Proposal to encode a combining diacritical mark for Low German dialect writing

2008-10-25

Karl Pentzlin, *karl-pentzlin@europatastatur.de*

The following combining diacritical mark is proposed:

U+1DFD   COMBINING TRIPLE BREVE BELOW

→ 035C combining double breve below
• to be applied to the first letter of any three-letter combination
• used with "sch", "Sch", and "SCH" in Low German dialect writing

The **combining triple breve below** is used in the "Rheinische Dokumenta" [2], an orthography promoted by the "Landschaftsverband Rheinland" (a cultural authority of the southern part of the German federal state Nordrhein-Westfalen) during the 1980s for the local Low German dialects.

It is used there to mark the trigraph "sch" (which is used to denote the voiceless ʃ sound, like in standard High German orthography) denoting the voiced ʒ sound. Thus, it may be applied to "sch" in lower case, "Sch" in title case, and "SCH" in upper case.

The proposed properties are:

1DFD;COMBINING TRIPLE BREVE BELOW;Mn;233;NSM;;;;;N;;;;;

The triple breve below is to be input after the first character of the combination it applies to, analogous to the double diacritics like U+035C COMBINING DOUBLE BREVE BELOW.

Thus, the correct way to enter a "sch" trigraph with triple breve below is:
U+0073 U+1Dxx U+0063 U+0068.

Bibliography:

[1]   CORNELISSEN, Georg, HONNEN, Peter , LANGENSIEPEN, Fritz (ed.):
       Das rheinische Platt · Eine Bestandsaufnahme
       (Handbuch der rheinischen Mundarten, Teil 1: Texte)
       Köln 1989, ISBN 3-7927-0689-X

[2]   HONNEN, Peter:
       Rheinische Dokumenta. Lautschrift für rheinische Mundarten
       Pulheim 1987, ISBN 3-7927-0947-3

[3]   http://de.wikipedia.org/wiki/Rheinische_Dokumenta as of 2008-10-25 *(German)*

[4]   SCHMITZ, Friedhelm, HINTZEN, Josef:
       Dę Lüschbangk. Plattdeutsche Anekdoten aus Jüchen und Umgebung
       Jüchen 1985, private edition, retrieved 2008-10-25 at:
        http://www.beepworld.de/memberdateien/members85/friedhelmschmitz/dielnschbank1.pdf

Der stimmhafte Reibelaut sch (Garage) wird durch einen untergesetzten Bogen gekennzeichnet. Dieser Laut ist in den Mundarten des südlichen Rheinlands häufig anzutreffen: intswische (inzwischen), gleisch (gleich).

Fig. 1: Sample from [2], p. 16, introducing the proposed diacritical mark.

Die **Rheinische Dokumenta** ist die Anfang der 1980er Jahre entwickelte und 1986 vom Landschaftsverband Rheinland vorgestellte Lautschrift für die rheinischen Mundarten des nördlichen und südlichen Niederrheins, des westlichen und zentralen Rheinlands, des Bergischen Landes, des Westerwalds, der westlichen und östlichen Eifel, der Unter- und der Obermosel und der Region um Hunsrück und Nahe.

Neben den im Hochdeutschen üblichen Buchstaben nebst Umlauten ä, ö und ü finden neun Sonderzeichen Verwendung:

a̦ ä̦ ch̦ e̦ o̦ ö̦ r̦ ș șch̦

Fig. 2: Sample from [3], explaining the Rheinische Dokumenta and showing the proposed diacritical mark.

Mundart: Mutterschied (Hunsrück)

Im Schpäätjoa, wän de̦ Ää̦re̦ in waa un di Șou fät ge̦nuuch schiin, is bai de̦ Doorfmätska e̦n Tärmiin foor de̦ Schlachtdaach ousge̦mach woa. Am O̦o̦me̦nt foorhea is di Muul med-e̦m Schupkare̦ ge̦huult woa, de̦ Wäschkäșe̦l met Wașa ge̦filt un Haitsmadriaal baraat ge̦läät, das am friie̦ Mo̦o̦rje̦ ange̦haitst weere̦ ko̦nt. Wän de̦ Mätska ko̦m is, hot mișe̦ de̦s Wașa ko̦ch̦e̦. Wän-s dan șo weit waa, is di Muda o̦o̦re̦ de̦ Fada in de̦ Schdal gang un ho̦t de Sou e̦n Schtrik aan e̦n Hinabään ge̦bun un șe̦ foa di Deer ge̦foort, wo șe̦ dan koats an-e̦ Ho̦o̦ge̦ o̦o̦re̦ Ring draan ge̦bun is woa. Intswișch̦e̦ hot șisch de̦ Mätska met de̦ Aks be̦wave̦nt un ho̦t do̦o̦met di Șou uf de̦ Schdeere̦ ge̦schlaan. Un wän șe̦ nit gleișch umge̦fal is, ho̦r-a no̦ch̦-e̦mo̦o̦ no̦o̦ge̦schlaan. Dan ho̦r-a das groos Mäșa ge̦nom un ho̦t di Șou in de̦ Hals e̦ninge̦schdo̦ch un di Schlaacho̦o̦ra doorischge̦schniit.*

Fig. 3: Sample from [2], p. 14, showing the proposed diacritical mark.

495 Mandel (Vbg. Rüdesheim)

S   Enck, Helmut, Jg. 1928, geb. in Mandel, lebt auch dort, Winzer.
A   1982, 3.54 min.
T   Weihnachten, Mandel um 1933, spontan gesprochen.

Doo waa de̦ alt O̦o̦ba no̦ch, de̦ °Ängge̦, alșo maim Fade̦r șai Fade̦r, dee̦r hat doo o̦ve̦ șai Schdup ge̦hat, un de̦ ho̦t doo ge̦wo̦o̦nt und alșo das Șäkrädärșch̦e̦ schdään [. . .], un isch waa doo șälișch̦e̦mo̦o̦[1] fine̦f, șäks Joe̦r alt. Ich ko̦nd-e̦ bisje̦ friie̦r lääșe̦ wii dii No̦rmaale̦, alșo ich ko̦nt mit finf Joe̦r lääșe̦ o̦o̦de̦r, däșch-e̦[2]: un läișe̦, das waa drum, isch ho̦n halt ge̦lääșe̦, șoo mach ich-s ho̦it no̦ch, alșo isch lääșe̦ tsäntne̦rwais Bișch̦e̦r

Fig. 4: Sample from [1], p. 495, showing the proposed diacritical mark.

E̦t wo̦o̦r tseen op tseen, Johan wo̦o̦r alt jans närvöös op de̦ Bangk e̦röm-am-rödșch̦e̦. Hä kre̦esch wie̦r e̦ns de̦ Ue̦r e̦ruus. Di wo̦o̦re̦n am Ärtsäle̦, e̦t wo̦o̦r alt tswansisch op tseen. Op eemo̦o̦l șäät Hain:

Fig. 5: Sample from [4], p. 14, showing the proposed diacritical mark.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and details before filling this form.**
Please ensure you are using the latest Form from http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest** *Roadmaps*.

A. Administrative

1. **Title:** *Proposal to encode three combining diacritical marks for German dialect writing*
2. Requester's name: *Karl Pentzlin*
3. Requester type (Member body/Liaison/Individual contribution): *Individual Contribution*
4. Submission date: *2008-10-25*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *yes*
   (or) More information will be provided later *no*

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters) *no*
      Proposed name of script:
   b. The proposal is for addition of character(s) to an existing block: *yes*
      Name of the existing block: *1DC0-1DFF Combining Diacritical Marks Supplement*
2. Number of characters in proposal: *1*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary   B.1-Specialized (small collection)   X   B.2-Specialized (large collection)
   C-Major extinct   D-Attested extinct   E-Minor extinct
   F-Archaic Hieroglyphic or Ideographic   G-Obscure or questionable usage symbols
4. Is a repertoire including character names provided? *yes*
   a. If YES, are the names in accordance with the "character naming guidelines"
      in Annex L of P&P document? *yes*
   b. Are the character shapes attached in a legible form suitable for review *yes*
5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for
   publishing the standard? *– tbd –*
   If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools
   used: *– tbd –*
6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources)
      of proposed characters attached? *yes*
7. Special encoding issues:
   Does the proposal address other aspects of character data processing (if applicable) such as input,
   presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? *yes*

8. Additional Information:
Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see http://www.unicode.org/Public/UNIDATA/UCD.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? *no*
   If YES explain
2. Has contact been made to members of the user community (for example: National Body
   user groups of the script or characters, other experts, etc.)? *yes*
   If YES, with whom? *Peter Honnen, Landschaftsverband Rheinland*
   If YES, available relevant documents: *n/a*
3. Information on the user community for the proposed characters (for example:
   size, demographics, information technology use, or publishing use) is included? *yes*
   Reference: *see proposal text*
4. The context of use for the proposed characters (type of use; common or rare
   Reference: *see proposal text*
5. Are the proposed characters in current use by the user community? *yes*
   If YES, where? Reference: *see proposal text*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely
   in the BMP? *yes*
   If YES, is a rationale provided?
   If YES, reference: *keeping them together with all other diacritics for the Latin script*
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? *n/a*
8. Can any of the proposed characters be considered a presentation form of an existing
   character or character sequence? no
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
9. Can any of the proposed characters be encoded using a composed character sequence of either
   existing characters or other proposed characters? *no*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
10. Can any of the proposed character(s) be considered to be similar (in appearance or function)
    to an existing character? *no*
    If YES, is a rationale for its inclusion provided?
    If YES, reference:
11. Does the proposal include use of combining characters and/or use of composite sequences? *yes*
    If YES, is a rationale for such use provided? *yes*
    If YES, reference: *see proposal text*
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? *yes*
    If YES, reference: *see proposal text*
12. Does the proposal contain characters with any special properties such as
    control function or similar semantics? *no*
    If YES, describe in detail (include attachment if necessary)

13. Does the proposal contain any Ideographic compatibility character(s)? *no*
    If YES, is the equivalent corresponding unified ideographic character(s) identified
    If YES, reference: