Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Working Group Document
Title: Proposal to encode a symbol "Capitalized Commercial At" in the UCS
Author: Karl Pentzlin
Status: Individual Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2012-04-17

Block: Miscellaneous Symbols and Arrows

Special orthographic element

Ⓐ   U+2Bxx   CAPITALIZED COMMERCIAL AT

→ 0040 commercial at
· used as a letter for the Koalib language (spoken in Sudan)

Properties:
```
2Bxx;CAPITALIZED COMMERCIAL AT;So;0;ON;;;;;N;;;;;
```

1. Introduction

In 2004, a Latin letter pair "LATIN CAPITAL/SMALL LETTER AT" was proposed in L2/04-349, due to its contemporary use in the Koalib language spoken in Sudan (Fig. 2 shows p.5 of that proposal, giving explanations and examples).
The encoding of such letters was rejected for security reasons. If there is a letter (the small at) which can be used to spoof the ubiquitous @ in e-mail addresses and other types of URLs, this is a security risk which by far outweighs the advances of a small user group having their special letters in a full-fledged way.

This proposal intends to serve that user group in another way.
It is not uncommon for languages to use non-letter characters (including true punctuation marks) as letters, without having the possibility to recur to letters which have the same appearance but true letter properties. A well-known example is Tlingit (of the Pacific coast area of Canada), which uses the full stop (U+002E) as letter (denoting the glottal stop).

Following this example, it imposes no special issues to advise the Koalib community to use as their small letter the existing U+0040 COMMERCIAL AT, which is a punctuation mark according to its Unicode properties.
Thus, they lack only the capital complement of it. As they have to deal with a character anyway which is not formally a letter, this complement may also a non-letter (to comply to the security issues). Therefore, a symbol as proposed here fits their needs.

Having the proposed character available, there are still some drawbacks left for the community:
They cannot use both letters in any identifier (including domain names or e-mail addresses), which they are used to anyway as they cannot do this for their small letter even now.
Also, they have no standard automatic case conversion.
However, these drawbacks are negligible compared with the real drawback not having a character representing their capital letter at all.

1.1 Security issue caused by the lack of a Capitalized Commercial At

A special issue also to be considered is the fact that the non-acceptance of L2/04-349 inspired several font suppliers which want to serve as many linguistic communities as possible to include the letters in the PUA (see e.g. fig. 1). Introducing a "small letter commercial at" by this way to communities using such fonts on a regular base, also raises the security considerations which were the original reason to not accept a "small letter commercial at".

Encoding the capitalized form now would unequivocally demonstrate these font suppliers that the needs of the Koalib community are in fact fulfilled by the Unicode repertoire, thus they no longer need to include a hazardous letter in the PUA.

2. Encoding Considerations

The proposed name contains the term "CAPITALIZED" rather than "CAPITAL", to emphasize the fact that the character is not a full-fledged letter.

It is proposed in a BMP block rather than the block 1F100...1F1FF Enclosed Alphanumeric Supplement, as it is aimed to an audience in a region where limited access to contemporary technical resources may prevail, which gives BMP characters an advantage over SMP ones in terms of user accessibility.

Also, the characters in that block have compatibility equivalences to the base letters they show as enclosed or modified forms, while such an equivalence is not given for the character proposed here due to security considerations.
Also, the character proposed here is primarily considered as a modified form of a symbol, rather than as an enclosed form of the capital letter A.

As the proposed character is an orthographic element, it may be regarded as a punctuation mark rather than a symbol. If it is decided such, it has to get the general category Po rather than So, and it has to be placed then in the block U+2E00...U+2E7F Supplemental Punctuation. Doing so would keep its general category in line with U+0040 COMMERCIAL AT which is a punctuation mark also by its general category Po.

3. Security Considerations

As no compatibility decomposition is given, the character cannot be used in any identifier (including domain names and e-mail addresses). Therefore, it cannot be used for any spoofing.

4. Examples and Figures

Fig. 1: Letters as proposed in L2/04-349 in the description of the PUA characters included in the Quivira font, retrieved 2012-04-17 from http://www.quivira-font.com/files/QuiviraPUA.pdf

Private Use Area: Additional Latin Supplement (0F000 – 0F00F)

4 characters since version 1.7

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0F00		ß	ʒ		Ⓐ	@										

0F001	ß	Latin Small Ligature Long S S	0F004	Ⓐ	Latin Capital Letter At
		≈ 0017F ſ 00073 s	0F005	@	Latin Small Letter At
0F002	ʒ	Latin Small Ligature Long S Z			• Koalib (spoken in Sudan)
		≈ 0017F ſ 0007A z			

Fig. 2: Page 5 from the original proposal L2/04-349, showing an explanation and examples.
The references used on this page are (according of p.6 of the original proposal):
Frontline Fellowship. 1993. Katikisim . (Koalib Catechism). Nairobi, Kenya: Sudan Literature Centre.
Sudanese Church of Christ. 1998. Rilŋa Rəthi Allah . (Hymnal in the Koalib language of Sudan).
Nairobi, Kenya: Sudan Literature Centre.

LATIN CAPITAL LETTER AT and LATIN SMALL LETTER AT are used as orthographic characters for in the Koalib language of Sudan. It is typically used for Arabic loan words. The Koalib Catechism and Hymnal were both printed with initial print runs of 1,000 and 1,500 respectively. They have been reprinted a number of times. Store records in the year 2000 showed that they had printed over 2,500 each at that time. The New Testament in Koalib also uses LATIN CAPITAL LETTER AT and LATIN SMALL LETTER AT. Although similar in appearance to COMMERCIAL AT, LATIN SMALL LETTER AT should have different character properties. The main concern is the similarity in appearance of LATIN SMALL LETTER AT to COMMERCIAL AT. There are potential implications for Internet protocols that use @.

@braaniiyyiin 4:13.

Kwomne kwiti kweere mac, kwəti luccuni kiyənə yəthi Allah; laakin nɔŋwəni kwurwənnəlɔ ndendet, nɔŋw naanalɔ por-por kiyənə yəthi ŋgwa kwindir ethi rilli kiyənə yuuŋwun, ŋgwa kwindi ethi iisacci nyuŋwsi ŋothɽor ŋəri.

Figure 24. LATIN CAPITAL LETTER AT (Frontline Fellowship, 1993, p. 1).

2   Allah wir kwimɽu
Kerreny nyiiŋə lurtunəgənə
Kirem nyiiŋə limaamigna
@alshaan shunu
Nyiiŋə liṭimica rom-rom.

Figure 25. LATIN CAPITAL LETTER AT (Sudanese Church of Christ, 1998, p. 141).

Ŋiɽaŋal Ŋəthi Kiṭaab Kirllinəlɔ Ter.

Isha@ya 45:5

Nyii kwir Kweeleny. Nə Allah ere naani wir ter kwokwony mac, illi ninyɔrɔ nyii tɔɽɔk.

Figure 26. LATIN SMALL LETTER AT (Frontline Fellowship, 1993, p. 3).

Ŋiɽaŋal ŋisaaw ŋəthi faathi

Ethi ne@mə kwəthi Kweeleny kwəri kwəni Yəcu
Kwɔrɔstɔ,
Nə ŋimɽi ŋəthi Allah tok,
Nə ŋɔmaṭ ŋəthi Tigɽim Tirllinelɔ ter tok,
Ethi nanni dəŋgəri nə dok-dok.
Imba.

Figure 27. LATIN SMALL LETTER AT (Sudanese Church of Christ, 1998, p. 143).

<table>
<tr><td colspan="2" align="center">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 <i>Roadmaps</i>.</td></tr>
</table>

A. Administrative

1. **Title:**	*Proposal to encode a symbol "Capitalized Commercial At" in the UCS*
2. Requester's name:	*Karl Pentzlin*
3. Requester type (Member body/Liaison/Individual contribution):	*Individual contribution*
4. Submission date:	*2012-04-17*
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:	

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:	*Miscellaneous Symbols and Arrows*
2. Number of characters in proposal:	*1*

3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary X	B.1-Specialized (small collection)	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. Fonts related:	
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	
	The author, on request
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	
	The author, on request
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)?	*No*

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: N3702-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, 2008-05, 2009-11)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? *Similarly*
   If YES explain *see text (L2/04-349 contained a letter similar to the symbol proposed here)*
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? *see L2/04-349*
   If YES, available relevant documents: *see L2/04-349*
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 L2/04-349*
4. The context of use for the proposed characters (type of use; common or rare) *Rare*
   Reference: *see L2/04-349*
5. Are the proposed characters in current use by the user community? *Yes*
   If YES, where? Reference: *see L2/04-349*
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? *Yes*
   If YES, reference: *see text*
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? *No*
    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:
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: