

# unicoding of math

# unicoding of math

Piotr Strzelczyk

# Unicode – what is it?

# Unicode – what is it?

Unicode is universal encoding

## Unicode – what is it?

Unicode is universal encoding,  
it assigns a unique number  
for every character.

(Including mathematical symbols.)

# Unicode – what else is it?

## Unicode – what else is it?

Unicode is international ‘industry’ standard. It is maintained by Unicode Consortium, ‘non-profit’ organization of: Adobe, Apple, Google, IBM, Microsoft, Oracle...

## Unicode – what else is it?

Unicode is international ‘industry’  
standard. It is maintained by  
Unicode Consortium,  
‘non-profit’ organization of:  
Adobe, Apple, Google, IBM,  
Microsoft, Oracle...  
and among other:  
Arthur Reutenauer

all legal gibberish about trademarks  
and copyrighted names was purposely omitted



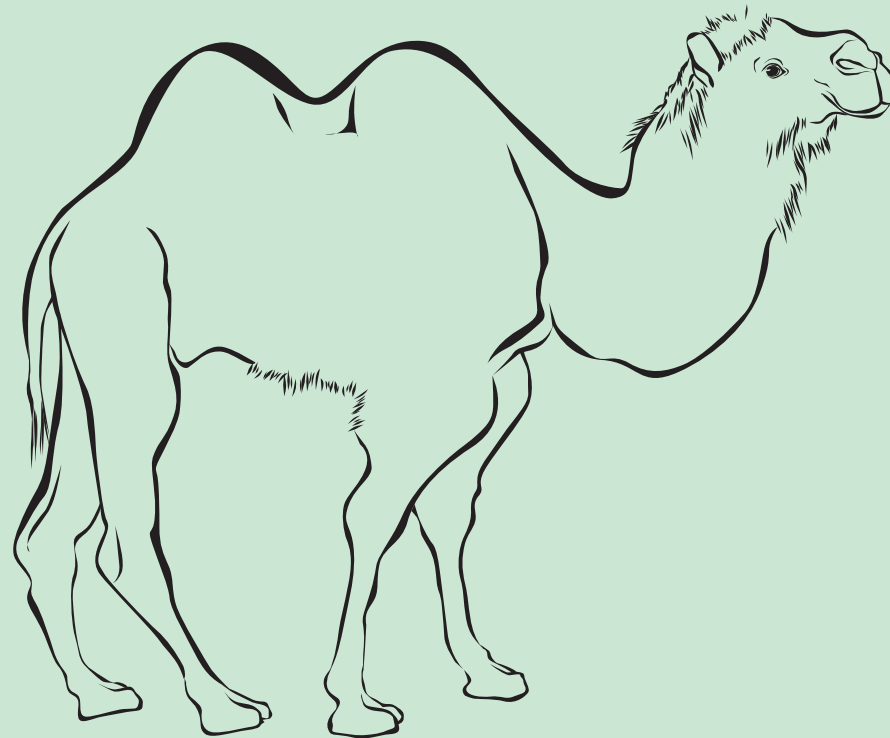
## hence, Unicode:

- is a standard ruled by drafts, reviews, releases and reports;
- tries to be up-to-date;
- want to include every character;

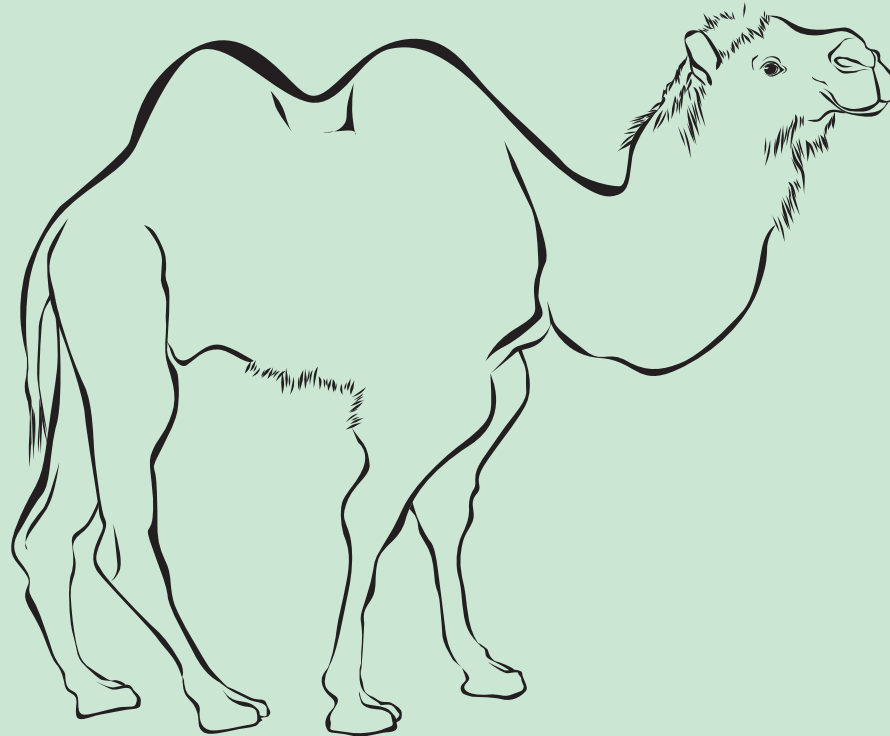
## hence, Unicode:

- is a standard ruled by drafts, reviews, releases and reports;
- tries to be up-to-date, but tries to be stable;
- want to include every character, and not to explode.  
(now it has only about 24000 chars!)

Do you know what is it?



Do you know what is it?



A horse designed by a commission

## Advantages of Unicode:

- it's really 'global'
- is rather stable
- has a wide choice of characters
- and supports all our needs  
(languages, math, music, etc.)
- UTF-8 is very convenient and compact

## Disadvantages of Unicode:

- uncustomizable
- very inert (peculiar legacy)
- unequivocal
- too extensive scope
- unpredictable, unguessable

# Math in Unicode

## history

First version of report  
'Unicode Support for Mathematics'  
was released in 2001-10-10  
(just before Unicode 3.2)

The current version is 2010-10-10  
(accompanying Unicode 6.0)

# Math in Unicode

contains:

- Mathematical Alphanumeric Symbols  
Latin, Greek, Punctuation
- Operators, Delimiters  
Mathematical Operators, Supplemental  
Math. Operators, Miscellaneous Technical
- Invisible Operators
- Arrows  
Arrows, Supplemental Arrows-A, -B
- Mathematical Alphabets  
Mathematical Alphanumeric Symbols
- Geometrical Shapes
- Superscripts and Subscripts
- Symbol Pieces



**Math in Unicode**  
looks good

# Math in Unicode

looks good

but...

# Math in Unicode

looks good

but...

peculiar legacy

U+1D5XX MATHEMATICAL FRAKTUR CAPITAL

VS

U+21XX BLACK-LETTER CAPITAL H, I, R, Z, C

U+1D5XX MATHEMATICAL SCRIPT LETTER

VS

U+21XX MATHEMATICAL SCRIPT LETTER

H, I, L, P, R, E, g, l, e, o

# Math in Unicode

looks good

but...

unequivocal

MICRO SIGN VS GREEK SMALL LETTER MU

ANGSTROM SIGN VS CAPITAL LETTER A WITH RING ABOVE

LEFT-POINTING ANGLE BRACKET VS LEFT ANGLE BRACKET  
OR MATHEMATICAL LEFT ANGLE BRACKET

BLACK DIAMOND VS BLACK DIAMOND SUIT  
OR BLACK LOZENGE

## Math in Unicode

looks good

but...

too extensive scope

Invisible operators

Supplemental Arrows-A, -B

Horizontal and vertical lines

delimiters, operators, combining chars, n-ary operators,  
terminal graphics, block elements

## Mathematical Alphanumeric Symbols

# Math in Unicode

looks good

but...

unpredictable, unguessable

U+23B4 TOP SQUARE BRACKET

U+23B5 BOTTOM SQUARE BRACKET

# Math in Unicode

looks good

but...

unpredictable, unguessable

U+23B4 TOP SQUARE BRACKET

U+23B5 BOTTOM SQUARE BRACKET

U+0346 COMBINING BRIDGE ABOVE

U+032A COMBINING BRIDGE BELOW

# Math in Unicode

looks good

but...

unpredictable, unguessable

U+23B4 TOP SQUARE BRACKET

U+23B5 BOTTOM SQUARE BRACKET

U+0346 COMBINING BRIDGE ABOVE

U+20E9 COMBINING WIDE BRIDGE ABOVE

U+032A COMBINING BRIDGE BELOW

U+???? COMBINING WIDE BRIDGE BELOW



# Math in Unicode

looks good

but...

unpredictable, unguessable

U+23B4 TOP SQUARE BRACKET

U+23B5 BOTTOM SQUARE BRACKET

U+0346 COMBINING BRIDGE ABOVE

U+20E9 COMBINING WIDE BRIDGE ABOVE

U+032A COMBINING BRIDGE BELOW

U+???? COMBINING WIDE BRIDGE BELOW

U+1F309 BRIDGE AT NIGHT

## Unpredictable Unicode continued

U+2329 LEFT-POINTING ANGLE BRACKET

U+27E8 MATHEMATICAL LEFT ANGLE BRACKET

U+3008 LEFT ANGLE BRACKET

# Unpredictable Unicode continued

U+2329 LEFT-POINTING ANGLE BRACKET

U+27E8 MATHEMATICAL LEFT ANGLE BRACKET

U+3008 LEFT ANGLE BRACKET

## Unpredictable Unicode continued

U+2329 LEFT-POINTING ANGLE BRACKET

U+27E8 MATHEMATICAL LEFT ANGLE BRACKET

U+3008 LEFT ANGLE BRACKET

U+276C MEDIUM LEFT-POINTING ANGLE BRACKET ORNAMENT

U+2770 HEAVY LEFT-POINTING ANGLE BRACKET ORNAMENT

U+27EA MATHEMATICAL LEFT DOUBLE ANGLE BRACKET

U+FE3F PRESENTATION FORM FOR VERTICAL LEFT ANGLE BRACKET

## Unpredictable Unicode continued

U+007C VERTICAL LINE

U+2223 DIVIDES

U+23B8 LEFT VERTICAL BOX LINE

U+23D0 VERTICAL LINE EXTENSION

U+258F LEFT ONE EIGHTH BLOCK

U+2758 LIGHT VERTICAL BAR

## Unpredictable Unicode continued

U+007C VERTICAL LINE

U+2223 DIVIDES

U+23B8 LEFT VERTICAL BOX LINE

U+23D0 VERTICAL LINE EXTENSION

U+258F LEFT ONE EIGHTH BLOCK

U+2758 LIGHT VERTICAL BAR

U+2016 DOUBLE VERTICAL LINE

U+2225 PARALLEL TO

U+2551 BOX DRAWINGS DOUBLE VERTICAL

## Unpredictable Unicode continued

U+007C VERTICAL LINE

U+2223 DIVIDES

U+23B8 LEFT VERTICAL BOX LINE

U+23D0 VERTICAL LINE EXTENSION

U+258F LEFT ONE EIGHTH BLOCK

U+2758 LIGHT VERTICAL BAR

U+2016 DOUBLE VERTICAL LINE

U+2225 PARALLEL TO

U+2551 BOX DRAWINGS DOUBLE VERTICAL

U+2980 TRIPLE VERTICAL BAR DELIMITER

U+2AF4 TRIPLE VERTICAL BAR BINARY RELATION

U+2AFC LARGE TRIPLE VERTICAL BAR OPERATOR

## Unpredictable Unicode continued

U+007C VERTICAL LINE

U+2223 DIVIDES

U+23B8 LEFT VERTICAL BOX LINE

U+23D0 VERTICAL LINE EXTENSION

U+258F LEFT ONE EIGHTH BLOCK

U+2758 LIGHT VERTICAL BAR

U+2016 DOUBLE VERTICAL LINE

U+2225 PARALLEL TO

U+2551 BOX DRAWINGS DOUBLE VERTICAL

U+2980 TRIPLE VERTICAL BAR DELIMITER

U+2AF4 TRIPLE VERTICAL BAR BINARY RELATION

U+2AFC LARGE TRIPLE VERTICAL BAR OPERATOR

U+1F36B CHOCOLATE BAR



## Unicode test



U+1F534



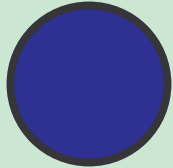
U+1F535

What color did you see?

## Unicode test



U+1F534 LARGE RED CIRCLE



U+1F535 LARGE BLUE CIRCLE

What color did you see?

Something better?

## Something better?

- named entites?  
with rules how to construct names?
- and some encodings  
for most popular tasks?
- more intelgence in programs  
and fonts (Typographical Objects)  
than in input text?

# The end



U+220E END OF PROOF

or, maybe:

U+25AE BLACK VERTICAL RECTANGLE

U+258C LEFT HALF BLOCK

U+275A HEAVY VERTICAL BAR