



# Bald kennt das Internet auch Umlaute

WWW-Adressen mit Sonderzeichen kommen noch in diesem Jahr / Verschiedene Standards

Freilich geht es auch ohne, schöner und korrekter aber ist es mit Umlauten in der Schrift. Bei Computern haben sich gute Umlautdarstellungen durchgesetzt, seit Dateinamen länger als acht Zeichen sein dürfen. Doch nicht die lange Geschichte umlautlicher Irrungen und Wirrungen wollen wir hier nachzeichnen, sondern die neueste Lösung bei Internet-Namen beschreiben.

Japaner, Chinesen, Koreaner, Araber, einfach alle mit einem nichtlateinischen Alphabet mußten bei Internet-Namen immer mühsam umdenken. Die eigentlichen Texte auf den Webseiten erscheinen schon lange in Landeszeichen, nur eben die Adresse oben im Eingabefeld des Browsers erlaubt nach wie vor nur eine Auswahl einfacher Buchstaben, keine Sonderzeichen, und Groß- und Kleinbuchstaben werden gleichgesetzt.

Die für De-Domänen zuständige Registrierungsstelle Denic wird noch in diesem Jahr Umlaute zulassen. Die Beliebigkeit der Großschreibung bleibt, Sonderzeichen wie etwa das für E-Mail reservierte Zeichen „@“ dürfen weiterhin nicht genutzt werden. Der Standard für Sonderzeichen liegt seit dem 11. Februar 2003 vor.

## Komplizierte Unicode-Zeichen

Bei der Sonderzeichenkodierung griff man auf die schier unendliche Zahl von Unicode-Zeichen zurück, die von 0 bis 11.534.335 (hexadezimal 10FFFF) reichen können und Platz für die exotischsten Alphabete bieten. Mit einem komplizierten, aber eindeutigen und umkehrbaren Verfahren, nach dessen Erfinder Adam Michael Costello, einem Studenten in San Franzisko, zunächst „AMC-ACE-Z“ genannt (ASCII compatible encoding Nummer 26 gleich Z, neuerdings „Punycode“), werden diese Unicode-Zeichen für das Internet auf eine etwas längere Kette einfacher, alter ASCII-Zeichen zurückgeführt – ASCII ist schon seit der Zeit der Fern-

schreiber der amerikanische Standardcode für „Information Interchange“. Die intern genutzte Zeichenkette ist natürlich auf den ersten Blick unverständlich. Weil im Internet sogenannte Domänennamenserver die jeweils angesteuerte Domäne, sagen wir [www.faz.de](http://www.faz.de), blitzschnell in eine aktuelle Internet-Protokollnummer umwandeln können müssen, bleiben die im Netz genutzten Domänennamen also wie bisher sonderzeichenfrei. Die Punycode-Wandlung passiert erst einmal lediglich im Browser, wohl zunächst nur im Microsoft-Explorer. Dafür ist ein Zusatz nötig. Ein solches Plugin verbreitet Verisign als „i-nav“ seit Ende 2002 und handelte sich sofort Proteste der Internet-Gemeinde ein, aber nicht, weil es statt Puny noch den alten Race-Code erzeugt, sondern weil es nicht protokollkonform arbeitet.

Im Netz sind die IDN-Domänen an einem klassischen Umschaltcode, einer Vorschaltzeichenfolge (Escape-Sequenz) erkennbar, so wie etwa in HTML das &-Zeichen ein Sonderzeichen einleitet. Der Umschaltcode für die IDN-Namen ist „xn--“. Er wurde mit einem kuriosen, aber öffentlich nachvollziehbaren Zufallsverfahren aus 18 möglichen Varianten von xb bis xy ausgelost – lauter bis jetzt selten vorkommende Zeichenverbindungen: Die am 11. Februar 2003 im „Wall Street Journal“ in Hunderten veröffentlichten Umsätze von zwölf Aktien mit I von IMS bis Intel wurden abgeschrieben, die Kette aus 81 Zeichen standardmäßig Message-Digest-5-verschlüsselt und dieses Signet dann durch die Variantenzahl geteilt, der Rest bestimmte endlich Variante 11: xn. Für et was muß die Börse ja gut sein!

Seit dem 25. Februar 2002 hatte die für deutsche Domänen zuständige Denic bereits Bindestriche auf Platz drei und vier verboten, damit IDN-Domänen auf jeden Fall erkennbar sind. Wer mag, kann sich mit einem auf Unicode gestellten Browser seine Umlautdomäne anzeigen lassen. Ein Konverter steht bei <http://josefsson.org/>

idn.php. So wird beispielsweise „Jörn“ richtig zu „xn--jrn-sna“. Der störende Umlaut wird zunächst einfach ausgelassen und am Ende nachcodiert.

## Ausreißer bei den Umlauten

Bei Com-Domänen hat Verisign seit etwa zwei Jahren Umlaute zugelassen, setzt sie allerdings mit einem anderen Code um, sogenanntem Race, und das mit einem umstrittenen Browser-Zusatzprogramm. Später – vielleicht 2004 – will Verisign ebenfalls auf Punycode übergehen. Die Codes sind nicht kompatibel. Im Netz selbst herrscht Zwiespalt, denn eigentlich müßten die Wandlungen dort die Domänennamenserver machen, wovon sie aber einstweilen wohl zurückschrecken.

Wieder andere, gut funktionierende Umlautdomänen gibt es schon seit ein paar Jahren, etwa [Malmö.Nu](http://Malmö.Nu) über die polynesischen Insel Niue, [Südtirol.cc](http://Südtirol.cc) über die 27 Kokos-Koralleninseln (630 Einwohner). Ein Hotel in Naturns nutzt frech mit [Südtirol.Bz](http://Südtirol.Bz) eine Domäne aus Bz, nicht Bozen, sondern aus Belize, dem früheren Britisch-Honduras. Die hier genutzten Codes sind weder Race noch Puny, dafür wandeln sie die Namenserver der jeweiligen Registrierungsstellen für x-beliebige, zusatzlose Browser. Das Durcheinander ist abzuwarten: Werden Browser nun Umlaute bei „com“ in Race, bei „De“ in Puny und bei „Bz“ gar nicht wandeln?

Die vorsichtige deutsche Zulassungsstelle Denic wird die nun standardisierten Umlautdomänen erst nach weiteren europäischen Klärungen freigeben, etwa bezüglich Groß- und Kleinbuchstaben. So mag das Ü klein stets zum ü werden, aus dem H wird aber griechisch ein Eta und nicht immer unser h. Allzuviel sollte man sich von Müller.De (xn--mller-kva.De), Schröder.de (xn--schrder-dla.de) und ähnlichem also nicht erwarten. Selbst E-Mail-Adressen sind noch nicht auf den neuen IDN-Code umgestellt, so daß [Info@Müller.De](mailto:Info@Müller.De) gewiß nicht bei Mueller landet. FRITZ JÖRN