Arabic mathematical symbols in Unicode

M. BENATIA, A. LAZREK and K. SAMI

Cadi Ayyad University, Faculty of Sciences, Marrakech, Morocco
www.ucam.ac.ma/fssm/rydarab/english/unicode.htm
k_sami@ucam.ac.ma

27^{th} Internationalization and Unicode Conference

Summary

The Unicode Standard provides a quite complete set of standard mathematical characters. Arabic alphabet based scripts make use of local ways for writing mathematics. There are a good deal of symbols found in Arabic mathematical handbooks that are not yet part of the Unicode Standard and can't be obtained through a simple mirroring nor through a simple implementation process. The contribution presents a *discussion* of the usual mathematics character repertoire used in mathematics written in an *Arabic presentation*. Some of such special characters are *to be proposed for inclusion* into the Unicode Standard. The proposal also contains some non-mathematical characters such as units and arrows...

Arabic mathematical notation (1/2)

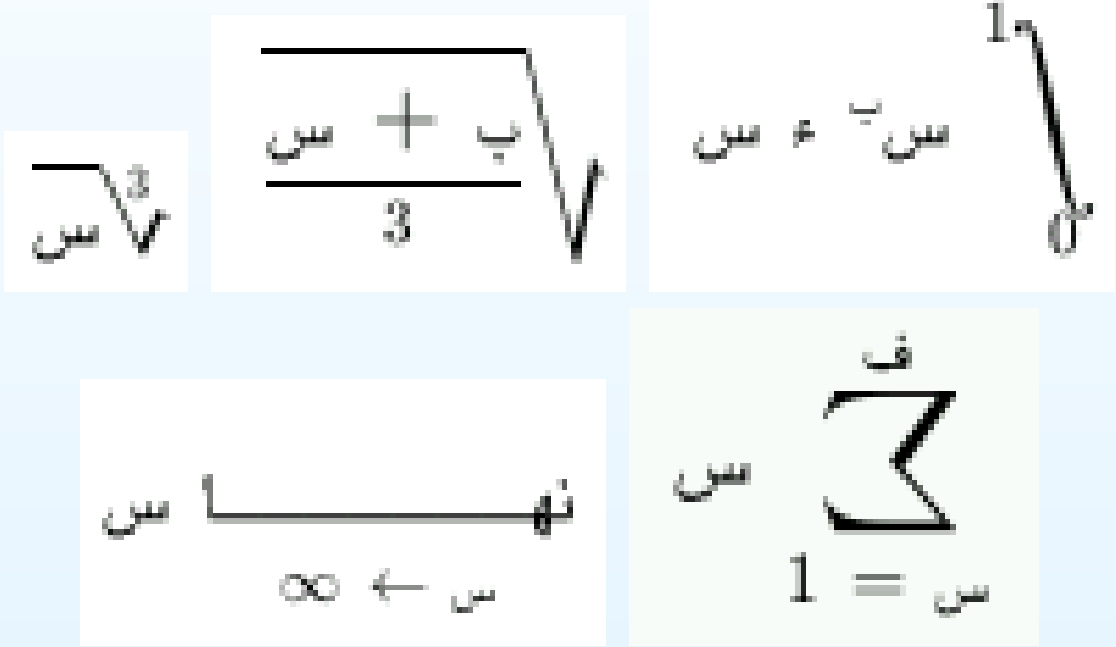
There are two main ways to denote mathematics in Arabic script based languages. In some contexts, mathematical texts use the usual mathematical symbols just as they are in Latin script based texts.

$$f(x) = \begin{cases} \sum_{i=1}^{s} x^i & \text{إذا كان} \quad x < 0 \\ \int_{1}^{s} x^i \, dx & \text{إذا كان} \quad x \in \mathrm{E} \\ \mathrm{tg}\ \pi & \text{غير ذلك} \ (\text{مع } \pi \simeq 3,141) \end{cases}$$

Mathematical expressions flow then from left to right against the stream of the natural language.

Arabic mathematical notation (2/2)

In other contexts, mathematical texts use specific symbols spreading out from right to left in accordance with the natural language writing.



Mathematical handbooks in Egypt, Syria, Libya, Iran, . . . adopt, generally, the second option.

Localized mathematical notations

Moreover, the use of a mathematical symbol can be localized. There can exist different local options to denote symbols. Both the symbols مجـ and ∑ can denote the operator *sum* and both the notations ٣.١٤١ and $3,141$ represent the same decimal number.

All the proposed symbols contained in this paper come from Arabic **handbooks of mathematics**. Most of these symbols had been adopted through official international conventions such as the *Amman's* 1987 *convention*.

As such usages seem to be widespread, these symbols may be proposed to be added to the Unicode Standard.

$$ت(س) = \begin{cases} \lim_{ب=١}^{ص} س^{ب} & \text{إذا كان } س > ٠ \\ \sum_{ب=١}^{ص} س^{ب} ئ & \text{إذا كان } س \ni م \\ ظا\ \pi & \text{غير ذلك (مع } \pi \backsimeq ٣,١٤١) \end{cases}$$

$$د(س) = \begin{cases} \lim_{ب=1}^{ص} س^{ب} & \text{إذا كان } س > 0 \\ \sum_{ب=1}^{ص} س^{ب} ئ & \text{إذا كان } س \ni م \\ tan\ \pi & \text{غير ذلك (مع } \pi \backsimeq 3{,}141) \end{cases}$$

$$f(x) = \begin{cases} \sum_{i=1}^{s} x^i & \text{if} \quad x < 0 \\ \int_{1}^{s} x^i \, dx & \text{if} \quad x \in \mathrm{S} \\ \tan\pi & \text{otherwise (with } \pi \simeq 3.141) \end{cases}$$

$$f(x) = \begin{cases} \sum_{i=1}^{s} x^i & \text{si} \quad x < 0 \\ \int_{1}^{s} x^i \, dx & \text{si} \quad x \in \mathrm{E} \\ \mathrm{tg}\ \pi & \text{sinon (avec } \pi \simeq 3,141) \end{cases}$$

$$(١-\sqrt[٥]{٧})^{\frac{١}{٤}} = x \text{ جا } \lim_{x \leftarrow \pi/١٠}$$

$$(1-\sqrt[5]{7})^{\frac{1}{4}} = x \text{ جا } \lim_{x \leftarrow \pi/10}$$

$$\lim_{x \to \pi/١٠} \sin x = \tfrac{١}{٢}(\sqrt{٥}-١)$$

$$\lim_{x \to \pi/10} \sin x = \frac{1}{4}(\sqrt{5}-1)$$

Mathematical alphabetical symbols

Mathematical notation uses a basic set of mathematical alphanumeric characters consisting basically on the sets:

- Arabic digits: 0 - 9;
- Arabic-Indic digits: ٠ - ٩;
- Arabic letters: ا - غ ;
- Some specific Arabic letters: ۿ, . . . ;
- uppercase and lowercase Latin letters: A - Z and a - z;
- uppercase and lowercase Greek letters: A - Ω and α - ω;
- plus other symbols.

Mathematical letter-like symbols

The alphabetic mathematical letter-like symbols used in Arabic mathematical handbooks are of six forms: *isolated*, *initial*, *tailed*, *stretched*, *looped*, and *double-struck*.

For Latin alphabet based Mathematical Alphanumeric Symbols, range 1D400-1D7FF, there are: Bold, Italic, Bold Italic, Script, Bold Script, Double-struck, Double Fracture, Sans-serif, Sans-serif bold, Sans-serif italic, Sans-serif bold italic and Monospace. For Greek, there are: Bold, Italic, Bold Italic, Sans-serif, Sans-serif bold, Sans-serif italic and Sans-serif bold italic.

ﺝ ﺟ MATHEMATICAL JEEM ISOLATED FORM

≈ <font> <isolated> 062C ﺝ Arabic letter jeem

ﺟ ﺠ MATHEMATICAL JEEM INITIAL FORM

≈ <font> <initial> 062C ﺟ Arabic letter jeem

≈ <font> FE9F ﺟ Arabic letter jeem initial form

ﺠ ﺤ MATHEMATICAL JEEM TAILED FORM

≈ <font> FE9F ﺟ 06C1 ﮧ

ﺟﺎ ﺤﺎ MATHEMATICAL JEEM STRETCHED FORM

≈ <font> FE9F ﺟ FE8E ﺎ

ﺝ ﺞ MATHEMATICAL JEEM LOOPED FORM

≈ <font> <isolated> 062C ﺝ Arabic letter jeem

ﺝ ﺟ MATHEMATICAL JEEM DOUBLE-STRUCK FORM

≈ <font> <isolated> 062C ﺝ Arabic letter jeem

Basic mathematical alphabetic symbols

In order to provide a big amount of mathematical symbols and to satisfy either the needs of those who use dot-less characters and those who prefer characters with dots, alphabetic mathematical symbols may be proposed in the two glyphs, with and without dots. So, the proposition will remain in the Unicode Standard spirit who recommends representing characters not glyphs. Actually, the symbols are presented with and without dots in the *a-b-j-d* order.

Basic mathematical alphabetic symbols

ISOLATED		INITIAL		TAIL		STRETCHED		
DOT	DOT-LESS	DOT	DOT-LESS	DOT	DOT-LESS	DOT	DOT-LESS	
ا								
ب	ب	بـ	ـد	بـه	ـده	بـا	ـدا	
ج	ح	جـ	ـح	جـه	ـحه	جـا	ـحا	
د	د							
ه		هـ	ـه	هـه	ـهه	هـا	ـها	
و	و							
ز	ر							
ح		حـ		حـه		حـا		
ط	ط			طـه	ـطه	طـا	ـطا	
ى	ى	ـيـ		ـيـه		ـيـا		

ISOLATED DOT	ISOLATED dot-less	INITIAL DOT	INITIAL dot-less	TAIL DOT	TAIL dot-less	STRETCHED DOT	STRETCHED dot-less	DOUBLE-STRUCK DOT	DOUBLE-STRUCK dot-less
ا								ا	
ب	ٮ	بـ	ٮـ	بــ	ٮــ	با	ٮا	ب	ٮ
ج	ح	جـ	حـ	جــ	حــ	جا	حا	ج	ح
د	د							د	د
ه		هـ	هـ	هــ	هــ	ها	ها	ه	
و	و							و	و
ز	ر							ز	ر
ح		حـ		حــ		حا		ح	
ط	ط			طــ	طــ	طا	طا	ط	ط
ي	ى	يـ		يــ		يا		ي	ى
ك		كـ	كـ	كــ	كــ	كا	كا	ك	
ل	ل	لـ	لـ	لــ	لــ			ل	ل
م	م	مـ	مـ	مــ	مــ	ما	ما	م	م
ن	ں	نـ		نــ		نا		ن	ں
س	س	سـ	سـ	ســ	ســ	سا	سا	س	س
ع	ع	عـ	عـ	عــ	عــ	عا	عا	ع	ع
ف	ڡ	فـ	ڡـ	فــ	ڡــ	فا	ڡا	ف	ڡ
ص	ص	صـ	صـ	صــ	صــ	صا	صا	ص	ص
ق	ٯ	قـ		قــ		قا		ق	ٯ
ر								ر	
ش		شـ		شــ		شا		ش	
ت		تـ		تــ		تا		ت	
ث		ثـ		ثــ		ثا		ث	
خ		خـ		خــ		خا		خ	
ذ								ذ	
ض		ضـ		ضــ		ضا		ض	
ظ				ظــ		ظا		ظ	
غ		غـ		غــ		غا		غ	

Table 4: Mathematical alphabetic symbols

Particular mathematical alphabetic symbols

Some glyphs used in Arabic mathematical presentation are not really Arabic letters but particular forms of mathematical alphabetic symbols used in Arabic mathematical handbooks

- The glyph of the letter ALEF ا can be taken for the Arabic-Indic digit ONE ١. Thus, it's replaced by ١.
- The glyph of the letter HEH ه can be taken for the Arabic-Indic digit FIVE ٥ in the isolated and double-struck forms. Thus, it is replaced by ھ.
- The glyph of the letter KAF ك results from the combination of two elements in the isolated and double-struck forms. Thus, it is replaced by either ك or ک.
- The glyph of the letter NOON can be found in various orientations and styles, with and without dot, (ن ں, ڼ ݨ, ں ن, ڼ ݨ, ﻦ ﻨ).

Exceptional mathematical alphabetic symbols

Some glyphs used in Arabic mathematical presentation are not really Arabic letters but symbols used in physics or in some Arabic alphabet based scripts such as Persian.

ﭒ MATHEMATICAL PEH ISOLATED FORM

≈ <font> <isolated> 067E ﭒ Arabic letter peh

ﭓ MATHEMATICAL PEH LOOPED FORM

≈ <font> <isolated> 067E ﭓ Arabic letter peh

ﭖ MATHEMATICAL TCHEH ISOLATED FORM

≈ <font> <isolated> 0686 ﭖ Arabic letter tcheh

ﭗ MATHEMATICAL TCHEH LOOPED FORM

≈ <font> <isolated> 0686 ﭗ Arabic letter tcheh

ﭪ MATHEMATICAL VEH ISOLATED FORM

≈ <font> <isolated> 06A4 ﭪ Arabic letter veh

ﭫ MATHEMATICAL VEH LOOPED FORM

≈ <font> <isolated> 06A4 ﭫ Arabic letter veh

ﮊ MATHEMATICAL AIN WITH TREE DOTS ABOVE ISOLATED FORM

≈ <font> <isolated> 06A0 ﮊ Arabic letter ain with tree dots above

Large mathematical alphabetic symbols

The Arabic n-ary summation operator is denoted by either $\sum$ and مجـ.

The Arabic n-ary product operator is denoted by either $\prod$ and جذ.

The Arabic limit operator is denoted by the symbol نهـا.

The Arabic factorial operator is denoted by either ! and لـ.

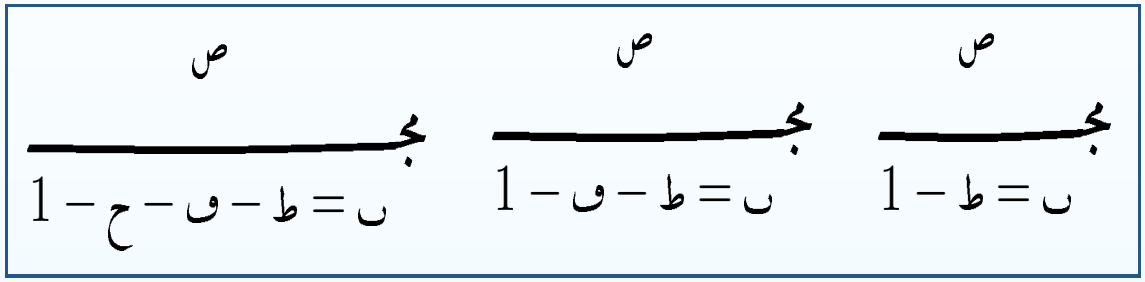
مجـ	محـ	ARABIC LARGEN-ARY SUMMATION
		≈ <font> FCCE مجـ Arabic ligature meem with jeem initial form
جذ	حد	ARABIC LARGE N-ARY PRODUCT
		≈ <font> FCCE جذ Arabic ligature jeem with thal
نهـا	هـا	ARABIC LARGE LIMIT
لـ		ARABIC LARGE FACTORIAL

The symbols Summation, Product and Limit can be denoted with dots (مجـ, جذ and نهـا) or without dots (محـ, حد and هـا).

Large mathematical alphabetic symbols

Some software tools, such as TEX or MathML, can help to combine any text string with any symbols as needed. WG2 adopted, in Resolution M38.12, not to add any more Arabic presentation forms to the standard. WG2 suggests users to make use of appropriate input methods, rendering and font technologies that meet their requirements. We propose these *large operators* for addition to the Unicode Standard even though they don't have similar entities in Latin. Actually, the shape of these ligatures is unusual compared to the layout in regular text. Moreover, these ligatures symbols are variable-sized (according to the covered expressions).

Large mathematical alphabetic symbols



Combined symbols

The Arabic operators "equal to by definition" and "equivalent to by definition" may be proposed as either compact symbols and element that can be combined with other symbols with dots or without dots.

COMBINING ARABIC DEFINITION

→ xxxx $\overset{\text{def}}{\bigcirc}$

ARABIC EQUAL TO BY DEFINITION

→ 225D $\overset{\text{def}}{=}$

ARABIC EQUIVALENT TO BY DEFINITION

→ xxxx $\overset{\text{def}}{\Longleftrightarrow}$

Accented characters

As is in Latin script based mathematical notation, only "unaccented" forms of the letters are used in Arabic mathematical notation. Mathematical symbols with diacritics are generally represented by combining characters sequences. Combining marks are used with mathematical alphabetic characters, instead of ready for use characters.

Mathematical functions names

In general, basic Arabic characters, not combined with other signs, are used to represent single-character Arabic variables. In contrast, mathematical function names like jib (sinus), jibtamam (cosinus), mamass (tangent), zill (tangent), etc., are represented by strings in usual text.

Units

The Arabic square unit, corresponding to the Latin square Km is marked with a special abbreviation ㎞. It represents an example of various signs for units that are worth of being included in the Unicode Standard.

㎞ ARABIC SQUARE KM

≈ <square>0643 0644 0645

→ 339E square km

Dates

Two main calendars are in use in the Islamic cultural area: the Hejry (Islamic calendar) and the MylAdy (Gregorian calendar). As the famous marks AC. and BC. used to distinguish years after and before the year zero. The marks ھ and م are used to make a distinction between the two calendars. These signs may be added as special characters.

ھ ARABIC HEJRY DATE

≈ <font> <isolated> 0647 ه Arabic letter heh

م ARABIC MYLADY DATE

≈ <font> <isolated> 0645 م Arabic letter meem

Mathematical symbols

In the Unicode standard, mathematical symbols are named in reference to their:

- appellation (e.g., $\sim$ *tilde*);
- representation (e.g., $\longrightarrow$ *right-arrow*);
- meaning (e.g., $<$ *less-than*);
- representation and meaning (e.g., "$($" *left parenthesis = open parenthesis*).

In the Unicode standard, the official character name is frozen once it's published. This allows the use of the character names in identifiers, and this is useful for correlating Unicode and other standards.

Aspects of mathematical symbols localization

The meaning of some mathematical symbols is not context free. For instance, in Arabic, the comma "،" as separator between items in a list is different from the one "," used to mark the limit between units and decimals. So, according to the context, the comma will have to be changed, into an Arabic comma or not. The algorithm of bidirectionality will have to transform the number's list "1,2" into "2،1" while it should leave unchanged the decimal number "1,2". Moreover, in Arabic, the comma symbol glyph "،" is not the mirrored form of the usual comma ",".

In order to get mirrored symbols as they are used in an Arabic mathematical expression, there are three possibilities:

- to eliminate the appellation in reference with the meaning. For instance, the symbol `0028` (becomes only *left parenthesis* and not also *open parenthesis*. Such eventuality is not allowed in the Unicode Standard where the name is frozen once introduced.
- to add the Arabic meaning appellation beside the old ones. For instance, the symbol `0028` (becomes *left parenthesis* = *open parenthesis* and also = *Arabic closing parenthesis*.
- to add those symbols. For instance, a new character `xxxx` (*left parenthesis* = *Arabic closing parenthesis*. The number of symbols will then be increased and there will be redundancy. The same symbol will have a double codifi cation with two opposite meanings with respect to the context.

Mirrored symbols

Many symbols used in the Arabic presentation of mathematics are only horizontally reversed Latin mathematical symbols. For instance, the symbol *contains as member*, in a left-to-right writing denotes the meaning *element of* in a right-to-left writing and vice-versa. Most of these symbols, with their initial meaning, are already contained in the Unicode Standard.

Concerning the names, there are three cases:

1. There is no clear relationship between names. For example, the symbol *element of* and its reversed form *contains as member*. It would be useful to add an equivalent new name.
2. The adjective `reversed` is added to the symbol's name. For example, the symbol *tilde* and its reversed form *reversed tilde*. Sometimes, some small changes are done.
3. The symbol's name contains the adjective `left` or `right` and its equivalent `open` or `closing`. For example, the symbol *left parenthesis* or *open parenthesis* and its reversed form *right parenthesis* or *closing parenthesis*. It may be useful to add an equivalent new name.

Mirrored symbols

2208 $\in$ ELEMENT OF

add = REVERSED CONTAINS AS MEMBER

add $\approx$ <reversed> 220B contains as member

220B $\ni$ CONTAINS AS MEMBER

add = REVERSED ELEMENT OF

add $\approx$ <reversed> 2208 element of

223C $\sim$ TILDE OPERATOR

223D $\backsim$ REVERSED TILDE

2243 $\simeq$ ASYMPTOTICALLY EQUAL TO

22CD $\backsimeq$ REVERSED TILDE EQUAL

0028 (LEFT PARENTHESIS

= OPEN PARENTHESIS

add = REVERSED RIGHT PARENTHESIS

add = ARABIC CLOSING PARENTHESIS

0029) RIGHT PARENTHESIS

For instance, the expression $(a - 2)$ in a left-to-right writing can be described as the sequence:
OPEN PARENTHESIS, a, $-$, 2, CLOSING PARENTHESIS,
or
LEFT PARENTHESIS, a,$-$, 2, RIGHT PARENTHESIS.
In a right-to-left context, for the equivalent expression $(2 - \backslash)$ the description will rather be:
OPEN PARENTHESIS, $\backslash$, $-$, 2, CLOSING PARENTHESIS,
or
RIGHT PARENTHESIS, $\backslash$, $-$, 2, LEFT PARENTHESIS.
The words `right` and `left`, referring to the written presentation of the symbol, may so be changed into `preceeding` and `following` when the reference is the meaning.

There are some frequently used Arabic symbols that have appropriate mirrored characters in the Unicode Standard. As, the number of these characters is not very important, we propose them for inclusion into the Unicode Standard. Some characters and their mirrored symbols (e.g., < and >, with an inversion in their names) are both already encoded ,
Other symbols are not frequently used. They can be obtained through mirroring.

𐆐	REVERSED THERE EXISTS
	≈ <reversed> 2203 ∃
⅁	REVERSED COMPLEMENT
	≈ <reversed> 2201 ∁
𝛛	REVERSED PARTIAL DIFFERENTIAL
	≈ <reversed> 2202 ∂
⅀	REVERSED N-ARY SUMMARY
	≈ <reversed> 2211 ∑
𝈜	REVERSED SQUARE ROOT
	≈ <reversed> 221A √
⌟	REVERSED RIGHT ANGLE
	≈ <reversed> 221F ⌞
⦣	REVERSED ANGLE
	≈ <reversed> 2220 ∠

Negated symbols

The oblique bar in negated symbols is oriented in the Arabic mathematical symbols just as it is in the Latin ones. So, mirroring the negated symbol can sometimes lead to mistakes. Generally, the division symbol is the slash symbol. Once mirrored, it will become a backslash instead of the usual slash.

2260 $\neq$ NOT EQUAL TO

$\equiv$ 003D $=$ 0338 $\not\circ$

$\not\exists$ REVERSED THERE DOES NOT EXISTS

$\approx$ <reversed> 2204 $\not\exists$

00BD $1/2$ VULGAR FRACTION ONE HALF

$\approx$ 0031 1 2044 $/$ 0032 2

Not mirrored symbols

There are some Arabic symbols with no appropriate mirrored characters in the Unicode Standard. Even tough the available fonts and rendering engines can provide the suitable mirrored glyphs of those symbols they can be proposed for inclusion into the Unicode Standard.


REVERSED CUBE ROOT

≈ <reversed> 221B


REVERSED FOURTH ROOT

≈ <reversed> 221C


REVERSED CUBE ROOT

≈ <reversed> 221B


REVERSED FOURTH ROOT

≈ <reversed> 221C

Miscellaneous symbols

There are some particular symbols that may be proposed for addition into the Unicode Standard.

Stars, asterisks and snowflakes

 OUTLINED WHITE STAR

- Five parts star sign

Miscellaneous symbols

 CRESCENT

Letter-like symbols

ﷺ ARABIC DIAMETER

General punctuation

٪ ARABIC-INDIC PERMILLE SIGN

→ `2030` ‰ permille sign

٪ ARABIC-INDIC PER TEN THOUSAND SIGN

→ `2031` ‱ per ten thousand sign

Arrows

The signs listed in the range U+2790-27FF don’t include symmetrical signs oriented right-to-left. So, supplemental arrows in Dingbat should be added. In particular, symmetrical signs for all symbols from 2794 to 27BE can be proposed with the specification LEFTWARDS, in contrast to RIGHTWARDS

Conclusions

In Arabic alphabet based scripts, mathematical expressions are, in many contexts, written from right-to-left. Specific symbols are then in use. A list of such established symbols are probably to include into the Unicode Standard since neither the bidi-algorithm nor the symbols mirroring process are enough to get these symbols. Some of alphabetic symbols, units, dates, particular mathematical symbols and arrows have also to be included into the Unicode Standard.

`http://www.ucam.ac.ma/fssm/rydarab/english/unicode.htm`