsort
BEI	Wohnort, Geburtsort
IM	Wohnort, Geburtsort
IN	Wohnort, Geburtsort
AM	Wohnort, Geburtsort
AN	Wohnort, Geburtsort
OB	Wohnort, Geburtsort

Minus in den Feldern:

Durch die Normierungsregeln wird ein einzelnes Minus bzw. ein doppeltes Minus im Feld „Tag der Geburt“ und im Feld „Monat der Geburt“ im Suchdatensatz zu NULL. Jede beliebige Eingabe im Vornamen bzw. Adresszusatz führt dadurch zu einem Treffer bei sonst gleichen anderen Attributen.

Ausnahme: Feld „Nachname“ und Feld „Jahr der Geburt“ Diese beiden Felder dürfen nicht nur aus einem „-„ bestehen (Nachname) bzw aus „----„ (vier Minuszeichen) bei „Jahr der Geburt“.

GEB in Vor- und Nach- und Geburtsname:

Bei einem transliterierten Vor- bzw. Nachnamen wird die Zeichenkette „GEB“ und alle darauf folgenden bzw. führenden Zeichen entfernt.

Anders als in „Vor-„ oder „Nachname“ werden im Feld „Geburtsname“ **bis zum ersten Auftreten** des Wortes „geb“ (Groß- / Kleinschreibung wird ignoriert) alle Zeichen (inkl. “geb“) abgeschnitten. Zum Beispiel wird aus „Müller geb. Fischer“ nun „Fischer“.

Ein alleine stehendes „geb“ (sonst keine weiteren Zeichen in diesem Feld) wird nicht ausgeschnitten.

STR und STRASSE im Feld Straße:

Abschließendes „STR“ und abschließendes „STRASSE“ werden entfernt. Alle Leerzeichen werden entfernt.