

# Ersatzkennzeichen (EKZ)

## Zulässige Zeichen in Merkmals-Ausprägungen (2023-11-08)

Grundsätzlich werden für alle Merkmals-Ausprägungen Zeichenketten von Unicode-Zeichen erwartet.

Im Bemühen um Einhaltung der österreichischen *E-Government-Konvention DZ-1.0* (mittlerweile leider nicht mehr online verfügbar) werden aber nicht alle möglichen Unicode-Zeichen akzeptiert:

Für sämtliche Merkmale bildet zunächst der gesamte Zeichenvorrat der erwähnten E-Government-Konvention die Menge der gültigen Zeichen (*lateinische Zeichen* und *Sonderzeichen* sowie Leerzeichen). Für die *Namens-Merkmale* wurden die gültigen Zeichen jedoch eingeschränkt: Es fallen Ziffern und *Sonderzeichen* - abgesehen von Bindestrich und Apostroph - weg. Als *Namens-Merkmale* werden hier *Vorname* und *Famname* verstanden.

Überblick zu zulässigen Zeichen nach Merkmalen:

	Merkmale ►	Namens-Merkmale	übrige Merkmale
	Beispiele ▼	<i>Vorname,</i> <i>Famname</i>	insbesondere <i>Ort, Strasse</i> (und Hausnummer), <i>PLZ</i>
● <b>lateinische Zeichen</b> (im westeuropäischen Kernsinn)		(teilweise erlaubt)	erlaubt
○ Ziffern	0, 1, 2, ...	nicht erlaubt	(erlaubt)
○ lateinische Buchstaben	A, b, ...	erlaubt	(erlaubt)
○ lateinische Buchstaben mit diakritischen Zeichen	Ä, ö, Á, ç, ñ, ...	erlaubt	(erlaubt)
○ sonstige lateinische Zeichen	ß, þ, Ɔ, Œ, Ʒ, ...	erlaubt	(erlaubt)
● <b>Sonderzeichen</b>		(teilweise erlaubt)	erlaubt
○ Bindestrich	-	erlaubt	(erlaubt)
○ Apostroph	'	erlaubt	(erlaubt)
○ sonstige Sonderzeichen	/, #, \$, %, &, ™, ...	nicht erlaubt	(erlaubt)
● <b>Leerzeichen</b>		erlaubt	erlaubt

In der Ersatzkennzeichen-Applikation werden folgende reguläre Ausdrücke für zulässige Zeichen verwendet (notiert als Java-Strings):

- für Namens-Merkmale:  
"[ A-Za-z\\u00C0-\\u00D6\\u00D8-\\u00F6\\u00F8-\\u02AF\\u1E00-\\u1EFF' - ]"
- für die übrigen Merkmale:  
"[ A-Za-z\\u00C0-\\u00D6\\u00D8-\\u00F6\\u00F8-\\u02AF\\u1E00-\\u1EFF0-9" +  
"\\u0021-\\u0026\\u0028-\\u002C\\u002E-\\u002F\\u003A-\\u003B\\u003D\\u003F-\\u0040" +  
"\\u005B-\\u0060\\u007B\\u007D-\\u007E\\u00A1-\\u00BF\\u00D7\\u00F7\\u20AC\\u2122'- ]"

(Zum Umfang des Zeichenvorrates der E-Government-Konvention sowie zur Klassifizierung von Zeichen als *lateinische Zeichen* bzw. *Sonderzeichen* bitte in dieser Konvention nachlesen.)

Wenn in einer EKZ-Anfrage an die Ersatzkennzeichen-Applikation bzw. an das REST-Webservice (für einen Fall) in einer Merkmals-Ausprägung nicht zulässige Zeichen auftreten, kommt es (für diesen Fall) im Response zu einem „Fehler in Merkmalen“ (Return-Code 10) mit Text „**MERKMAL X enthält zumindest ein Zeichen, das ...**“, wobei statt *MERKMAL X* der Name des Merkmals angeführt ist.

Nachtrag von 2020-09-01:

**Für die Namens-Merkmale Vorname und Famname wird zusätzlich geprüft, ob zulässige Folgen von Zeichen vorliegen.**

Als *zulässige Folgen von Zeichen* wurden festgelegt: Folgen von *Wörtern*, die jeweils entweder durch ein einzelnes Minus oder ein einzelnes Leerzeichen voneinander getrennt sind. Als *Wörter* sind hier alle Folgen von zulässigen Zeichen - ohne Minus und Leerzeichen - mit Mindest-Länge 1 zu verstehen, wobei darin keine zwei Apostrophe unmittelbar aufeinander folgen dürfen und die Apostrophe nicht an erster oder letzter Stelle dieser Wörter stehen dürfen - mit Ausnahme von [niederländisch] 't in: in/op/van 't.

Folgender *Regulärer Ausdruck* ist geeignet zu prüfen, ob eine Ausprägung durch derzeit automatisch stattfindende Korrekturen mit diesen Regeln in Übereinstimmung gebracht werden kann (Notation als Java-String):

```
"(?i)^([\^ -]+([\^'][\^ -]+)*)([\^ -]+([\^'][\^ -]+)*)*$"
```

Wenn die direkte Überprüfung (also ohne vorherige Auto-Korrekturen) geprüft werden soll, kann folgender *Regulärer Ausdruck* angewandt werden:

```
"(?i)^(((?!IN|OP|VAN)[][T][ -])[\^ -]+([\^'][\^ -]+)*)|((IN|OP|VAN)[][T][ ][\^ -]+))" +  
"([\^ -]+(((?!IN|OP|VAN)[][T][ -])[\^ -]+([\^'][\^ -]+)*)|((IN|OP|VAN)[][T][ ][\^ -]+)))*$"
```