

Universal Multiple-Octet Coded Character Set
Universeller mehrfach-8-bit-kodierter Zeichensatz
International Standards Organisation
Internationale Standardisierungsorganisation
Organisation Internationale de Normalisation
Международная организация по стандардизации

Doc. type: Working group document
Titel: Proposal to encode Latin Capital Letter Sharp S to the UCS
Source: Deutsches Institut für Normung (DIN),
Technical Committee NIA-01-29-01 Coded Character Sets
Status: Official ISO NB Member Request
Date: 2007-03-21

This proposal requests the encoding of the Latin uppercase letter Sharp S in ISO/IEC 10646.

The German Institute for Standardization (DIN) has noted the existence of this character in German print publications and has identified a need for encoding the Latin uppercase letter Sharp S in ISO/IEC 10646. We regard the incorporation of the character in ISO/IEC 10646 as necessary.

Uppercase Sharp S is to be seen as the typographical equivalent to LATIN SMALL LETTER SHARP S (U+00DF).

ß LATIN CAPITAL LETTER SHARP S

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹

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 Latin Capital Letter Sharp S to the UCS**
2. Requester's name: *Cord Wischhöfer, on behalf of DIN-Committee on Codes Character Sets (NIA-01-29-01)*
3. Requester type (Member body/Liaison/Individual contribution): *German ISO Member Body*
4. Submission date: *2007-03-21*
5. Requester's reference (if applicable): *E-Mail: cord.wischhoefer@din.de . Fon: +49 (0)30 2601 2535*
6. Choose one of the following:
- This is a complete proposal: Yes
- (or) More information will be provided later:

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: *Latin Extended Additional, Latin general use extension*
2. Number of characters in proposal:
3. Proposed category (select one from below - see section 2.2 of P&P document):
- | | | | |
|--|--|---|--|
| A-Contemporary <input checked="" type="checkbox"/> | B.1-Specialized (small collection) <input type="checkbox"/> | B.2-Specialized (large collection) <input type="checkbox"/> | |
| C-Major extinct <input type="checkbox"/> | D-Attested extinct <input type="checkbox"/> | E-Minor extinct <input type="checkbox"/> | |
| F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/> | G-Obscure or questionable usage symbols <input type="checkbox"/> | | |
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? *Andreas Stötzner (as@signographie.de)*
- If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: *Contact as@signographie.de when required*
6. References:
- a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? Yes
- 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
- (see attached)*

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.

¹ Form number: N3102-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	<i>Yes</i> <i>A proposal for "Latin Capital Letter Double S" was forwarded to the UTC by Andreas Stötzner in 2004</i>
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.)? If YES, with whom? If YES, available relevant documents:	<i>Yes</i> <i>National Body, National Government</i> <i>See attached document</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	<i>Yes</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>common</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>Yes</i> <i>See attached</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	<i>Yes</i> <i>No</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<i>No</i>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>No</i>
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>No</i>
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>No</i>
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:	<i>No</i>
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	<i>No</i>
13. Does the proposal contain any Ideographic compatibility character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference:	<i>No</i>

Rationale

The character capital sharp S has been existing in printed matters for 128 years. Thus it is a part of recent German script history. Its encoding is required in order to match original text lettering in present editions of old publications. In recent lettering praxis there is a need to have the uppercase sharp S available for some special tasks like branding, shop-signing and book titles.

DIN has thoroughly considered possible consequences of a capital letter sharp S for existing data stock and data processing which might be caused by the introduction of the new character. The encoding of Capital Sharp S does *not* interfere with the stability of case folding policy as outlined in doc. RESOLUTION M47.1, STABILITY OF CASE FOLDING, DOC. N 2954. Capital Sharp S should be encoded without case folding to ß (U+00DF). No corruption of existing data is to be expected, user expectations and orthographical customs *will not be touched*. See the following chapter for a detailed discussion of these aspects of our proposal.

The frequently used German letter ß (sharp s or “Eszett”) is regularly replaced by “SS” in using uppercase lettering. Yet, if a word has been written in uppercase letters (e.g. WEISSE), one can not anymore detect if its proper original spelling is *Weiß*e or *Weisse*. This being a sensitive matter particularly with regard to personal names, many users see the need for a capital sharp S to avoid ambiguity and confusion. This applies mainly to the fields of name writing, company branding and title design. Because a real capital letter has not been available in coded character set so far, in many instances the corruption of an uppercase line of characters by a lowercase ß is found. This is obviously an undesirable makeshift solution, a problem to be solved by the encoding of the new character.

Without a sensible solution the present situation constitutes an anomaly German script culture and an even impediment to its further development.

Discussion of technical aspects

(Citations are in cursive writing)

Michael Everson [message to as@signographie.de, 23. 10. 2006]:

*Because I believe that when it is available, people will want Weiß*e to go to WEIßE [WEIßE].

We do not follow that assumption. From our point of view it is important to make clear that the proposition is not aimed at a reformation of the German orthography. The official capitalisation of “ß” remains “SS”. Furthermore it seems imperative to make clear that all existing data stock that uses “SS” instead of “ß” in German names *will remain valid*.

We need the capital ß exclusively for the following special purposes:

- to retain the distinction between “SS” and “ß” in capital letters in those cases where the owner of the name wants it (especially in proper names);
- to represent text correctly where the capital ß has already been used (historic text or title-, product- or company-names, logos etc.).

This does neither exclude the introduction of a capital ß on the basis of a future reform of orthography of the German language with making the “SS” capitalisation an ancient (old-fashioned) variant, nor the usage of the then available capital ß in future orthographic propositions for higher or lower German dialects. It is not before such a point in time that we need consider the introduction of new locales etc.

We propose the following:

The capital ß is considered a special representation of the sequence of letters “SS” (which *can* be applied when “SS” is corresponding to a “ß”), which is an individual character as opposed to a ligature.

It should be discussed if the compatibility decomposition to “U+0053 U+0053” should be assigned to the new Unicode character. That would provide for the equivalence of the character sequences “capital ß” and “SS” in those applications that use the Normalization Form KD or KC for the detection of sameness of names etc. (see also below regarding the problem of the usage in domain-names (IDN). It seems not sufficient to solve the IDN-problem with compatibility decomposition.)

The capital ß may under no circumstances change the case mappings of existing data stock:

– toupper (“ß”) remains “SS”;

– tolower (“capital ß”) results in “ß”: there is no rule that toupper (tolower(X)) must give X again, and existing data stock is not affected as there is no capital ß yet.

Furthermore the “case foldings” of existing data stock may not be changed:

According to the rules of Unicode 5.0 p.188 (5.18 Case Mappings, Caseless Matching) the following entry in CaseFolding.txt (<http://www.unicode.org/Public/UNIDATA/Casefolding.txt>) results from the above-mentioned toupper- and tolower-values:

```
XXX;F;0073 0073;# LATIN CAPITAL LETTER SHARP S
```

Unicode-“Full case folding (mappings that cause strings to grow in length):” (see CaseFolding.txt): results in string “ss” for each “capital ß” with the above-mentioned toupper- and tolower values as well as for each lower letter ß.

Application to existing data stock leave data unconditionally unchanged. Unicode-“simplecase folding, mappings to single characters” (see CaseFolding.txt):

A small letter ß is produced for any capital ß that has the above-mentioned tolower-value.

Application to existing data stock leave data unconditionally unchanged.

MICROSOFT-case-folding (no official standard, but mentioned by Asmus Freytag [message to as@signographie.de, 4. 5. 2006]): ... does not apply e.g. to the case folding used in the WINDOWS operating system. Therein a case folding results in an uppercase letter for reasons of compatibility with DOS. In that case e.g. WINDOWS could not support the equivalence of a normal ß and a capital ß, because that would mean that all ß would have to be assigned to the uppercase ß.

As not only the OS but also many applications make use of that kind of case folding, that would also affect the interfaces of the programs. Compatibility prohibits off-hand changes therein.

As in UnicodeData.txt for the small letter “ß” (U+00DF) for the property Simple_Uppercase_Mapping an empty value is registered, the “single capital letter” corresponding to the “ß” is again the small-letter “ß”. Because the existing property is not changed by the method described here (which under no circumstances can be altered with regard to the current stability policies) everything remains the same. Therefore the application to existing data stock leave data unconditionally unchanged in this case, too.

The results are *anyway incorrect* according to German orthography rules, which makes the result in a similar manner acceptable with the existence of a capital ß.

Unlike the Unicode-full-case-folding Microsoft-case-folding distinguishes between “ss” and “ß”. One can create files “Maße.txt” and “MASSE.txt” in the same Windows directory (otherwise upper- and lower-case letters are not distinguished in Microsoft-Windows-Filesystems).

For the capital ß, a “curious” consequence of the demand that the outcome of a Microsoft-case-folding for upper- and lowercase letters with same resulting-string-lengths should give the same result is that it is mapped to a small letter “ß” (whereas all other letters become uppercase letters). This can be accepted, because this kind of folding is relevant for technical purposes only, as e.g. comparison of filenames.

A result similar to the Microsoft-case-folding can be found in many SQL-implementations for the function UPPER, i.e. for UPPER (‘Maße’) the string “MAßE” is returned. According to standard ISO 9075 this is wrong in any case; the result has to comply with the Unicode-full-case-folding. For correct SQL-implementations surely no problems will come up with the introduction of the capital ß, as the ISO-compliant application of UPPER on “ß” still produces “SS”. Future extensions of the SQL-standard may provide mechanisms for mapping an “ß” on a capital ß, but this needs not to be discussed here and now.

It is important to clarify the use of a capital ß in domain-names.

Michael Everson claims – justly in our opinion – unambiguously:

But the UTC has rules (which have to do with IDN specifications) and this MUST be taken into account. This is NOT just a matter of typography. It affects processing, and must be done with all the costs and benefits in mind. [message to as@signographie.de, 23. 10. 2006]

As is the case with all pairs of uppercase-/lowercase pairs, capital ß and lowercase ß must be equivalent. Following the IDN-rules, the lowercase ß is equivalent to the string “ss”. That implies that a capital ß in a domain-name is equivalent to the string “ss”.

This cannot be concluded directly from the formerly proposed “compatibility decomposition” for a capital ß (IDN-mechanism, <http://tools.ietf.org/html/rfc3490>), because the mechanism “nameprep” first performs the substitution by table (and there are only characters from Unicode 3.2) and after that applies Normalization Form KC (in reverse sequence the capital ß would have been mapped to “SS” in the first step and to “ss” in the second step).

This context requires an update of RFC3454 <http://tools.ietf.org/html/rfc3454>; without it the usage of a capital ß would be inadmissible. This RFC has to be adapted to the Unicode-version containing the capital ß, and there in Table B.2 an entry has to be supplemented, as we want to point out (albeit the changes in IDN are carried out by another institution and are not automatically adjusted after the introduction of a capital ß in Unicode):

xxxx;0073 0073; Case map
(xxxx=Hex-Code of the capital ß in Unicode)

On the other hand it is advantageous that the capital ß is inadmissible in domain names: it means no trouble from that side.

Proposed codepoint: 1E9C

Block: Latin Extended Additional, Latin general use extension

	Latin Extended Additional															
	1E00	1E01	1E02	1E03	1E04	1E05	1E06	1E07	1E08	1E09	1EA0	1EB0	1EC0	1ED0	1EE0	1EF0
0	À	Ā	Ġ	Ķ	Ĭ	Ò	Š	Ŧ	Ẁ	Ẃ	Ȧ	Ǻ	È	Õ	Ǿ	Ư
1	à	ā	ġ	ķ	ĭ	ò	š	ŧ	ẁ	ẃ	ȧ	ǻ	è	õ	ǿ	ư
2	Ḃ	Ḅ	Ḥ	Ḷ	Ḝ	Ḑ	Ḓ	Ḙ	Ẃ	Ẅ	Ȧ	Ǻ	Ë	Ö	Ơ	Ỡ
3	ḃ	ḅ	ḥ	ḷ	ḝ	ḑ	ḓ	ḙ	ẃ	ẅ	ȧ	ǻ	ë	ö	ơ	ỳ
4	Ḭ	È	Ḧ	Ḹ	Ḛ	Ḕ	Ḗ	Ḙ	Ẅ	Ẇ	Ȧ	Ǻ	Ë	Ö	Ư	Ỡ
5	ḭ	è	ḧ	ḹ	ḛ	ḕ	ḗ	ḙ	ẅ	ẇ	ȧ	ǻ	ë	ö	ư	ỳ
6	Ḭ	È	Ḧ	Ḹ	Ḛ	Ḕ	Ḗ	Ḙ	Ẅ	Ẇ	Ȧ	Ǻ	Ë	Ö	Ư	Ỡ
7	ḭ	é	ḡ	Ḻ	Ḟ	Ḕ	Ḗ	Ḙ	ẅ	ẇ	ȧ	ǻ	ê	õ	ủ	ý
8	Ç	Ē	Ḥ	Ḹ	Ḓ	Ḕ	Ḗ	Ḙ	Ẅ	Ẇ	Ȧ	Ǻ	Ë	Ö	Ư	Ỡ
9	ç	ē	ḥ	ḻ	ḟ	ḕ	ḗ	ḙ	ẅ	ẇ	ȧ	ǻ	ë	ö	ư	ỳ
A	Ḭ	È	Ḧ	Ḹ	Ḛ	Ḕ	Ḗ	Ḙ	Ẅ	Ẇ	Ȧ	Ǻ	Ë	Ö	Ư	
B	ḭ	è	ḧ	ḹ	ḛ	ḕ	ḗ	ḙ	ẅ	ẇ	ȧ	ǻ	ê	õ	ử	
C	Ḭ	È	Ḧ	Ḹ	Ḛ	Ḕ	Ḗ	Ḙ	Ẅ	Ẇ	Ȧ	Ǻ	Ë	Ö	Ư	
D	ḭ	è	ḧ	ḹ	ḛ	ḕ	ḗ	ḙ	ẅ	ẇ		ȧ	ǻ	ê	õ	ử
E	Ḭ	È	Ḧ	Ḹ	Ḛ	Ḕ	Ḗ	Ḙ	Ẅ	Ẇ		ȧ	ǻ	ê	õ	ử
F	ḭ	è	ḧ	ḹ	ḛ	ḕ	ḗ	ḙ	ẅ	ẇ		ȧ	ǻ	ê	õ	ử

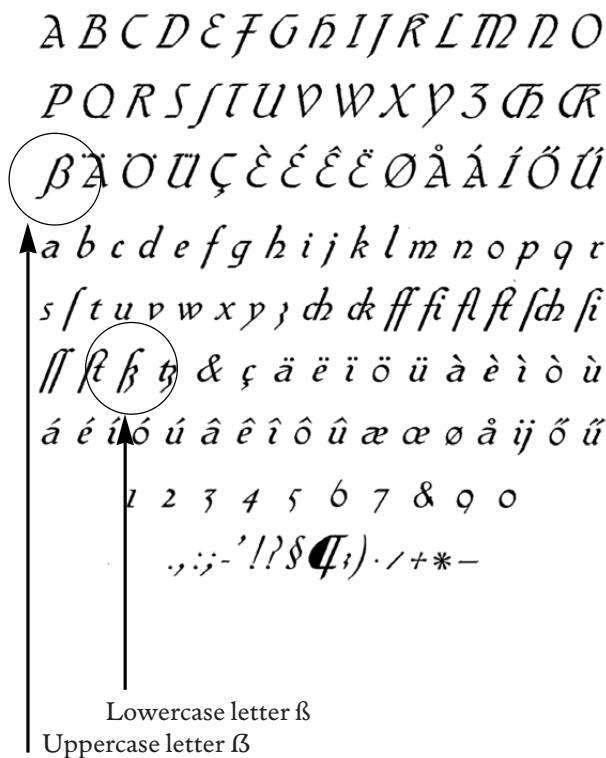
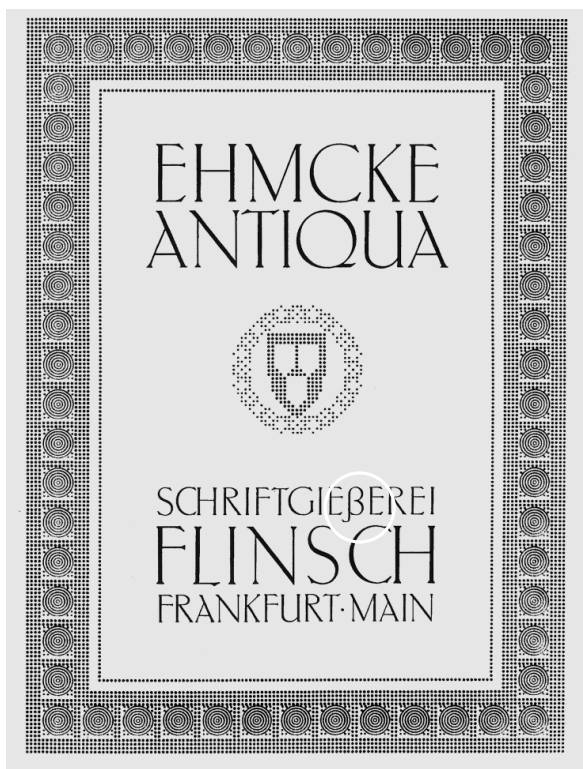
References

A number of publications clearly show that the uppercase ß was used in printing. It should be remembered that the actual letter shapes happen to be of quite variable quality, they even might easily be mistaken for lowercase ß .

Note that in all cases which are documented here the intended capitality of the letter is verified.

PREUBEN — REUß — STRABBURG — GIEßEN
 — GROßHERZOG — WEIß — MAßMANN —
 FRIDRICHSTRABE — GROßBEERENSTRABE.

1. *Regeln für die deutsche Schreibung*. Ed. Verein für deutsche Rechtschreibung. Berlin 1879, p. 18;
 The very first capital ß in German typographic history.



2. EHMCKE, F. H.: *Ehmcke-Antiqua*. Type specimen of foundry Schriftgießerei Flinsch. Frankfurt (Main) 1909

3. EHMCKE, F. H.: *Ehmcke-Antiqua Kursiv*. Type specimen of foundry Schriftgießerei Flinsch. Frankfurt (Main) 1910

z ä ö ü ß ch ck fi fi fl ssi fi ß
CDEFGHIJKLMNOPQ

abcdefghijklmnopqrstu
vwxyz ä ö ü ch ck .,:;?!') & ß
ABCDEFGHIJKLMNO

xyzäöüßchckfffi fl ssi fi
BCDEFGHIJKLMNO
RSTUVWXYZ Ä Ö Ü
.,:;?!') & ß 123456789

á é í ó ú â ê î ô û ch ck f f f i f l s s i f i ß ç
KLMNOPQRSSTUVW
ò ù á é í ó ú â ê î ô û ch ck f f f i f l s s i f i ß
IJKLMNOPQRSSTUV

ä ö ü ß ch ck f f f i f l s s i f i ß
3CDEFGHIJKLMNOPQ

ß h ch ck ß .,:;?!')
ABCDEFGHIJKL

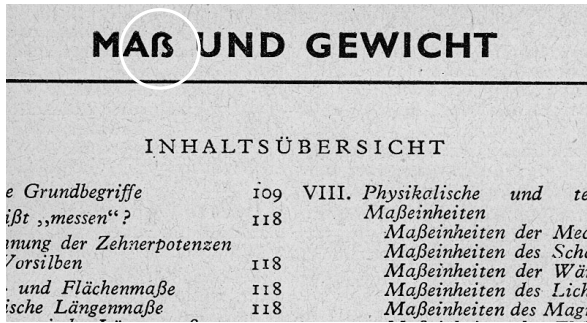
ô û ç ch ck f f f i f l s s i f i ß ç
OPQRSSTUVWXYZ

4. Variuos lead typefaces (1913–1928), each contain-
ing either minuscule and majuscule characters for ß
(source: Signa 9/2006, p. 52).

◇ Lowercase ○ Uppercase



5. From various print samples, around 1915; source: Deutsche Bücherei Leipzig



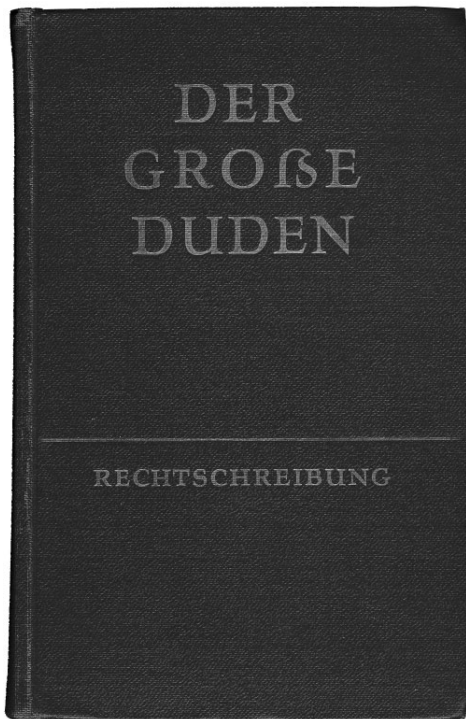
6. *Schlag nach / Natur*. Bibliographisches Institut, Leipzig 1956 (left, bottom left)



7. *Schlag nach / Natur*. Bibliographisches Institut, Leipzig 1963 (bottom)



8. *Kleine Enzyklopädie / Natur*. Bibliographisches Institut, Leipzig 1963; p. 139, 157



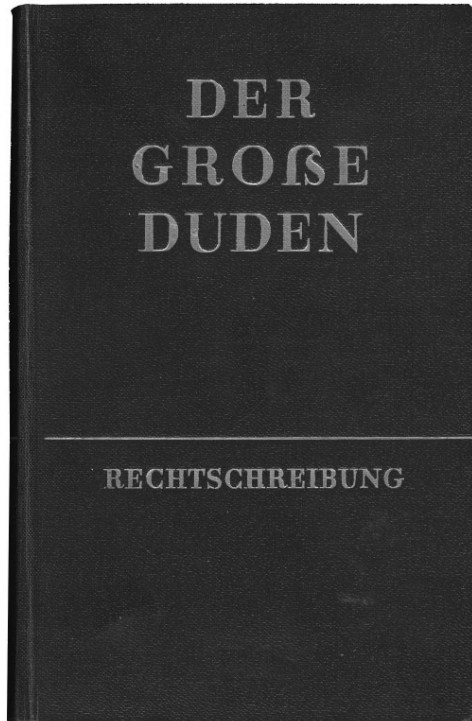
DER GROßE DUDEN

Wörterbuch und Leitfaden
der deutschen Rechtschreibung

15. Auflage

Herausgegeben von Horst Klien

9. *Der Große Duden*. 15th edition, Bibliographisches Institut, Leipzig 1957 (cover and main title)



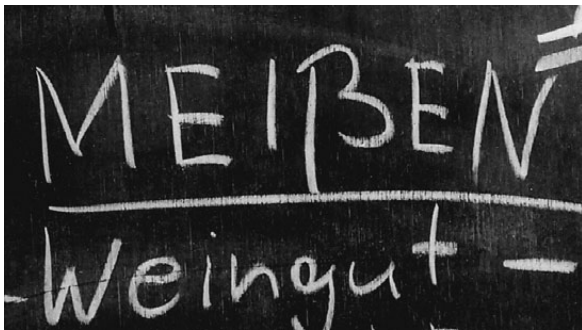
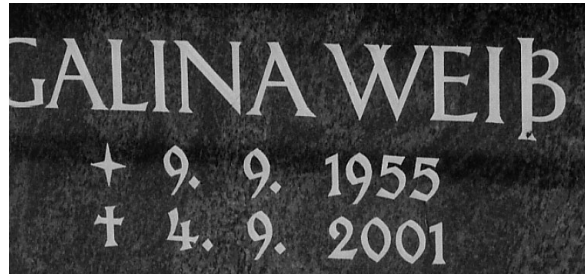
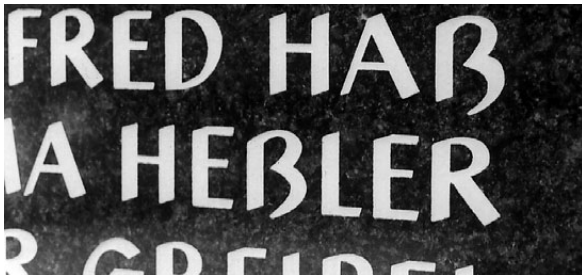
DER GROßE DUDEN

Wörterbuch und Leitfaden
der deutschen Rechtschreibung

15. Auflage

Herausgegeben von Horst Klien

10. *Der Große Duden*. 15th edition, Bibliographisches Institut, Leipzig 1960, 1964 (cover and main title)



11. Those engraved and handwritten samples show how the capital sharp S is part of typographic culture. (Photography: A. Stötzner, I. Preuß)



GROßE SOßE
GROßE SOßE

wrong (lowercase ß)

right (uppercase ß)

12. Capital sharp S is needed for typesetting of commercial advertisements, packaging and titling.



Bundesministerium
des Innern

DIN NI

1.6. OKT. 2006

POSTANSCHRIFT Bundesministerium des Innern, 11014 Berlin

DIN Deutsches Institut für Normung e.V.
10772 Berlin

HAUSANSCHRIFT Alt-Moabit 101 D, 10559 Berlin

POSTANSCHRIFT 11014 Berlin

TEL +49 (0)1888 681-3197

FAX +49 (0)1888 681-53197

BEARBEITET VON RD Süßmuth

E-MAIL

INTERNET

DATUM Bonn, 26. September 2006

AZ IT 4 - 127763/2

BETREFF **Normung des versalen "ß"**

BEZUG Ihr Schreiben vom 21.09.2006 -whf / chr -

Sehr geehrte Damen und Herren,

seitens des Bundesministeriums des Innern bestehen gegen die Normung eines versalen ß (Großbuchstabe ß) keine Bedenken.

Mit freundlichen Grüßen

Im Auftrag

Reisen

Proposal to encode Latin Capital Letter Sharp S to the UCS

CONTENTS

Title	1
Proposal Summary Form	2
Rationale	4
Discussion of technical aspects	4
Proposed codepoint	7
References	8
Letter from Bundesministerium des Inneren	13

CREDITS

Contributors to this document are:

Paul Dettmer, Dr. Hans-Werner Eisermann, Karl Pentzlin, Andreas Stötzner, Cord Wischhöfer