

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Международная организация по стандартизации

Doc Type: Working Group Document
Title: Additional Latin characters for the UCS
Source: Michael Everson
Status: Working Document
Date: 1998-05-25

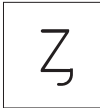
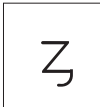




This document proposes several Latin characters for inclusion in the UCS, and discusses the rationale for their inclusion.

Two of these characters, LATIN CAPITAL LETTER KRA and LATIN CAPITAL LETTER YR, are capital forms of characters already in the UCS and derive from ISO/TC46/SC4 standard ISO 5426-2:1996.

Two of the characters, LATIN CAPITAL LETTER Z WITH HOOK and LATIN SMALL LETTER Z WITH HOOK, are used in Middle High German texts and scholarly apparatus.

Two of the characters, LATIN CAPITAL LETTER OU and LATIN SMALL LETTER OU, were used in the Wendat language and are currently used to write the Algonquin language in Canada.

The characters proposed to be added here are given below, with proposed code positions. The proposal summary form is appended thereafter.

0220	LATIN CAPITAL LETTER KRA	0222:	
0221	LATIN CAPITAL LETTER YR		
0222	LATIN CAPITAL LETTER Z WITH HOOK	0223:	
0223	LATIN SMALL LETTER Z WITH HOOK		
0222	LATIN CAPITAL LETTER OU	0224:	
0223	LATIN SMALL LETTER OU		
0220:		0225:	
0221:			

A. Administrative

1. Title

Additional Latin characters for the UCS

2. Requester's name

Michael Everson, EGT (WG2 member for Ireland)

3. Requester type

Expert contribution

4. Submission date

1998-05-25

5. Requester's reference

6a. Completion

This is a complete proposal.

6b. More information to be provided?

No

B. Technical -- General

1a. New script? Name?

No.

1b. Addition of characters to existing block? Name?

Yes. Latin Extended-A

2. Number of characters

6

3. Proposed category

Category A

4. Proposed level of implementation and rationale

Level 1 noncombining characters.

5a. Character names included in proposal?

Yes

5b. Character names in accordance with guidelines?

Yes

5c. Character shapes reviewable?

Yes

6a. Who will provide computerized font?

Michael Everson, Everson Gunn Teoranta

6b. Font currently available?

Yes

6c. Font format?

TrueType

7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?

Yes. ¶ ISO 5426-2:1996. ¶ Karl Weinhold, Gustav Ehrismann, & Hugo Moser, *Kleine mittelhochdeutsche Grammatik* (Stuttgart: Wilhelm Braumüller, 1972). ¶ Albert Bachmann, *Mittelhochdeutsches Lesebuch* (Zürich: Beer & Cie., 1970). ¶ Jacques Maurais, ed. *Les langues autochtones du Québec*. (Québec: Publications du Québec, 1992)

See also <http://www.indigo.ie/egt/standards/iso10646/pdf/iso-5426-2.pdf>

7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?

Hardcopy is provided to WG2 for distribution. See also pictures given below.

8. Does the proposal address other aspects of character data processing?

No

C. Technical -- Justification

1. Contact with the user community?

Yes

2. Information on the user community?

Linguists, librarians, publishers. ISO/TC46/SC4, U.S. Library of Congress.

3a. The context of use for the proposed characters?

KRA is used in older Greenlandic orthography. YR is used in linguistic descriptions and texts dealing with primitive Norse. Z WITH HOOK is used in linguistic descriptions and texts dealing with Old and Middle High German. OU was used to write the Wendat language and is currently used to write the Algonquin language. COmpatibility with ISO 5426-2:1996.

3b. Reference

See above

4a. Proposed characters in current use?

Yes

4b. Where?

KRA and YR are in ISO 5426-2:1996. Z WITH HOOK is used in academic contexts. OU is used commonly in Algonquin words; two photographs showing their use on public signage in Canada are provided.

5a. Characters should be encoded entirely in BMP?

Yes.

5b. Rationale

Contemporary use and keeping them with the other Latin characters.

6. Should characters be kept in a continuous range?

No, but the casing pairs should be kept together.

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

No. KRA could be decomposed to either K' or K'.

7b. Where?

7c. Reference

8a. Can any of the characters be considered to be similar (in appearance or function) to an existing character?

No. Glyph for OU can look like a digit 8 (with no break at the top). Some implementations currently use the digit 8 to represent this letter, but this is a feature of the paucity of existing character sets.

8b. Where?

8c. Reference

9a. Combining characters or use of composite sequences included?

No

9b. List of composite sequences and their corresponding glyph images provided?

10. Characters with any special properties such as control function, etc. included?

No

D. SC2/WG2 Administrative

To be completed by SC2/WG2

1. Relevant SC 2/WG 2 document numbers:

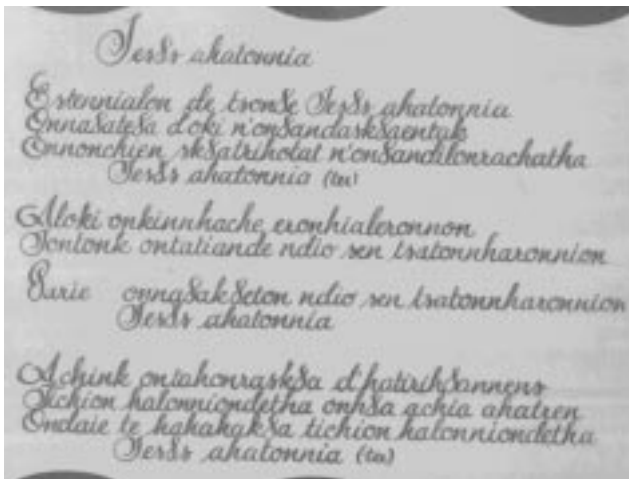
2. Status (list of meeting number and corresponding action or disposition)

3. Additional contact to user communities, liaison organizations etc.

4. Assigned category and assigned priority/time frame

Other Comments

School in Québec showing the name “ÉCOLE” and “H8TAIE” in capital letters in French and Algonquin.



Beginning of the canticle *Le Noël huron*, composed in 1648 by Fr. Jean de Brébeuf. At the beginning of the eighth line CAPITAL LETTER OU is shown; SMALL LETTER OU is shown throughout.



Roadsign in Quebec in French and Algonquin, showing LATIN SMALL LETTER OU.