

I. Administrative

1) Title

Combining diacritics for medievalistics

2) Requester's name

- Deutsches Institut für Normung (DIN)
- Monumenta Germaniae Historica. Constitutiones et acta publica imperatorum et regum (MGH)
- Berlin-Brandenburgische Akademie der Wissenschaften (BBAW)
- Arbeitsstelle »Deutsche Texte des Mittelalters« an der BBAW
- Arbeitsstelle »Mittelhochdeutsches Wörterbuch« und Projekt »Mittelhochdeutsche Wörterbücher auf CD-ROM und im Internet« der Universität Trier in Verbindung mit der Akademie der Wissenschaften und der Literatur Mainz
- Zentrum für Datenverarbeitung der Universität Tübingen, Abteilung Literarische und Dokumentarische Datenverarbeitung (LDDV)

3) Requester type

- DIN: Member body
- Remainder: Academic institutions

4) Submission date

February 2nd, 2000

5) Requester's reference

see Annex C

6) Kind of proposal

This is a complete proposal

II. Technical – General

1) Issue 1

The proposal is for addition of character(s) to an existing block.
Name of the existing block: Combining diacritics

2) Number of characters in proposal

52

3) Proposed category

Combining diacritics

4) Proposed level of implementation

Level 3

Rationale: The proposal contains combining diacritics

5) Repertoire including character names

Repertoire and character names are provided and are in accordance with Annex K of ISO/IEC 10646-1

Character shapes are attached

6) Font

TTF-font to be provided by

Marc Wilhelm Küster

Zentrum für Datenverarbeitung der Universität Tübingen

Abteilung Literarische und Dokumentarische Datenverarbeitung

Wächterstr. 76

72074 Tübingen

marc.kuester@zdv.uni-tuebingen.de

on the basis of the »MGH Constitutiones« font of the BBAW

This address serves also as point of contact for all questions which concern this proposal.

7) Special encoding issues

Currently used input methods used for these combining diacritics are explained in annex B

III. Technical – Justification

1) Has this proposal been submitted before?

No

2) Has contact been made to members of the user community?

Yes. The proposal was initiated by relevant institutions in the user community (cf. above) and has their full support. Furthermore, it reflects the current implementation in TUSTEP, a software widely used in long-term projects in the user community.

Reviews are attached.

3) Information on the user community

- Medieval studies (primarily for Middle High German)
- Long running projects such as the MGH

User community estimated at 25.000 in Germany and several thousand specialists in other countries.

Courses in Middle High German are obligatory for students of the Germanic languages in most German universities. The number of beginners in the study of the Germanic languages is estimated at three to four thousand students per year.

4) Character in current use?

Yes. They are regularly used in ongoing projects in the field (cf. Annex C)

5) Character to be encoded in the BMP?

Preferably yes, as these diacritics are currently in use by a sizable user community.

6) Should the proposed characters be kept together in a contiguous range?

Preferably yes, as these characters form an organic whole.

7) Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No. While they are similar in appearance to the basic Latin letters a to z, they are used as diacritics proper.

8) Can any of the proposed character(s) be considered to be similar in appearance or function to an existing character?

Cf. above

9) Does the proposal include use of combining characters and/or use of composite sequences?

The proposal includes the use of combining diacritics. They are essential to the task at hand, as it must be possible to combine any of these diacritics with any (basic) Latin letter.

10) Does the proposal contain characters with any special properties such as control function or similar semantics?

No

IV. Annex A: Repertoire

Glyph Code-point Name

^a o	0x0	COMBINING LATIN SMALL LETTER A ABOVE
^b o	0x0	COMBINING LATIN SMALL LETTER B ABOVE
^c o	0x0	COMBINING LATIN SMALL LETTER C ABOVE
^d o	0x0	COMBINING LATIN SMALL LETTER D ABOVE
^e o	0x0	COMBINING LATIN SMALL LETTER E ABOVE
^f o	0x0	COMBINING LATIN SMALL LETTER F ABOVE
^g o	0x0	COMBINING LATIN SMALL LETTER G ABOVE

0x0	COMBINING LATIN SMALL LETTER H ABOVE
0x0	COMBINING LATIN SMALL LETTER I ABOVE
0x0	COMBINING LATIN SMALL LETTER J ABOVE
0x0	COMBINING LATIN SMALL LETTER K ABOVE
0x0	COMBINING LATIN SMALL LETTER L ABOVE
0x0	COMBINING LATIN SMALL LETTER M ABOVE
0x0	COMBINING LATIN SMALL LETTER N ABOVE
0x0	COMBINING LATIN SMALL LETTER O ABOVE
0x0	COMBINING LATIN SMALL LETTER P ABOVE
0x0	COMBINING LATIN SMALL LETTER Q ABOVE
0x0	COMBINING LATIN SMALL LETTER R ABOVE
0x0	COMBINING LATIN SMALL LETTER S ABOVE
0x0	COMBINING LATIN SMALL LETTER T ABOVE
0x0	COMBINING LATIN SMALL LETTER U ABOVE
0x0	COMBINING LATIN SMALL LETTER V ABOVE
0x0	COMBINING LATIN SMALL LETTER W ABOVE
0x0	COMBINING LATIN SMALL LETTER X ABOVE
0x0	COMBINING LATIN SMALL LETTER Y ABOVE
0x0	COMBINING LATIN SMALL LETTER Z ABOVE
0x0	COMBINING LATIN SMALL LETTER A BELOW
0x0	COMBINING LATIN SMALL LETTER B BELOW
0x0	COMBINING LATIN SMALL LETTER C BELOW
0x0	COMBINING LATIN SMALL LETTER D BELOW
0x0	COMBINING LATIN SMALL LETTER E BELOW
0x0	COMBINING LATIN SMALL LETTER F BELOW
0x0	COMBINING LATIN SMALL LETTER G BELOW
0x0	COMBINING LATIN SMALL LETTER H BELOW
0x0	COMBINING LATIN SMALL LETTER I BELOW
0x0	COMBINING LATIN SMALL LETTER J BELOW
0x0	COMBINING LATIN SMALL LETTER K BELOW
0x0	COMBINING LATIN SMALL LETTER L BELOW
0x0	COMBINING LATIN SMALL LETTER M BELOW
0x0	COMBINING LATIN SMALL LETTER N BELOW
0x0	COMBINING LATIN SMALL LETTER O BELOW
0x0	COMBINING LATIN SMALL LETTER P BELOW
0x0	COMBINING LATIN SMALL LETTER Q BELOW
0x0	COMBINING LATIN SMALL LETTER R BELOW
0x0	COMBINING LATIN SMALL LETTER S BELOW
0x0	COMBINING LATIN SMALL LETTER T BELOW
0x0	COMBINING LATIN SMALL LETTER U BELOW
0x0	COMBINING LATIN SMALL LETTER V BELOW
0x0	COMBINING LATIN SMALL LETTER W BELOW
0x0	COMBINING LATIN SMALL LETTER X BELOW
0x0	COMBINING LATIN SMALL LETTER Y BELOW
0x0	COMBINING LATIN SMALL LETTER Z BELOW

V. Annex B: Input method

TUSTEP, a software much in use by long-term projects in the field, has been supporting the proposed combining diacritics for many years. They are keyed in as #;xy (resulting in $\overset{x}{y}$) and #!xy (resulting in $\underset{x}{y}$) respectively (where x and y can be arbitrary letters). The input is then rendered appropriately.

VI. Annex C: References and supporting documentation

Please refer to the following pages of supporting documentation.

BERLIN-BRANDENBURGISCHE AKADEMIE DER WISSENSCHAFTEN

- *Präsident* -

POSTFACH 238, 10106 BERLIN
TEL.: +49/ (0)30/ 20 370 645
FAX: +49/ (0)30/ 20 370 622
Berlin, den 16.09.99

BBAW * Postfach 238 * 10106 Berlin

Universität Tübingen
Zentrum für Datenverarbeitung
Marc Wilhelm Küster
Wächterstraße 76
D-72074 Tübingen

Dear Mr. Küster,

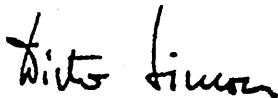
the Berlin-Brandenburg Academy of Sciences and Humanities (formerly Prussian Academy of Sciences and Humanities) is overseeing about 30 cultural, historical and scientific long-term Academy projects.

In these Academy projects looking back on a long scientific tradition the problem of the „special characters“ often is very important. It is a matter of concern of the Berlin-Brandenburg Academy to modernize the methods of working of these so-called Academy projects and thereby to improve on the time to deal with these projects.

With regard to their publications - working aim of nearly all Academy projects - scientists are very often confronted with the limits and considerable hindrances caused by the character set „ASCII“ traditionnally used. Therefore, the Berlin-Brandenburg Academy of Sciences and Humanities and its management are firmly supporting the application for including special characters used in medieval and early modern German language in the caracter set „UNICODE“.

We would very much appreciate if you could realize this technical standard important to our cultural, historical and scientific projects as soon as possible.

Sincerely yours,



Dieter Simon

**BERLIN-BRANDENBURGISCHE
AKADEMIE DER WISSENSCHAFTEN**

MONUMENTA GERMANIAE HISTORICA
PROF. DR. ECKHARD MÜLLER-MERTENS

10117 Berlin
Jägerstr. 22/23
Telefon 030/20370239
Telefax 030/20370600

Postanschrift: BBAW, Monumenta Germaniae Historica, Postfach 238, 10106 Berlin

Herrn
M. W. Küster
Zentrum für Datenverarbeitung
der Universität Tübingen
Wächterstr. 76
72074 Tübingen

Berlin, den 18. September 1999

Sehr geehrter Herr Küster,

anbei in kürzester Form der Bedarf der MGH Constitutiones an einem Zeichensatz „Combining Diacritics“ und in einem zweiten Absatz ein Ergänzungsvorschlag. Wir haben beides in hoffentlich nicht allzu fehlerhaftem Englisch formuliert:

According to the editorial rules of the MGH Constitutiones series any letter combination, which is found in the charters to be edited (at the moment of emperors Louis IV, 1314-1347, and Charles IV, 1346-1378), has to be reproduced as carefully and exactly as possible. Documents written in early modern German are especially concerned by that principle. The MGH Constitutiones team has bought a special font „MGH Constitutiones“ which contains the most common letter combinations, but we are fully aware of the fact that new ones may appear at any time. Therefore free combining of diacritics is highly desirable for our purposes, if in electronic publishing we want to maintain the standards used in our printed editions.

The proposed category „combining diacritics“ however should not only offer the possibility of combining letters, but also of combining letters and diacritic signs as dashes, points and accents (see the hints in item 3 of Prof. Gärtner' s expertise). Thus not only editors of medieval German texts, but editors of any text printed in Latin letters (e. g. classicists and

medieval latinists) could be enabled to constitute critical apparatus that show abbreviations, additions or corrections in the way they appear in manuscripts. Because of that we further suggest and recommend to get into contact with medievalists, classicists and experts in editorial technique - as well scientific (i. e. „Editionswissenschaften“, for example Prof. J. Roloff, FU Berlin) as applied in publishing houses (e. g. B. G. Teubner, Stuttgart, or Brepols Publishers, Turnhout) or even at the Monumenta Germaniae Historica in Munich, where Dr W. Setz does most of the layout of MGH editions). Some of these persons and institutions might also be able to supply more combinations of letters and/or diacritics.

Wir hoffen, daß „Combining Diacritics“ auf dem Weg in den Unicode auch die internationalen Hürden recht bald mit Erfolg zu nehmen vermag.

Mit freundlichen Grüßen

A handwritten signature in black ink, appearing to read 'Eckhard Müller-Mertens', written in a cursive style.

(Prof. Dr. Eckhard Müller-Mertens)

Berlin-Brandenburgische Akademie der Wissenschaften
Deutsche Texte des Mittelalters

Professor Dr. Kurt Gärtner

Universität Trier
FB II: Sprach- und
Literaturwissenschaften
D-54286Trier
☎ 0651/201-2323
Fax: 0651/201-3909
e-mail: gaertner@uni-trier.de

Herrn
Marc Wilhelm Küster
Computing Centre of the University of Tübingen
Dept. Literary and Linguistic Computing
Waechterstraße 76
D-72074 Tübingen

Dear Mr. Küster,

in support of our proposal to ISO/ISE JTC1/SC2/WG2 I summarize the main points of my expertise of 22.03.1999 for the Berlin-Brandenburg Academy of Sciences and Humanities. I support the application in my capacity as project director of the Academy's large-scale editorial series *Deutsche Texte des Mittelalters* (German Text of the Middle Ages) which was founded in 1904 and has over the years produced close to 90 editions of medieval German texts. The project also comprises the biggest collection of descriptions of medieval German manuscripts. For the editorial series and for the manuscript descriptions we are planning digitized versions of all those items which are of importance to scholars all over the world. I support the application also in my capacity as director of various research projects carried out at the University of Trier in connection with the Academy of Sciences and Literatures in Mainz (1. New Middle High German Dictionary; 2. *Middle High German Dictionaries Interlinked* and *Deutsches Wörterbuch* der Brüder Grimm, 1st Edition, for the Internet and on CD-ROM).

After having been working in the field of Humanities Computing for nearly three decades I from the beginning encountered a major problem for the processing of medieval vernaculars: the considerable variety of combined characters in medieval texts was difficult to deal with adequately. Especially the placing of most of the Latin characters and various diacritics (acute, grave, circumflex, abbreviations for *er*, *re*, *ra*, *ri*, *ro*, nasal strokes etc.) above and below another character or above and to the right resp. below and to the right of any character was essential for reproducing an exact transcription of a medieval text or manuscript.

The use of letters as diacritics in medieval German manuscripts is very important from a linguistic and philological aspect: the diacritics were used by medieval scribes to indicate the umlaut or other vowel changes which could not be represented by the restricted inventory of the Latin alphabet. Diacritics were also used in certain areas in a specific manner that helps considerably to localize a medieval vernacular manuscript. In most cases letters used as diacritics do not represent mere graphic variants of the letter above which they are placed, but are distinctive from the letter without diacritics; in sorting alphabets the letters with diacritics usually have to be treated as subletters to the main letter i. e. the letter without diacritic.

A general systematic solution for the problems mentioned has been found by the developers of TUSTEP; there is –as far as I know – no other software package at hand which offers a similar systematic representation of graphic variants which the medievalist has to deal with in all stages of his work. Therefore I support the application in its present form wholeheartedly.

Yours sincerely,

A handwritten signature in black ink, appearing to read 'K. Gärtner', written in a cursive style.

Dr. Kurt Gärtner

Professor of German (University of Trier, Germany)

Project Director *Deutsche Texte des Mittelalters* (Berlin-Brandenburg Academy of Sciences and Humanities, Berlin)

dort verre zuzin gahen
 ubir velt, zuhant sprachen sie:
 9200 'secht der troumer kumit hie.
 kumit den slahin, und besen
 waz im gutis sul geschen,
 sint im so wol getroumit si.
 und stet ein cisterne bi
 9205 di sichte und ane wazzir ist,
 dar in werfe wir in an dirre vrist
 und senden heim sin gewant,
 tun unsern vatur irkant:
 ein tier habe in irbizzen,
 9210 vrezzen und zurizzen;
 so wirt danne wol schin,
 waz in helfen di troume sin
 und wilchir selicheite gewin.'
 do sprach der eldist under in,
 9215 Ruben, der e uch ist genant,
 der wolde in losen von ir hant:
 [73^{7a}] 'wir virgizen nicht sin blut;
 werfin in lebende – daz ist gut –
 in dise durre alde cisternen,
 9220 daz unsir vatur nicht lernen
 sul· nach im iemerliche sterbin
 und immir in clagende virterbin.'
 Joseph zu sinen brudern quam,⁶⁰
 di im i nidic und gram
 9225 in irn herzen lange waren.
 daz begonden si im offenbaren:

9198 Do K. zv in *KM2*. 9200 nu kvmet der trömer *PL*. 9201 kumit *G* (venite *Gn* 37,20)]
 Wol dan *KM2P*. suln wir *P*; sla wir vñ *L*. 9204 Vnz *Ny*, Hie *KM2P*. 9205 trucken *M2W*.
 9206 werfen in *KP*, wir in (*f. W*) werffen *M2W*. 9207 brengen (+ hin *PL*) heim *KM2PL*.
 9208 Vnd tun *NyKM2*. daz erkant (be- *M2P*) *KM2P*. 9209 In habe ein tier erbizzen (ge- *M2*,
 zerrizzen *PL*) *KM2WPL*. 9210 Gefrezzen *M2*, Gezzen *PL*. vñ ouch *KM2*. er(ge- *L*)bizzen
PL. 9211 wol] an im *KM2P*. 9212 Waz heiles im gebēt *PL*. im *KM2*. 9215 Ruben| *G*. den
 ich han e *KM2P*. ew e *Ny*. 9216 Er *KM2*. 9217 Lat stē nicht v̄gizzet (vergiezzet[-en *PW*])
 niht *M2WP*) *KM2WP*. 9218 Werfet *KM2*. 9219 In die *KM2*. alde *f. KM2P*. 9220 icht
KM2P. lerne *KP*, m̄g gelern (lern *W*) *M2W*. 9221 In iamere nach im sterben *KM2P*. 9222
 immir in clagende *G*, clegelichen *KM2*, in clagender clage *PL*. in clagen *Ny*. 9224 Daz sie im
 nidic *KM2P*. 9225 le in *M2P*. lange *f. KM2P*. 9226 Daz zeigete (+ im *M2*, + in *WP*) ir
 gebaren *KM2WPH*, Daz ge zeigete sie in ir gebare *L*.

⁶⁰ 9223–30.33–36.42–56 *Gn* 37,23–27; 9231f. *Hist. schol.*, *Gn Cap.* 87 (*PL* 198,1126A); 9237–41
Hist. schol., *Gn Cap.* 87 (*PL* 198,1126A).