

Reinhold Heuvelmann

Diakritika in Unicode
































































Ç ↔ C + ¸

Code Charts

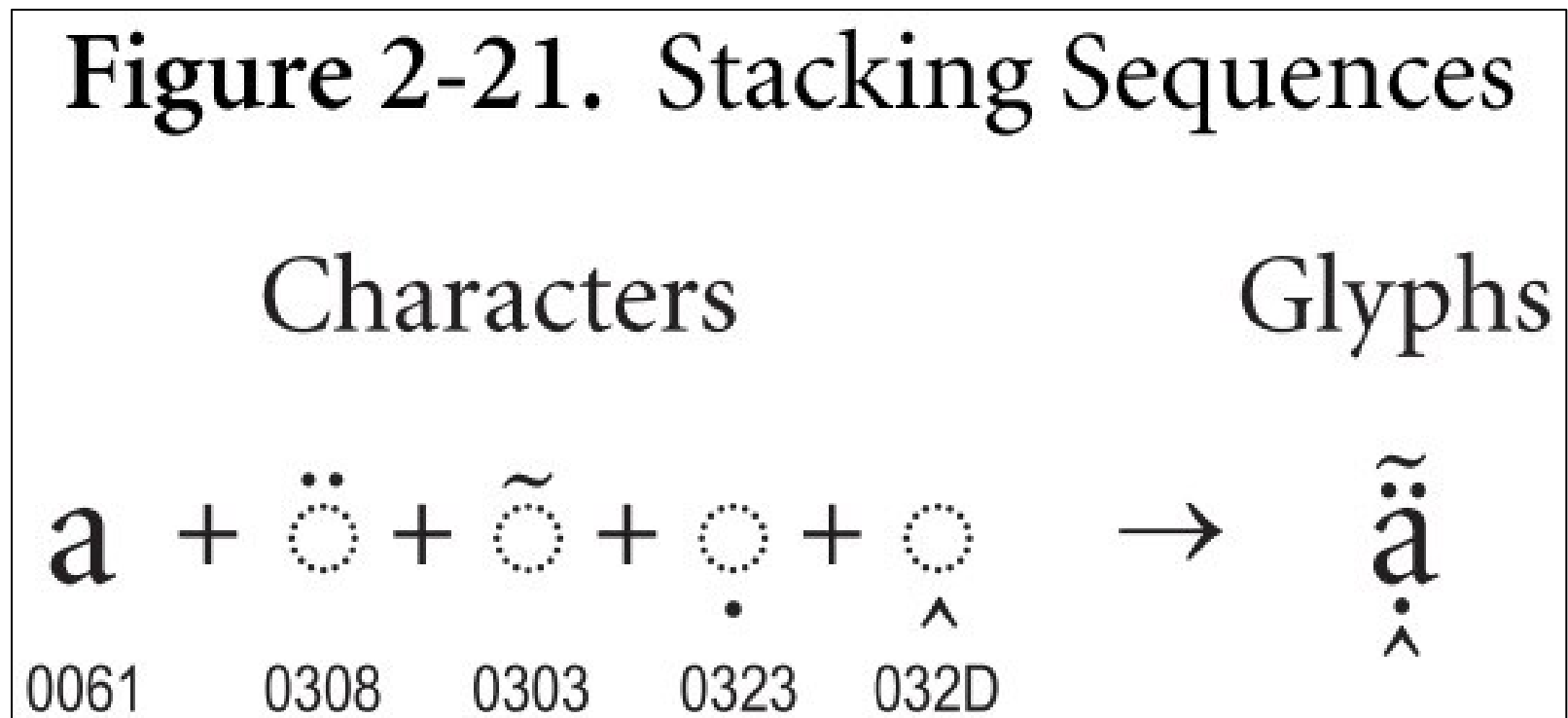
<http://www.unicode.org/charts/> , Stichwort "combining"

<i>Superscripts and Subscripts</i>	P
Combining Marks	P
<i>Combining Diacritical Marks</i>	S
Combining Diacritical Marks Extended	S
Combining Diacritical Marks Supplement	C
<i>Combining Diacritical Marks for Symbols</i>	M
<i>Combining Half Marks</i>	M

- <http://www.unicode.org/charts/PDF/U0300.pdf>
- <http://www.unicode.org/charts/PDF/U1AB0.pdf>
- <http://www.unicode.org/charts/PDF/U1DC0.pdf>
- <http://www.unicode.org/charts/PDF/U20D0.pdf>
- <http://www.unicode.org/charts/PDF/UFE20.pdf>

	030	031	032	033	034	035	036
0	 0300	 0310	 0320	 0330	 0340	 0350	 0360
1	 0301	 0311	 0321	 0331	 0341	 0351	 0361
2	 0302	 0312	 0322	 0332	 0342	 0352	 0362
3	 0303	 0313	 0323	 0333	 0343	 0353	 0363
4	 0304	 0314	 0324	 0334	 0344	 0354	 0364
5	 0305	 0315	 0325	 0335	 0345	 0355	 0365
6	 0306	 0316	 0326	 0336	 0346	 0356	 0366
7	 0307	 0317	 0327	 0337	 0347	 0357	 0367
8	 0308	 0318	 0328	 0338	 0348	 0358	 0368

Stacking Sequences, Beispiel 1



<http://www.unicode.org/versions/Unicode9.0.0/ch02.pdf>

Stacking Sequences, Beispiel 2

á

LATIN SMALL LETTER A WITH CIRCUMFLEX AND ACUTE

â

LATIN SMALL LETTER A WITH CIRCUMFLEX + COMBINING ACUTE

LATIN SMALL LETTER A + COMBINING CIRCUMFLEX + COMBINING ACUTE

â

LATIN SMALL LETTER A ACUTE + COMBINING CIRCUMFLEX

LATIN SMALL LETTER A + COMBINING ACUTE + COMBINING CIRCUMFLEX

Aus den FAQ zu Characters and Combining Marks

Q: Why are **new combinations of Latin letters with diacritical marks** not suitable for addition to Unicode?

A: There are several reasons. First, Unicode encodes many diacritical marks, and the **combinations can already be produced**, as noted in the answers to some questions above. If precomposed equivalents were added, the number of multiple spellings would be increased, and **decompositions would need to be defined and maintained** for them, adding to the **complexity of existing decomposition tables** in implementations.

...

Aus den FAQ zu Characters and Combining Marks

...

Finally, **normalization form NFC (the composed form favored for use on the Web) is frozen**—no new letter combinations can be added to it. Therefore, the normalized NFC representation of any new precomposed letters would still use decomposed sequences, which can already be expressed by combining character sequences in Unicode. Nothing would be gained by adding the letter with diacritical mark as a precomposed character; on the contrary, adding such a letter would add one or more multiple spellings to be reckoned with, incrementally complicating all Unicode implementations for no net gain.

Kombinationen, Kombinationen

List of precomposed Latin characters in Unicode

From Wikipedia, the free encyclopedia

This is a list of **precomposed Latin characters in Unicode**.^{[1][2]} Unicode typefaces (e.g. Fixedsys Excelsior) may be needed for these to display correctly.

Contents [hide]

- 1 Letters with diacritics
- 2 Digraphs and ligatures
- 3 Other characters
- 4 See also
- 5 References

Letters with diacritics [edit]

	Aa	Ææ	Bb	Cc	Dd	Ee	Ff	Gg	Hh	Ii	Jj	Kk	Ll	Mm	Nn	Oo	Øø	Pp	Qq	Rr	Ss	Tt	Uu	Vv	Ww	Xx	Yy	Zz	Ʒ
acute	Áá	Ææ		Ćć		Éé		Óó		Íí		Ķķ	Ĺĺ	Ļļ	Ńń	Óó	Øø	Þþ		Ŕŕ	Śś		Úú		Ŵŵ		Ýý	Žž	
acute and dot above																					Šš								
breve	Ăă					Ĕĕ		Œœ		Ĭĭ						Ŏŏ							Ūū						
breve and acute	Ȧȧ																												
breve and dot below	Ȱȱ																												
breve and grave	Ȫȫ																												
breve and hook above	Ȭȭ																												
breve and tilde	Ȯȯ																												
breve below									Ҁҁ																				
cedilla	ȺȻ			ȼȽ	Ⱦȿ	ȼȽ	ȼȽ	ȼȽ	ȼȽ	ȼȽ	ȼȽ	ȼȽ	ȼȽ	ȼȽ	ȼȽ	ȼȽ	ȼȽ	ȼȽ	ȼȽ	ȼȽ	ȼȽ	ȼȽ	ȼȽ	ȼȽ	ȼȽ	ȼȽ	ȼȽ	ȼȽ	ȼȽ

https://en.wikipedia.org/wiki/List_of_precomposed_Latin_characters_in_Unicode

In der Anwendung durch die DNB

- aktuell keine Zeichensatz-Konversion bei der Erstellung von bibliografischen Daten
- "garbage in - garbage out"
- MARC 21 ist neutral in Bezug auf Unicode composed vs. decomposed
- Wo beginnen, wo enden?
- Tools sind vorhanden:
`"uconv -f utf-8 -t utf-8 -x NFC [Datei]"`
 (mit Dank an Johann Rolschewski)

Danke

r.heuvelmann@dnb.de

MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media, CHARACTER SETS AND ENCODING OPTIONS:
Part 3: Unicode Encoding Environment

<https://www.loc.gov/marc/specifications/speccharucs.html>

Part 4: Conversion Between Environments

<https://www.loc.gov/marc/specifications/speccharconversion.html>

Assessment of Options for Handling Full Unicode Character Encodings in MARC21

<https://www.loc.gov/marc/marbi/2004/2004-report01.pdf>

<https://www.loc.gov/marc/marbi/2005/2005-report01.pdf>