





















UNCOD	e.	Unicode A	llocatio	on -	BMP
0000			0000		
			0100		
1000		General Scripts	0200		Latin
			0300		Greek
2000		Symbols and	0400		Cyrillic
		General Scripts	0500		Armenian, Hebrew
3000		CJK Miscellaneous	0600		Arabic
			0700		Syriac, Thaana
4000			0800		
			0900		Devanagari, Bengali
5000			0A00		Gurmukhi, Gujarati
			0B00		Oriya, Tamil
6000			0C00		Telugu, Kannada
			0000		Malamlan Cinhala









42nd Internationalization and Unicode Conference (#) Santa Clara, September 2018



CO Controls and Basic Latin CI Controls and Latin-1 Supporting Times Developing 0	UNCODE. Org	ganized into	blocks	by script
$ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0$	C0 Controls and Basic Latin	C1 Controls and Latin-1 Supplement	Tifinagh	Devanagari
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	000 001 002 003 004 005 006 007	008 009 00A 00B 00C 00D 00E 00F	2D3 2D4 2D5 2D6 2D7	090 091 092 093 094 095 096 097
$ \begin{bmatrix} 1 & 1 & A & Q & a & q \\ 1 & 1 & A & Q & a & q \\ 2 & 1 & 1 & A & Q & a & q \\ 1 & 1 & A & Q & a & q \\ 2 & 1 & 1 & A & Q & a & q \\ 1 & 1 & A & Q & a & q \\ 2 & 1 & 1 & A & Q & a & q \\ 1 & 1 & A & Q & a & q \\ 2 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & A & Q & a & q \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1$	0 NUL DLE SP 0 @ P p p	0 XXX DCS NB ° A Đ à ð		ऐठरी अर्थका क
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	1 SOM DC1 ! 1 A Q a q	1 XXX PU1 j ± Á Ñ á ñ		ं ऑ ड र ु ं ॡ आ । विषे कि । कि । कि
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	2 STX DC2 " 2 B R b r	2 19 PH PU2 C 2 Â Ò â Ò 1000 000 1000 1000 1000 000 000 000	2 \bigoplus : 3 5 2	ं ऒ ढ ल ू ॒ ॢ
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	3 ETX 003 # 3 C S c S	3 NBH STS £ 3 Ã Ó ã Ó	3 🛛 🔨 🖇 🕺 3	ः ओ ण ळ ृ े ॣ
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	4 KOT DOG \$ 4 D T d t	4 [IN3] [CCH] X ´ Ä Ô ä Ô	4 X I O I X 4	ऄ ॵ त ऴ ॄ ॔ ।
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	5 ENG NAK % 5 E U e u	5 NEL MW ¥ µ Å Õ å õ	5 X X Q X s	अ क थ व ॅ <u> </u> ॥
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	6 ACK SYN & 6 F V f V	6 SSA SPA ¶ Æ Ö æ Ö	6 T :: Y	आ खदश ेे ००
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	7 DELL ETB ' 7 G W g W	7 (ESA) (EPA) S · C × C ÷	7 A Z :	डगधपे हा स्थ
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	8 BS CAN (8 H X h X	8 HTS SOS È Ø è Ø	8 V ···· :: 8	^{∞नं} लग खा खा खा खा ई घन स ै क २
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	9 HT EM) 9 I Y i Y	9 HTJ XXX © 1 É Ù é ù	- πα πα - F ξ Ο	<u>ज्व लग लग लग लग लग</u> उड़न हॉख ३
	A LF SUB * : J Z i Z	A VTS SCI ^a ^o Ê Ú ê ú		ू ता प्राप्त के कि उस्च प्राप्त के स
	в (VT) ESC + ; К [k {	B PLD CSI « » Ë Û ë û		अस्त कर
	C	c PLU ST ¬ 1/4 1 Ü 1 Ü		
	CRIGS - = MII m }	$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$		(전 VI 역 · · · · · · · · · · · · · · · · · ·
	$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$			·····································

42nd Internationalization and Unicode Conference (#) Santa Clara, September 2018



Cover	rage – Script	S	
 Arabic Armenian Bengali Bopomofo CJK Unified Ideographs Combining Diacritical Marks and symbols Currency Symbols Cyrillic Devanagari Dingbats Punctuation and Symbols 	 Georgian Greek Gujarati Hangul Jamo Hangul Syllables Hebrew Hiragana IPA Extensions Kanbun Kannada Katakana Lao Latin 	 Malayalam Mathematical Operators Oriya Private Use Spacing Modifier Letters Superscripts and Subscripts Surrogates Tamil Telugu Thai Tibetan 	
Unicode Tutorial Internation	nalization and Unicode Conference 42 – Santa C	lara – September 2018	20





































42nd Internationalization and Unicode Conference (#) Santa Clara, September 2018















Character Properties

Name

- General Category

 basic partition into letters
- Numbers
- Symbols
- punctuation,
- Other important general characteristics
 - Whitespace
 - Dash
 - Ideographic
 - Alphabetic
 - Noncharacter
 - deprecated

Display-related properties

- bidirectional class
- Shaping
- Mirroring
- Width

- Casing
 _ Upper
 - Lower
 - Title
 -
 - folding—both simple and full)
 - Numeric values and types
- Script and Block
 - Normalization properties
 - Decompositions
 - decomposition type
 - canonical combining class
 - composition exclusions
- Age (version of the standard in which the code point was first designated)

45

- Boundaries
 - grapheme cluster
 - Word
 - Line

•

Internationalization and Unicode Conference 42 - Santa Clara - September 2018

- sentence
- Standardized variants

Unicode Tutorial

Properties - General Lu = Letter, uppercase LI = Letter, lowercase Lt = Letter, titlecase Letters Name Lm = Letter, modifier Name_Alias Lm = Letter, modinier Lo = Letter, other Mn = Mark, nonspacing Mc = Mark, spacing combining Me = Mark, enclosing Md = Number, decimal digit NI = Number, letter Block Marks Age Numbers General_Category No = Number, other Pc = Punctuation, connector Pd = Punctuation, dash Script Punctuation Ps = Punctuation, open Pe = Punctuation, dose Pi = Punctuation, initial quote (may behave like Ps or Pe depending on White Space Alphabetic • usage) Pf = Punctuation, final quote (may behave like Ps or Pe depending on usage) Hangul_Syllable_Type Po = Punctuation, other Sm = Symbol, math Sc = Symbol, currency Noncharacter_Code_Point Default_Ignorable_Code_Point Sk = Symbol, modifier So = Symbol, other Symbols Deprecated Zs = Separator, space ZI = Separator, line Separators Zp = Separator, paragraph Cc = Other, control Logical Order Exception • Variation_Selector Cf = Other, format • Cs = Other, surrogate Co = Other, private use Other Cn = Other, not assigned (including noncha Internationalization and Unicode Conference 42 - Santa Clara - September 2018 Unicode Tutorial 46









42nd Internationalization and Unicode Conference (#) Santa Clara, September 2018













42nd Internationalization and Unicode Conference (#) Santa Clara, September 2018



Unicode Tutorial



Internationalization and Unicode Conference 42 - Santa Clara - September 2018

57























UNICODE.	Un	icode Coll	ation (UCA)	
• UT: • Mul	S#10 Iti-lev	– Unicode C el compariso	ollation Algorithn	٦
		Compariso	on Levels	
	Level	Description*	Examples	
	LI	Base characters	role < roles < rule	
	L2	Accents	role < r <u>ô</u> le < role <mark>s</mark>	
	L3	Case	role < <u>R</u> ole < rôle	
	L4	Punctuation	role < <u>"</u> role" < Role	
	Ln	Tie-Breaker	$role < ro\underline{\Box}le < "role"$	
Unicode Tutorial		Internationalization and Unicode Co	onference 42 – Santa Clara – September 2018	70



0				
A	Norma		re compa	INSON
Original	Form C	Form D	Form KC	Form KD
U+01FA	U+01FA	U+0041 U+0301 U+030A	U+01FA	U+0041 U+0301 U+030A
U+00C5 U+0301	U+01FA	U+0041 U+0301 U+030A	U+01FA	U+0041 U+0301 U+030A
U+212B U+0301	U+212B U+0301	U+212B U+0301	U+01FA	U+0041 U+0301 U+030A
U+0041 U+030A U+0301	U+01FA	U+0041 U+0301 U+030A	U+01FA	U+0041 U+0301 U+030A








UCA Dem	0		
CU = Demonstrations = Loade Epicer > Rost > French France France Service Service Service France France Service Service Service Sebscule Bielan, Bens Roberts Responder Company, Careta Album Resulti, Cost. Company, Careta Service	No. Comp Ferdman, Killer Charles, Dilloud, Ecuntoria Games	, France, Gabon, Guadelloupe, G	(Bigle yeaknet
Bearson, Rewords, Saint Marchelemy, Saint Marchelem, Saint Marchelem, Topo This summer demonstrates scoring (collation) in this locate.	(
Source	Hide Collation Key	Original	Collated
lundi		00: dimanche	18: bnef
janvier.	Ontions	01: lundi	171 boef
føvrier mare	Default (L3 = L2 + Case)	021 mardi	22 43 25 37 31 30 41 04 00
dei ja Nome	Default (Den't force case)	031 Janvier	14: Boef 24 49 39 37 31 08 01 86 07 00
deja méne	Default (Parchation = Base) Default (French accents: On)	05: mars	16: boef
dela lose f	Default (Add case level: OH)	Dei déjà	13 i bæg
beg Bosť	Default (Full normalization mode: Off) Default (Add Hinagana Level: Off)	07: Meme	22 49 25 29 11 85 e8 24 04 01 85 89 89 24 05 20
Piden d	Default (Nameric collation: Off)	08: deja	26 46 25 27 81 88 81 68 60
bon f		09: mēme	091 deja 33.35 2f 2d 81 08 63 08 00
ntital péche	Reset Rules and Options	101 deja	11: dejà
peche	Fetch rules for locale	12: beeg	07: déjà
Search for: (or leave blank for no search) Overlap?		13: Boef	13 35 37 20 31 87 85 96 04 31 86 99 011 dimanche
		14: Mémé	13 34 45 26 47 31 36 35 01 le 11 le 11
	1	15: boef	17 25 57 64 34 35 47 65 12 14 55 50 00
	-1	16: D001	04: janvier 16: 26:41:51:36:35:47:01:06:01:10:00
	1	18; pêche	i02 i lundi
		191 pèché	03: mardi
	1	20: pěché	45 24 45 33 34 81 83 61 09 00
	1	21: pēche	45 24 45 51 11 10 41 00 00
		221 peche	08: Meme
	1		10: même
	1	1	· · · · · · · · · · · · · · · · · · ·
Unicode Tutorial Internationalization and	Unicode Conference 42 - Sant	a Clara – Septe	mber 2018 77



42nd Internationalization and Unicode Conference (#) Santa Clara, September 2018

















42nd Internationalization and Unicode Conference (#) Santa Clara, September 2018















Unicode Security Considerations and Mechanisms



























42nd Internationalization and Unicode Conference (#) Santa Clara, September 2018









42nd Internationalization and Unicode Conference (#) Santa Clara, September 2018





UNICODE.	http://www.unicode.org	
🗽 The Unicode (Consortium	Contact Us Site Map Search
New to Unicode General Information Students, Educators, NGOs Conferences and Tutorials FAQ The Consortium The Unicode Standard Unicode Locales (GLDR) Publications and Data Projects and Committees UTC Document Register Proposed Changes For Members	Welcome! The Unicode Consortium enables people around the world to use computers in any language. Our freely-available specifications and data form the foundation for software internationalization in all major operating systems, search engines, applications, and the World Wide Web. An essential part of our mission is to educate and engage academic and scientific communities, and the general public.	Quick Links Code Charts The Unicode Standard CLDR UTC Document Register ICU TC Calendar of Meetings Unicode Blog
Press Contact US DIALE DONATE Some of our Members:	Recent News: • MKM [UC.42: Keynote Sceaker Announced (2018.08.23) • MKM [UC.42: Keynote Sceaker Announced (2018.08.23) • MKM [UC.42: Keynote Sceaker Announced (2018.08.09) • MKM [UL moves to GitHub and JIRA (2018.07.18) • MKM Unicode 11.0 Paperback Available (2018.07.13) • Unicode Consortium Announces Version 11.0 and Version 12.0 Cover Designs (2018.07.05) Recent Releases: • Unicode CLOR 33.1 Released (2018.06.20) • Unicode CLOR 30.1 Metased (2018.06.20)	Adopt a Character
	Announcing The Unicode Standard, Version 11.0 (2018.06.05) For Public Review (Open, About): Mark Proposed Updates of Unicode Standard Annexes for Version 12.0 (2018.09.10) #44 Draft UTR #53. Unicode Arabic Mark Rendering (2018.07.16) Internationalization and Unicode Conference 42 – Santa Clara - Soctember 2018	113
Unicode rutorial	internationalization and Unicode Conference 42 – Santa Clara – September 2018	113







42nd Internationalization and Unicode Conference (#) Santa Clara, September 2018

UNICODE.

Adopt-a-Character

(funds help encode digitally disadvantaged languages)

Sponsors of Adopted Characters							
Character sponsors help support the work of the Unicode Consortium, to help modern software and computing systems support the widest range of human languages. More than 120,000 characters							
The Unicode Consortium gratefully admowledges the following generous character sponsors. Each adoption is permanent.							
35 Gold Sponsors							
Mark Davis and Anne Gundelfinger	Elastic	} Elastic	& Adobe Systems Incorporated	${\mathfrak a}$ Ann Lewnes and Greg Welch			
💩 Jason Jenkins	🖖 Vinton G. Cerf	iscourse.org	R. Martin Chavez & Adam Norbury	White Unicom Agency			
Qakland Athletics	💘 Indian Type Foundry	Candice DeStefano	Qakland Athletics	USA Pears			
S&A Group	Zespri Kiwifruit	Avocados From Mexico	Scott Spears	Fresno Grizzlies/Fresno Tacos Baseball			
Ballantine's	Ford Motor Company	Ford Motor Company	S IBM	San Jose Earthquakes			
Oakland Athletics	SMU Guildhall	dtSearch Corp.	dtSearch Corp.	MediaLab. Inc.			
JMP software from SAS	s comprigo	🚯 💷 com	™ Anne Gundelfinger	CSRA Inc.			
e Tutorial Internationalization and Unicode Conference 42 – Santa Clara – September 2018							







42nd Internationalization and Unicode Conference (#) Santa Clara, September 2018











Characters, not glyphs							
The Unicode Standard encodes characters, not glyphs							
	Character	Code Points	Linguistic Usage				
	ch	0063 0068	Slovak, traditional Spanish				
	t ^h	0074 02B0					
	x	0078 0323	Native American languages				
	Ŕ	019B 0313	10110000000				
	á	00E1 0328	Lithuanian				
	ź	0069 0307 0301	Littitualitati				
	۴	30C8 309A	Ainu (in kana transcription)				
Inicode T	icode Tutorial Internationalization and Unicode Conference 42 – Santa Clara – September 2018						





















42nd Internationalization and Unicode Conference (#) Santa Clara, September 2018









42nd Internationalization and Unicode Conference (#) Santa Clara, September 2018


























Common Locale Data Repository (CLDR)











UCD - Unicode Data

- Name
- General_Category
- Canonical_Combining_Class
- Bidi_Class
- Decomposition
 - Decomposition_Type
 - Decomposition_Mapping

Numeric

Unicode Tutorial

- Numeric_Type
- Numeric_Value
 - decimal digit
 - Digit
 - numeric

- Bidi_Mirrored
- Comments
 - Unicode_1_Name
 - ISO_Comment
- Case Mapping
 - Simple_Uppercase_Mapping
 - Simple_Lowercase_Mapping

159

- Simple_Titlecase_Mapping

UCD Property Files ArabicShaping NameAlias NormalizationCorrections BidiMirroring • Blocks PropertyAliases ٠ CompositionalExclusions PropertyValueAliases CaseFolding • Scripts Derived Age SpecialCasing EastAsianWidth • Unihan DerivedCoreProperties HangulSyllableType Jamo DerivedNormalizationProps LineBreak PropList ٠ UnicodeData • 160 Unicode Tutorial Internationalization and Unicode Conference 42 - Santa Clara - September 2018









42nd Internationalization and Unicode Conference (#) Santa Clara, September 2018

CODe.	Does
CODe.	DUCS

oes toNFC(x)=toNFC(y) and toNFD(x)=toNFC(y)?

Normalization Browser

Input string: Enter a string with \uhhhh and \Uhhhhhhh escapes Résumé or enter code points (e.g. 0061 0308 0304 ac01 f900 50000)

Decomposition exclusions:
Hangul CJK Compat.
Unicode version:
Normal Unicode 3.2
(Show Results) (Help)

Norma	lization Results	1	
Mode	Quick Check	Normalized	Text
Input		0052 00e9 0073 0075 006d 00e9	Résumé
NFD	NO	0052 0065 0301 0073 0075 006d 0065 0301	Résumé
NFC	YES	0052 00e9 0073 0075 006d 00e9	Résumé
NFKD	NO	0052 0065 0301 0073 0075 006d 0065 0301	Résumé
NFKC	YES	0052 00e9 0073 0075 006d 00e9	Résumé
FCD	YES	0052 00e9 0073 0075 006d 00e9	Résumé

Normalization Browser

Input string: Enter a string with \uhhhh and \Uhhhhhhhh escapes

Re\u0301\sume\u0301| or enter code points (e.g. 0061 0308 0304 ac01 f900 50000)

Decomposition exclusions:
Hangul CJK Compat.
Unicode version:
Vinicode version:
Vinicode 3.2
(Show Results) (Help)

Quick Check	Normalized	Text
	0052 0065 0301 0073 0075 006d 0065 0301	Résumé
YES	0052 0065 0301 0073 0075 006d 0065 0301	Résumé
MAYBE	0052 00e9 0073 0075 006d 00e9	Résumé
YES	0052 0065 0301 0073 0075 006d 0065 0301	Résumé
MAYBE	0052 00e9 0073 0075 006d 00e9	Résumé
YES	0052 0065 0301 0073 0075 006d 0065 0301	Résumé
	YES MAYBE YES MAYBE YES	Jurick Check Normalized 0052 0065 0301 0073 0075 0064 0065 0065 0061 0073 0075 0064 0065 0061 0073 0075 0064 0065 0061 0073 0075 0064 0065 0061 0073 0075 0064 0065 0061 0062 0065 0061 0073 0075 0064 0065 0061 0062 0065 0061 0065 0061 0062 0062 0062 0062 0062 0062 0062 0062 0062 0062 0062 0062 0062 0065 0061 0065 0061 0065 0061 0065 0061 0065 0061 0064 0065 0061 0065 0061 0065 0061 0065 0061 0065 0061 0065 0061 0065 0061 0065 0061 0065 0061 0065 0061 0065 0065 0061

Unicode Tutorial

UN

Internationalization and Unicode Conference 42 - Santa Clara - September 2018

165

Uicode.	Demo NF(C&D)	
	Normalization Browser Input string: Enter a string with \uhhhh and \Uhhhhhhhh escapes ö or enter code points (e.g. 0061 0308 0304 ac01 f900 50000) Decomposition exclusions: Hangul CJK Compat.	
	Unicode version: Current Unicode 3.2 Show Results (Help) Normalization Results	
	Mode Quick Check Normalized Text	
	Input 006f 0303 031b 0323 õ	
	NFD NO 006f 031b 0323 0303 oŢ	
	NFC NO 1ee3 0303 \tilde{q}	
	NFKD NO 006f 031b 0323 0303 o	
	NFKC NO 1ee3 0303 ϕ	
	FCD NO 006f 031b 0323 0303 o	
Unicode Tutorial	Internationalization and Unicode Conference 42 – Santa Clara – September 2018	166

Nicode.	Cc (or	ompati why th	bility Demo ne Web is NF) C ai	nd not	NFKC)	
	Nor	malizat	tion Browser				
	Input st	ring: Enter a str	ing with \uhhhh and \Uhhhhh	hhh esca	pes		
	E=mc ²						
	or enter	code points (e.	g. 0061 0308 0304 ac01 f900	0 50000)			
	Decomposition exclusions: Hangul CJK Compat. Unicode version: Current Unicode 3.2 (Help)						
	Norma	lization Results					
	Mode	Quick Check	Normalized	Text			
	Input		0045 003d 006d 0063 00b2	E=mc ²			
	NFD	YES	0045 003d 006d 0063 00b2	E=mc ²			
	NFC	YES	0045 003d 006d 0063 00b2	E=mc ²			
	NFKD	NO	0045 003d 006d 0063 0032	E=mc2			
	NFKC	NO	0045 003d 006d 0063 0032	E=mc2			
	FCD	YES	0045 003d 006d 0063 00b2	E=mc ²			
nicode Tutorial		Internationalization	ation and Unicode Conference 42 – Sa	nta Clara – S	September 2018		167



NCODE.	Word on Accent Foldings… why we need those silly accents)
	<text><text><text><text><text><text><text><text><text><text><text><text><text><text><text><text><text></text></text></text></text></text></text></text></text></text></text></text></text></text></text></text></text></text>	
Unicode Tutorial	Internationalization and Unicode Conference 42 – Santa Clara – September 2018	169







UNICODO	Unicode Compression	
Ex	ample: Öl fließt	
	Unicode code points (9 code points): 00D6 006C 0020 0066 006C 0069 0065 00DF 0074	
	Compressed (9 bytes):	
	D6 6C 20 66 6C 69 65 DF 74	
Ex	ample: Москва	
	Unicode code points (6 code points):	
	041C 043E 0441 043A 0432 0430	
	Compressed (7 bytes):	
	12 9C BE C1 BA B2 B0	
Unicode Tutorial	Internationalization and Unicode Conference 42 – Santa Clara – September 2018	173



