

Zur Ordnung und Codierung der Umlautbuchstaben

B. Eversberg, Universitätsbibliothek Braunschweig

11.2.1998

Anhang 2

Behandlung von Diakritika in anderen Sprachen

Nur Sprachen mit lateinischem Alphabet sind hier berücksichtigt, und nur solche mit einem gewissen Aufkommen in deutschen Bibliothekskatalogen, also beispielsweise nicht Baskisch oder Maltesisch, auch nicht das sehr akzentreiche Vietnamesisch.. Sprachen ohne Besonderheiten der Ordnung, wie Französisch, Italienisch, Niederländisch und Portugiesisch, sind hier nicht eigens aufgeführt.

Die hier genannten Regeln beziehen sich auf Wörterbücher. In der Regel ist wohl die Praxis in Bibliothekskatalogen gleich, ob dies aber auch auf heutige, EDV-gestützte Kataloge zutrifft, konnte nicht so schnell eruiert werden; einzelne Beispiele gibt es für beides. Die Zusammenstellung soll vielmehr nur zeigen: eine Parallele zur deutschen Umlautproblematik, d.h. Gleichwertigkeit eines Sonderbuchstabens mit einer Kombination aus zwei Buchstaben, gibt es in anderen Sprachen nicht. Deutlich wird aber auch: es kann keinen Sortieralgorithmus geben, der in einer mehrsprachigen Datenbank (wie es Bibliothekskataloge typischerweise sind) alle Einträge sprachspezifisch korrekt behandelt. Das geht schon deshalb nicht, weil dann sogar jedes Datenfeld eine Sprachkennung haben müßte, denn es gibt ja z.B. (aber nicht nur) bei Übersetzungen jede denkbare Kombination: deutscher Name und ungarischer Titel, türkischer Name und deutscher Titel, usw. Bis heute hat aber nicht einmal jeder Datensatz eine Sprachkennung.

Mit dem Zeichen '<' wird angedeutet, wie die Buchstaben angeordnet sind, "z < zh" bedeutet also, daß "zh" hinter allem kommt, was mit "z" beginnt. a < á < b heißt, daß á zwischen a und b steht.

Mit '=' wird Gleichordnung bezeichnet, z.B. o = ó.

Albanisch

c < ç < d < dh, g < gj, l < ll, n < nj, s < sh, t < th, x < xh, z < zh

Dänisch und Norwegisch

z < æ < ø < å

Estnisch

z < õ < ä < ö < ü

Finnisch

z < ä < ö

Isländisch

a < á < b und Entsprechendes gilt für e, i, o und u

d < ð < e und y < ý < z < þ < œ

In unseren Altkatalogen ist ð = d , þ = th und œ = oe gesetzt worden.

Polnisch

alle akzentuierten Buchstaben ordnen hinter den Grundbuchstaben:

a < ą < b, e < ę < f, c < ć < d, l < ł < m
n < ń < o < ó < p, s < ś < t, z < ź < ż,

Rumänisch

a < ă < b, s < ș < t, t < ț < u.

Schwedisch

z < å < ä < ö

Spanisch (Castellano)

Akzent und Trema haben keine Ordnungswirkung (o = ó = ö); das Trema soll wie im Französischen nur andeuten, daß der Buchstabe getrennt vom vorangehenden Vokal ausgesprochen wird.

Es gibt folgende Sonderregeln:

c < ch < d, l < ll < m, n < ñ < o, r < rr < s

(Ein Beispiel, damit die ungewohnte Besonderheit des ch ganz klar wird: lectura < lecha < leda)

Jedoch wurde um 1995 offiziell die Sonderrolle von ch, ll und rr abgeschafft, nicht aber die des ñ.

Das Katalanische teilt diese Besonderheiten nicht. Zwar gibt es das "ele geminado" l·l als Sonderbuchstabe, doch dieser wird wie ll geordnet.

Tschechisch und Slowakisch, ähnlich auch Kroatisch, Slowenisch, Litauisch und Lettisch

(nur das Tschechische besitzt alle diese Zeichen. Litauisch und Lettisch haben ansonsten mit den anderen wie auch miteinander sehr wenig gemeinsam!)

c < č < d, r < ř < s < š < t = t', z < ž

Kroatisch zusätzlich: c < č < ć < d < dž < đ < e

Slowakisch zusätzlich: a < ä (ein kurz ausgesprochenes a)

Sonstige Akzente haben keine Wirkung, also z.B. e = é = ě, n = ň, u = ů,

Die ungewöhnlichste Eigenheit des Tschechischen und Slowakischen: h < ch < i.

Türkisch

c < ç < d, g < ğ < h < ı < i,
o < ö < p, s < ş < t, u < ü < v.

Ungarisch

o = ó < ö = ő < p, u = ú < ü = ű < v
