

.INFO IDN Language Table & Policy (German - DE)

Language Code (based on BCP 47):	DE
Version Number:	1.0
Effective Date:	March 16, 2004

This document provides a description of the IDN (Internationalized Domain Names) Language Table and Policy to be used by Afiliias in the .INFO TLD registry for the registration of German language .INFO domains. The policy is consistent and compliant with the ICANN IDN Registry Implementation Guidelines [ICANN] and is intended to be published publicly and registered at the IANA Internationalized Domain Names Language Character Variant Tables Registry [IANA].

1.0 German Language Variant Table

The following is the German Language Variant Table that will be used for the .INFO registry. Note that only the “sanitized” character based on the Nameprep [RFC3491] specifications are included. It is expected that implementers first “sanitize” an IDN according to the Nameprep specifications before applying this table.

The table will use the convention U+XXXX to denote Unicode codepoints for each of the accepted characters for German Language IDN registrations. The table includes entry rows with 2 columns: 1. Unicode Codepoint; and 2. Description; each, separated by a semicolon: “;”.

Registry: .INFO
Language Code: DE
Version Number: 1.0
Effective Date: 2004-03-16

U+002D; HYPEN-MINUS
U+0030; DIGIT ZERO
U+0031; DIGIT 1
U+0032; DIGIT 2
U+0033; DIGIT 3
U+0034; DIGIT 4
U+0035; DIGIT 5
U+0036; DIGIT 6
U+0037; DIGIT 7
U+0038; DIGIT 8
U+0039; DIGIT 9
U+0061; LATIN SMALL LETTER A
U+0062; LATIN SMALL LETTER B
U+0063; LATIN SMALL LETTER C
U+0064; LATIN SMALL LETTER D
U+0065; LATIN SMALL LETTER E
U+0066; LATIN SMALL LETTER F
U+0067; LATIN SMALL LETTER G
U+0068; LATIN SMALL LETTER H
U+0069; LATIN SMALL LETTER I
U+006A; LATIN SMALL LETTER J
U+006B; LATIN SMALL LETTER K

U+006C; LATIN SMALL LETTER L
U+006D; LATIN SMALL LETTER M
U+006E; LATIN SMALL LETTER N
U+006F; LATIN SMALL LETTER O
U+0070; LATIN SMALL LETTER P
U+0071; LATIN SMALL LETTER Q
U+0072; LATIN SMALL LETTER R
U+0073; LATIN SMALL LETTER S
U+0074; LATIN SMALL LETTER T
U+0075; LATIN SMALL LETTER U
U+0076; LATIN SMALL LETTER V
U+0077; LATIN SMALL LETTER W
U+0078; LATIN SMALL LETTER X
U+0079; LATIN SMALL LETTER Y
U+007A; LATIN SMALL LETTER Z
U+00E4; LATIN SMALL LETTER A WITH DIAERESIS (A-UMLAUT)
U+00F6; LATIN SMALL LETTER O WITH DIAERESIS (O-UMLAUT)
U+00FC; LATIN SMALL LETTER U WITH DIAERESIS (U-UMLAUT)

2.0 The Use of Umlauts in the German Language

In determining the Language Policies for German .INFO IDNs, Afilias has engaged with a local authority in Germany regarding the use of German umlaut characters in domain names. This section is an extract of the report provided by Dr. Hardarik Blühdorn and Dr. Klaus Heller at the Institut für Deutsche Sprache (Institute of German Language), Mannheim.

Usually, such combinations of letters as ae, oe and ue are used only to substitute the letters ä, ö and ü ("umlauts"). This is particularly the case when umlauts are not available. The following equivalents are used:

ae ↔ ä
oe ↔ ö
ue ↔ ü

However, cases also exist where the combinations of ae, oe and ue do not function as substitute letters, but occur in regular use. An impression of the frequency of such words in relation to the total number of words in the German language may be gained from a widely used dictionary, e.g. the Duden-Universalwörterbuch (3rd Edition, 2003) [Duden Universal Dictionary], which lists more than 100,000 words. Of these, 147 (a good 1‰) are written with ae, 129 with oe (also a good 1‰) and 1766 with ue (just on 2%). Combinations using ue occur more than ten times more frequently than those with ae and oe.

The following case groups may be distinguished:

1. Words borrowed from other languages that are usually only written with ae, oe or ue and not with ä, ö or ü but where the umlauts are pronounced, e.g. Aerobic, Praesens; Oenothera (evening primrose); Fondue, Revue. This is a very small group containing few words, frequently from technical usage, that only occur infrequently in general speech.
2. Names, particularly surnames, that are usually only spelt with ae, oe or ue and not with ä, ö or ü but where the umlauts are also pronounced, e.g. Haeckel; Goethe; Mueller. This group is difficult to quantify, particularly as the distribution of family names varies from region to region. A random sample from the city of Mannheim's telephone book revealed that 1100 entries of the name Müller also included eight entries of

Mueller (less than 1%). There were also five entries of Moeller alongside 37 entries of Möller (more than 10%). Also, in regard to the beginning of surnames, Oe (Oehler, Oechsler, Oehschläger, Oehmann, etc.; a total of 36 different names in the Mannheim telephone book) occurs significantly more frequently than Ae (in Mannheim only Aehnel) and Ue (eight names, e.g. Uebelhör, Ueberle, Uelzhöffer). As shown above, some surnames use ae, oe or ue alongside ä, ö or ü as in Oehschläger or Uelzhöffer. Generally, speaking, surnames in which ae, oe or ue occur regularly constitute a group that is not insignificant where quantity is concerned.

3. Words from other languages and names (from other languages) where ae, oe or ue are regularly pronounced as separate vowels, e.g. Aerodynamik, Paella, Michael, Raffael; Boethius, Noetik, Oboe, Poesie; Buenos Aires, Duell, Lues, Muezzin. This group is a little larger than Group 1 and some of the words and names it includes occur more frequently. But generally, this group of words is also small.
4. Names in which the e indicates an elongation of the previous a, o or u (surviving from former spelling conventions in Lower German), e.g. in such place names as Aertsen, Raesfeld; Coesfeld, Soest; Klues (district of Flensburg), Kluess (district of Güstrow) and in such people's names as Goes and Lueg. This group is the smallest of those listed here. The names it includes are very rare special cases.
5. Words from other languages or names from other languages in which ae, oe or ue are pronounced according to non-German rules, as in Kim Dae Jung, Taekwondo; Boeing; Baguette, Barbecue, Blues, Cinquecento, Gigue, oblique. This group is also not large and consists in the main of rarely occurring words. Ue occurs somewhat more frequently than ae and oe because words with the ue spelling have arrived in the German language from a variety of other languages (English, French, Italian, etc.)
6. Words in which u and e are used in conjunction where, however, the u is part of one of the diphthong spellings au or eu or the qu combination of letters, e.g. Litauen, Dauer, lauern; Abenteuer, Feuer, steuern; Quelle, quetschen, quer, bequem. This group constitutes around 80% of the words using the ue combination of letters in the Duden Universalwörterbuch. It is a significant explanation why the ue combination of letters occurs more frequently than ae and oe.
7. Word forms whose root ends with o or u and whose inflection suffix commences with e. Here, oe and ue are always pronounced as separate vowels, e.g. Protozoen; Individuen, Kontinuen, genauer, neues. Such word forms occur rather infrequently although most German inflection suffixes end with e (-er, -en, -es, etc.). But only few words whose roots end with o or u may acquire such endings. Except for the extremely rare foreign substantives with -oon and -uum (as in Protozoon, Individuum, etc.), these are almost exclusively substantives and adjectives with -au and -eu, e.g. Frau and scheu, a total of around 40 word roots.
8. Derivations with the ending -er, particularly for place names, which then end with a, o or u, e.g. Danaer(-geschenk), Riesaer; Togoer; Korfuier, Turkuier, Krakauer, but also some verbs, e.g. Geheimnistuer, Fleischhauer, Wiederkäuer. Place names ending with a, o or u are not rare, but only few create derivations to -er. A consonant is more frequently inserted before the -e in order to prevent two vowels from following each other, e.g. in Amerikaner or Tokioter. Verbs whose root ends in a, o or u are extremely rare in German. So generally, Group 8 is also small. The word creations belonging to this group are rarely used.
9. (Technical) derivations to -eder, -edrisch, -esk, -ell, -ent, -end, enz, ente and a few other endings whose roots end with a, o or u, e.g. Tetraeder (four-surfaced body), oktaedrisch (eight-surfaced body), kafkaesk; aktuell, visuell, sexuell; Distribuent, Obstruent, kongruent; Minuend; Konfluenz; Konstituente. Adjectives ending with -uell are relatively frequent and used often, according to the Rückläufiges Wörterbuch der deutschen Gegenwartssprache by Erich Mater (1967), the total is around 20. However, Group 9, is also small in relation to the total number of words in the German language.

10. Word compositions whose first component ends with one of the a, o or u vowels and whose second component commences with e and in which ae, oe and ue are always pronounced as separate vowels, e.g. Afrikaexperte, Sofaecke, Sienaerde; Kinoerlebnis, soeben; Kanuexkursion, zuerst. Mater (1967) records 604 words that end with a, 432 words with o and 646 words with u. The Duden-Universalwörterbuch lists more than 5000 words that commence with e. Due to the fact that in German as opposed to other languages it is normal to create word compositions spontaneously and frequently, Group 10 is theoretically very large, the largest of all the groups listed here. However, most of the theoretically possible compositions are probably not created due to a lack of context in which they might be used, e.g. Guttaperchaenthusiasmus or Kakaduerschwernis. Also, particularly word compositions where an a, o or u ending encounters an e at the beginning of a word, are frequently written with a hyphen to facilitate reading (e.g. Afrika-Experte). That's why in practice this group is considerably smaller than could initially be assumed. The Duden-Universalwörterbuch records only six compositions as words in which a encounters e, 48 compositions where o encounters e, and 17 where u encounters e. The relative frequency of compositions with o and e is due to series-creating prefixes, such as elektro-, mikro-, mono-, photo-, stereo-, thermo-, etc., which are used to create technical compositions, such as Elektroenzephalogramm, Mikroelektronik, Photoelement and so on.

All the above cases occur occasionally in relation to the total number of words in the German language. They are not characteristic for current spelling in German.

The Group 1 to 5 words are particularly difficult to recognize and unsystematic in their behaviour in regard to the umlauts ä, ö and ü. In comparison, the words and word forms in Groups 6 to 10 are more easily recognized, i.e. combinations from au, eu and qu with e, and inflection forms, derivatives and word compositions where a, o and u occur at the end of the root or the first syllable and e at the beginning of the ending or the second syllable. In our opinion, Groups 6 to 10 may from the outset and with relatively little effort be excepted from substitution with ä, ö and ü spellings.

3.0 The Eszet Character: “ß”

Based on the Nameprep [RFC3491] specification for a Stringprep [RFC3454] profile for IDNA [RFC3490], the Eszet “ß” character used in the German script will be mapped to a double “s” string (i.e. “ss”).

4.0 Conclusion

Due to the nature of the umlaut characters (ä, ö and ü) with its counterparts (ae, oe and ue), as explained in Section 2.0, which are not bilaterally definitively interchangeable and equivalent, no character variant mapping will be provisioned for German Language IDNs registered.

Normal domain requirements, such as the prohibiting of the use of “hyphen” (U+002D) as the first or last character will continue to be enforced. Domain length limit of 63 characters will be determined based on the IDNA “punycode” domain generated.

5.0 References

[RFC3491] P. Hoffman. and M. Blanchet, "Nameprep: A Stringprep Profile for Internationalized Domain Names (IDN)", RFC 3491, March 2003.

- [RFC3492] A. Costello, "Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA)", RFC 3492, March 2003.
- [RFC3454] P. Hoffman, M. Blanchet, "Preparation of Internationalized Strings", RFC 3454, December 2002
- [RFC3490] Faltstrom, P., Hoffman, P. and A Costello, "Internationalizing Domain Names in Applications (IDNA)", RFC 3490, March 2003.
- [UNICODE] The Unicode Consortium, "The Unicode Standard",
<http://www.unicode.org/unicode/standard/standard.html>
- [ICANN] Guidelines for the Implementation of Internationalized Domain Names, Version 1.0, 20 June 2003, <http://www.icann.org/general/idn-guidelines-20jun03.htm>
- [IANA] Internationalized Domain Names Language Character Variant Tables Registry,
<http://www.iana.org/assignments/idn/>

Contact Information:

Contact: Ram Mohan

Address: Building 3, Suite 105, 330 Welsh Road, Horsham, PA 19044, USA

Tel: +1 215 706 5700

Fax: +1 215 706 5701

Email: rmohan@afiliass.info

Website: <http://www.afiliass.info>

Document URL: http://www.afiliass.info/register/idn/DE-INFO_IDN_Language_Table-v1-1.pdf