

ORGANISATION INTERNATIONALE DE NORMALISATION
Международная организация по стандартизации
INTERNATIONAL ORGANIZATION FOR STANDARDIZATION

ISO/CEI JTC 1/SC 2/GT 2
ISO/IEC JTC 1/SC 2/ WG 2

Jeu universel de caractères codés sur plusieurs octets (JUC)

Universal Multiple-Octet Coded Character Set (UCS)

ISO/IEC JTC 1/SC 2/WG 2 N2738
2004-01-19

Title: Proposal to add a Roman Numeral "Canopy" Character
Source: Canada (prepared by P. Andries – pandries@videotron.ca)
Action: For WG2 consideration and adoption in the next Amendment to the Standard
Distribution: ISO/IEC JTC 1/SC 2/WG 2 and Liaison Organizations

Titre: Proposition d'ajouter un caractère « dais » numéral romain
Source: Canada (préparé par P. Andries – pandries@videotron.ca)
Action: Pour considération par le GT2 et adoption dans le prochain amendement de la Norme
Distribution: ISO/CEI JTC 1/SC 2/GT 2 et les organismes de liaison

ISO/CEI JTC 1/SC 2/GT 2

**FORMULAIRE RÉSUMÉ DEVANT ACCOMPAGNER TOUTE PROPOSITION D'ADDITION
AU RÉPERTOIRE DE L'ISO/CEI 10646**

Veuillez compléter les sections A, B et C ci-dessous.

(Veuillez lire le document intitulé « Principles and Procedures Document for guidelines and details » avant de compléter ce formulaire.)

Voir <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> pour la dernière version de *Principles and Procedures*.

Voir <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> pour la dernière « feuille de route ».

A. Administratif

- | | | |
|--|---|--|
| 1. Titre: | Proposition d'ajout d'un caractère numéral romain « dais » | |
| 2. Nom du demandeur: | Canada | |
| 3. Type de demande (organisme membre/liaison/contribution individuelle): | Organisme membre | |
| 4. Date de la demande: | 2004-04-19 | |
| 5. Référence du demande (le cas échéant): | | |
| 6. (Choisir une des alternatives suivantes:) | | |
| Il s'agit d'une proposition complète: | Oui | |
| ou | Plus d'information sera fournie par la suite: | |

B. Technique — général

- | | | |
|---|------------------|--|
| 1. (Choisir une des alternatives suivantes:) | | |
| a. Ce document propose une nouvelle écriture (jeu de caractères): | | |
| Nom proposé de l'écriture: | | |
| b. Ce document propose l'addition d'un ou plusieurs caractères à un bloc existant: | | |
| Nom du bloc existant: | Formes numérales | |
| 2. Nombre de caractères proposés: | 1 | |
| 3. Catégorie suggérée (voir section II, <i>Character Categories</i>): | D | |
| 4. Niveau de mise en œuvre suggéré (1, 2 ou 3) (voir article clause 14, ISO/CEI 10646-1: 2000 (F)) : | 3 | |
| Une justification pour ce choix est-elle fournie ? | | |
| Si oui, renvoi: | Voir ci-dessous | |
| 5. Un répertoire comprenant des noms de caractères est-il fourni ? | Oui | |
| a. si OUI, les noms se conforment-ils aux <i>Conseils pour le choix des noms de caractère à l'Annexe L de l'ISO/CEI 10646-1: 2000 (F)</i> ? | Oui | |
| b. La forme des caractères est-elle jointe de manière lisible afin de permettre sa révision ? | Oui | |
| 6. Qui fournira la police informatique appropriée (ordre de préférence des formats : TrueType puis PostScript) pour la publication de la norme ? | Patrick Andries | |
| Si elle est d'ores et déjà disponible, préciser la ou les sources de la police (inclure l'adresse postale, de courriel, le site ftp,...), préciser également les outils utilisés : | | |
| __ http://pages.infinit.net/hapax/polices/dais.ttf | | |
| __FontLab 4.5 | | |
| 7. Références et renvois: | | |
| a. Des références (à d'autres jeux de caractères, des dictionnaires, des textes descriptifs, etc.) sont-elles fournies ? | Oui | |
| b. Des exemples publiés (comme des coupures de presse) des caractères proposés sont-ils fournis ? | Oui | |
| 8. Aspects particuliers du codage : | | |
| Cette proposition aborde-t-elle d'autres aspects du traitement informatique des caractères, tels que la saisie, la représentation, le tri, la recherche, l'indexation, la translittération... (si oui, veuillez joindre l'information) ? | Oui | |
| 9. Renseignements supplémentaires : | | |
| Il est suggéré aux demandeurs de fournir tout renseignement supplémentaire relatif aux propriétés des caractères proposés ou de l'écriture proposée qui permettront de bien les comprendre et de les traiter correctement. Parmi les propriétés : casse, renseignement numérique, renseignement monétaire, comportement au rendu comme la chasse ou la sécabilité des caractères, etc. Comportement combinatoire, directionnel, d'approches, valeur de tri par défaut, sa pertinence dans un contexte balisé, son équivalence de compatibilité et autres renseignements de normalisation Unicode. Voir le standard Unicode < http://www.unicode.org > pour des informations de ce type pour d'autres écritures. Voir également < http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html > et les rapports techniques Unicode correspondants qui stipulent l'information nécessaire pour que le comité technique Unicode considère l'inclusion de ces caractères dans le standard Unicode. | | |

C. Technique — Justification

1. Cette proposition d'addition de caractère(s) a-t-elle soumise auparavant ?	Non
Si OUI, expliquer	
2. Est-on rentré en contact avec des membres de la communauté des utilisateurs (par exemple, organisme national, utilisateurs de l'écriture ou de l'écriture, d'autres experts, etc.) ?	Non
Si OUI, avec qui ?	
Si OUI, documents pertinents disponibles :	
3. Des renseignements sur la communauté d'utilisateurs des caractères proposés (par exemple : profil démographique, utilisation en informatique, utilisation sous forme publiée) sont-ils fournis ?	
Référence:	
4. Fréquence d'utilisation des caractères proposés (habituel ou rare)	Rare
Référence:	
5. Les caractères proposés sont-ils actuellement utilisés par la communauté des utilisateurs ?	Oui
Si OUI, par qui? Référence :	Érudits
6. Après avoir dûment considéré les principes dans <i>Principles and Procedures</i> (un document permanent du GT 2), les caractères proposés doivent-ils tous être codés dans le PMB ?	Oui
Si OUI, une justification est-elle fournie ?	Oui
Si OUI, renvoi :	D. 5. Justification au codage dans le PMB
7. Les caractères proposés doivent-ils rester groupés dans un intervalle contigu (plutôt que dispersés) ?	S/O
8. Un des caractères proposés peut-il être considéré comme une forme de présentation d'un caractère ou d'une suite de caractères existants ?	Non
Si OUI, une raison pour son inclusion est-elle fournie ?	
Si OUI, renvoi :	
9. Un des caractères proposés peut-il se coder à l'aide d'une suite de caractères composés préexistants ou proposés?	
Non	
Si OUI, une raison pour son inclusion est-elle fournie ?	
Si OUI, renvoi :	
10. Peut-on considérer un des caractères proposés comme semblable (en apparence ou en fonction) à un caractère préexistant ?	Oui
Si OUI, une raison pour son inclusion est-elle fournie ?	
Si OUI, renvoi :	D. 7. Caractères similaires
11. Cette proposition comprend-elle des caractères combinatoires ou des suites composites (voir articles 4.4 et 4.34 dans ISO/CEI 10646-1:2000(F)) ?	Oui
Si OUI, une raison pour son inclusion est-elle fournie ?	
Si OUI, renvoi :	
Une liste des suites composites et des œils correspondants est-elle fournie ?	
Si OUI, renvoi :	
12. Cette proposition comprend-elle des caractères ayant une propriété particulière comme une fonction de commande ou une signification semblable ?	Non
Si OUI, décrire en détail (joindre pièce si nécessaire)	
13. Cette proposition comprend-elle un ou plusieurs caractères idéographiques de compatibilité ?	Non
Si OUI, le ou les caractères idéographiques unifiés correspondants sont-ils désignés ?	
Si OUI, renvoi :	

D. Proposition détaillée

D. 1. Le caractère « dais »

Les Romains utilisèrent plusieurs conventions pour exprimer les grands nombres. Une consistait à ajouter une ligne horizontale au-dessus des chiffres pour les multiplier par 1 000. Cette convention était quelque peu ambiguë puisqu'une convention préalable utilisait ce même trait suscrit pour différencier les valeurs numérale et alphabétique des symboles utilisés comme chiffres et lettres (IIIVIR = *duumvir*, IIIIVIR = *triumvir*).

Ceci explique sans doute pourquoi, lors du règne de l'empereur Hadrien (2^e siècle av. J.-C.), le facteur 1 000 fut parfois indiqué à l'aide de deux barres verticales et une barre horizontale suscrit qui entourait le nombre ainsi multiplié. Selon cette convention, DLVII représentait 557 000 ($557 \times 1\,000$). Pour des raisons de concision, nous appellerons ces barres englobantes un « dais ».

Toutefois, le dais représentait habituellement un facteur de 100 000. Un exemple typique se trouve sur les abaque romains où le signe [X] se trouve gravé devant la rangée des millions (voir l'illustration ci-dessous).

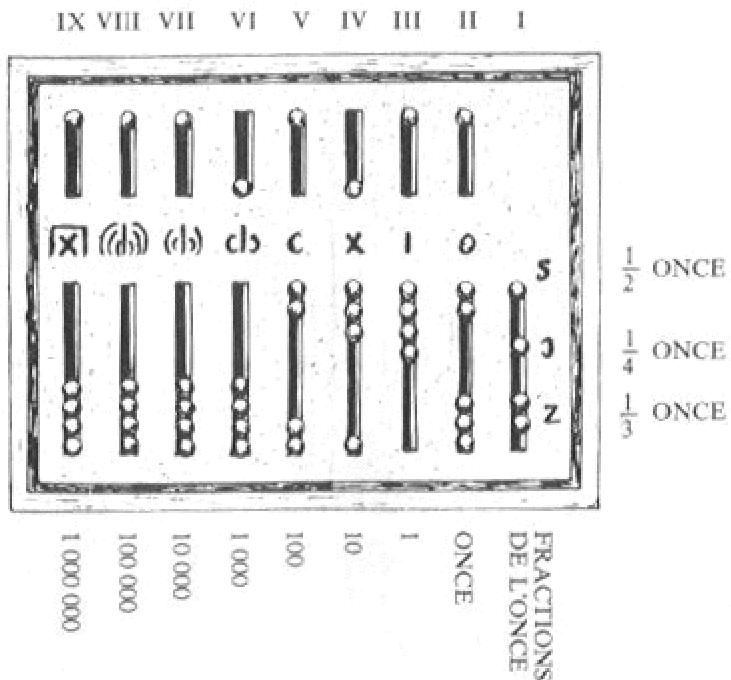


Fig. 16.95. – *Le principe de l'abaque romain portatif. Le dispositif représenté sur cette figure avait appartenu au jésuite allemand Athanase Kircher (1601-1680). Musée des Thermes de Rome.*

D. 2. Noms de caractère

Afin de représenter le dais (quelle que soit sa valeur), le présent document propose un nouveau caractère combinatoire ainsi que ses noms français et anglais.

U+21XX		DIACRITIQUE DAIS NUMÉRAL ROMAIN COMBINING ROMAN NUMERAL CANOPY
--------	--	---

Un DIACRITIQUE DAIS NUMÉRAL ROMAIN placé après un caractère numéral romain indique que ce numéral est multiplié par la valeur du dais. Au rendu, le caractère numéral devrait être encadré de deux jambages de part et d'autre et d'un toit horizontal (parfois facultatif), les trois traits se touchant et formant un dais. En présence d'une suite de caractères de base chacun suivi d'un DIACRITIQUE DAIS NUMÉRAL ROMAIN, les jambages intermédiaires devraient disparaître, le rendu consistant alors en un jambage au début de la suite, d'un autre jambage à la fin de la suite et d'un toit recouvrant la suite entière.

Exemples:

$$\begin{array}{ccc} \text{X}\square\text{XII} & \rightarrow & \overline{\text{X}}\text{XII} \\ \text{X}\square\text{I}\bar{\square}\text{I} & \rightarrow & \overline{\text{X}}\bar{\square}\text{I} \\ \text{D}\square\text{L}\square\text{V}\square\text{I}\square\text{I}\square\text{CCLXXIV} & \rightarrow & \overline{\text{DLVI}}\text{CCLXXIV} \end{array}$$

D. 3. Propriétés de caractères

Le DIACRITIQUE DAIS NUMÉRAL ROMAIN est de catégorie générale « Me ». L'entrée complète du caractère dais dans le fichier de données Unicode se lit :

21xx; COMBINING ROMAN NUMERAL CANOPY; Me;0;NSM;;;;N;;;;

D. 4. Valeur associée au dais

Voir ci-dessus — il n'existe pas de valeur unique intrinsèque associée au dais si ce n'est qu'il multiplie le chiffre précédent par un facteur donné (par exemple, mille ou cent mille selon le contexte).

D. 5. Justification au codage dans le PMB

Pour se retrouver dans le même bloc que les autres chiffres et symboles numéraux romains.

D. 7. Caractères similaires

U+0488	DIACRITIQUE CYRILLIQUE SIGNE CENTAINE DE MILLIERS COMBINING CYRILLIC HUNDRED THOUSANDS SIGN
U+0489	DIACRITIQUE CYRILLIQUE SIGNE MILLIONS

La principale différence entre le dais et ces deux signes combinatoires réside dans le fait que – lorsque plusieurs DIACRITIQUE DAIS NUMÉRAL ROMAIN se suivent et ne sont séparés que par une seule lettre de base – le dais ne ceint pas complètement ces lettres au rendu.

D. 8. Formes prises par le caractère dais

482

HISTOIRE UNIVERSELLE DES CHIFFRES

Exemples restitués :

M	D	C	L	I	L	X	V	I		C	C	X	V	I
1 651 × 100 000 + 78 × 1 000 + 316
165	178	316
→ 165 178 316														

Selon la source [2], récemment, le caractère dais se résume parfois en deux lignes verticales et omet la ligne horizontale suscrite.

51. The thousand-fold value of a number was indicated in some instances by a horizontal line placed above it. Thus, Aelius Lampridius (fourth century A.D.) says in one place, “CXX, equitum Persarum fudimus: et mox X in bello interemimus,” where the numbers designate 120,000 and 10,000. Strokes placed on top and also on the sides indicated hundred thousands; e.g., |X|CLXXXDC stood for 1,180,600. In more recent practice the strokes sometimes occur only on the sides, as in |X| · DC.XC., the date on the title-page of Sigüenza's *Libra astronomica*, published in the city of Mexico in 1690. In antiquity, to prevent fraudulent alterations, XXXM was written for 30,000, and later still C|O took the place of M.³ According to

Cette seconde forme → variante de glyphe (police)

E. Bibliographies

[1] GEORGES IFRAH, *Histoire universelle des chiffres*, Collection Bouquins, Éditions Robert Laffont,

Paris, 1994. ISBN 2-221-07838-1.

[2] FLORIAN CAJORI, *A History of Mathematical Notations*, La Salle (Illinois), The Open Court Publishing Company, 1928-1929 (2 volumes), réimprimé sous la forme d'un fac-similé en un volume, New York, Dover Publications, 1993, § 51 (vol. 1, p. 32).

[3] GILDERSLEEVE & LODGE, *Latin Grammar (1895)* réimprimé, Bolchazy-Carducci Publishers, Wauconda (Illinois) 1997, p. 52.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646

Please fill all the sections A, B and C below.

(Please read Principles and Procedures Document for guidelines and details before filling this form.)

See <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html> for latest Form.

See <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for latest Principles and Procedures document.

See <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest roadmaps.

A. Administrative

1. Title:	Proposal to add a Roman Numeral “Canopy” Character		
2. Requester's name:	Canada		
3. Requester type (Member body/Liaison/Individual contribution):	Member Body		
4. Submission date:	2004-04-03		
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):			
Proposed name of script:			
b. The proposal is for addition of character(s) to an existing block:			
Name of the existing block:	Number Forms		
2. Number of characters in proposal:	1		
3. Proposed category (see section II, Character Categories):	D		
4. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000 (E)):	3		
Is a rationale provided for the choice?			
If Yes, reference:	See below		
5. 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 ISO/IEC 10646-1: 2000 (E)?	Yes		
b. Are the character shapes attached in a legible form suitable for review?	Yes		
6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?	Patrick Andries		
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:			
__ http://pages.infinit.net/hapax/polices/dais.ttf			
__FontLab 4.5			
7. 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		
8. 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		
9. 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/UnicodeCharacterDatabase.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.)?	No
If YES, with whom?	
If YES, available relevant documents:	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	
Reference:	
4. The context of use for the proposed characters (type of use; common or rare)	rare
Reference:	
5. Are the proposed characters in current use by the user community?	Yes
If YES, where? Reference:	Scholars
6. After giving due considerations to the principles in <i>Principles and Procedures document</i> (a WG 2 standing document) must the proposed characters be entirely in the BMP?	Yes
If YES, is a rationale provided?	Yes
If YES, reference:	D. 5. Rationale to code character in BMP
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?	Yes
If YES, is a rationale for its inclusion provided?	Yes
If YES, reference:	D. 7. Similar characters
11. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000 (E))?	Yes
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:	

D. Detailed proposal

D. 1. The “Canopy” character

The Romans used several conventions to express large numbers. One was a horizontal line over them, which meant to multiply the number by 1,000. This convention was somewhat ambiguous as an older convention used the same overbar to distinguish the numeral from the phonetic values of symbols used as letters and numerals ($\overline{I}IVIR = duumvir$, $\overline{III}IVIR = triumvir$).

This may explain why, during Emperor Hadrian's reign (2nd century A.D.) the 1,000 factor was sometimes specified by surrounding the number multiplied with two vertical bars and one horizontal bar. Using this convention, $\overline{\overline{D}L\overline{VII}}$ represented 557,000 ($557 \times 1,000$). For brevity's sake, we shall call the surrounding bars a canopy.

However, the canopy was usually reserved to represent a 100,000 factor. A typical example is found on Roman abaci that had a million sign $\overline{\overline{X}}$ engraved in front of the million-slot (see figure below).

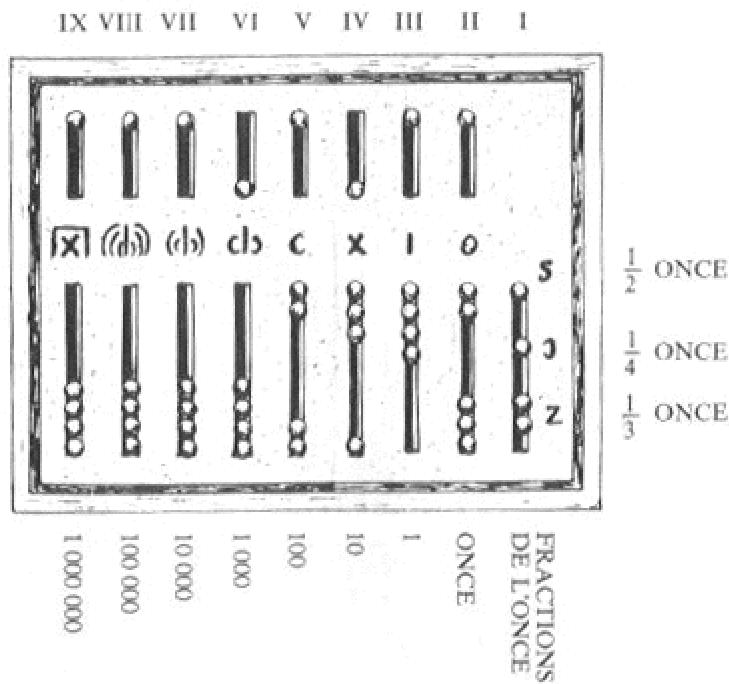


Fig. 16.95. – *Le principe de l'abaque romain portatif. Le dispositif représenté sur cette figure avait appartenu au jésuite allemand Athanase Kircher (1601-1680). Musée des Thermes de Rome.*

D. 2. Character names

In order to represent the canopy (whatever its value), a combining character is proposed below with its suggested English and French names.

U+21XX		DIACRITIQUE DAIS NUMÉRAL ROMAIN COMBINING ROMAN NUMERAL CANOPY
--------	---	---

A COMBINING ROMAN NUMERAL CANOPY placed after a Roman numeral character specifies that this numeral character be multiplied by the canopy value. When rendering, the numeral character shall be surrounded by two descenders and a (sometimes optional) horizontal overbar; these three strokes will touch each other and form a canopy. When several base characters are each followed by a COMBINING ROMAN NUMERAL CANOPY, the intermediate descenders shall not be rendered, the overall rendering being made out of an initial descender at the start of the sequence, a terminal descender at the end of the sequence and an overbar covering the whole sequence.

Examples:

X <small>□</small> XII	→	[X]XII
X <small>□</small> I <small>—</small> I	→	[X]I <small>—</small> I
D <small>□</small> L <small>□</small> V <small>□</small> I <small>□</small> I <small>□</small> CCLXXIV	→	[DLV]I <small>—</small> CCLXXIV

D. 3. Character properties

COMBINING ROMAN NUMERAL CANOPY has general category of Me. The suggested full entry of the canopy character in the Unicode Data file is:

21xx; COMBINING ROMAN NUMERAL CANOPY; Me;0;NSM;;;;N;;;;;

D. 4. Value associated to the canopy

See above — there is no single intrinsic value except that the canopy multiplies numeric values by a given factor (for instance, one thousand or one hundred thousand according to the context).

D. 5. Rationale to code character in BMP

To be found in the same block as other Roman numerals.

D. 7. Similar characters

U+0488	DIACRITIQUE CYRILLIQUE SIGNE CENTAINE DE MILLIERS COMBINING CYRILLIC HUNDRED THOUSANDS SIGN
U+0489	DIACRITIQUE CYRILLIQUE SIGNE MILLIONS COMBINING CYRILLIC MILLIONS SIGN

The main difference with these combining signs is that, when several successive COMBINING ROMAN NUMERAL CANOPY are found, the canopy does not completely surround the base characters when rendered.

D. 8. Alternatives shapes taken by the canopy character

482

HISTOIRE UNIVERSELLE DES CHIFFRES

Exemples restitués :

MDCLI	LXXVIII	CCCXVI
1 651 × 100 000	+ 78 × 1 000	+ 316
.....>		
165 178 316		

According to source [2], in recent practice the canopy character was sometimes represented by only opening and closing strokes.

51. The thousand-fold value of a number was indicated in some instances by a horizontal line placed above it. Thus, Aelius Lampridius (fourth century A.D.) says in one place, “CXX, equitum Persarum fudimus: et mox X in bello interemimus,” where the numbers designate 120,000 and 10,000. Strokes placed on top and also on the sides indicated hundred thousands; e.g., |X|CLXXXDC stood for 1,180,600. In more recent practice the strokes sometimes occur only on the sides, as in |X| · DC.XC., the date on the title-page of Sigüenza's *Libra astronomica*, published in the city of Mexico in 1690. In antiquity, to prevent fraudulent alterations, XXXM was written for 30,000, and later still CIO took the place of M.³ According to

E. References

[1] GEORGES IFRAH, *Histoire universelle des chiffres*, Collection Bouquins, Édition Robert Laffont, Paris, 1994. ISBN 2-221-07838-1.

[2] FLORIAN CAJORI, *A History of Mathematical Notations*, La Salle (Illinois), The Open Court Publishing Company, 1928-1929 (2 volumes), reprinted as a facsimile in one volume, New York, Dover Publications, 1993, § 51 (vol. 1, p. 32).

[3] GILDERSLEEVE & LODGE, *Latin Grammar* (1895) reprinted, Bolchazy-Carducci Publishers, Wauconda (Illinois) 1997, p. 52.