



Курс повышения квалификации

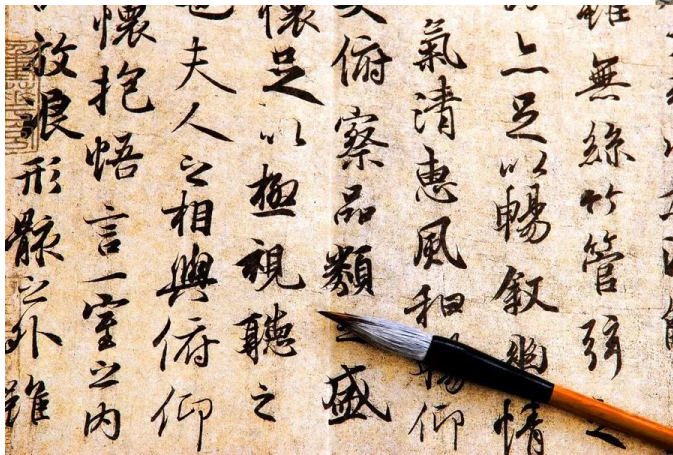
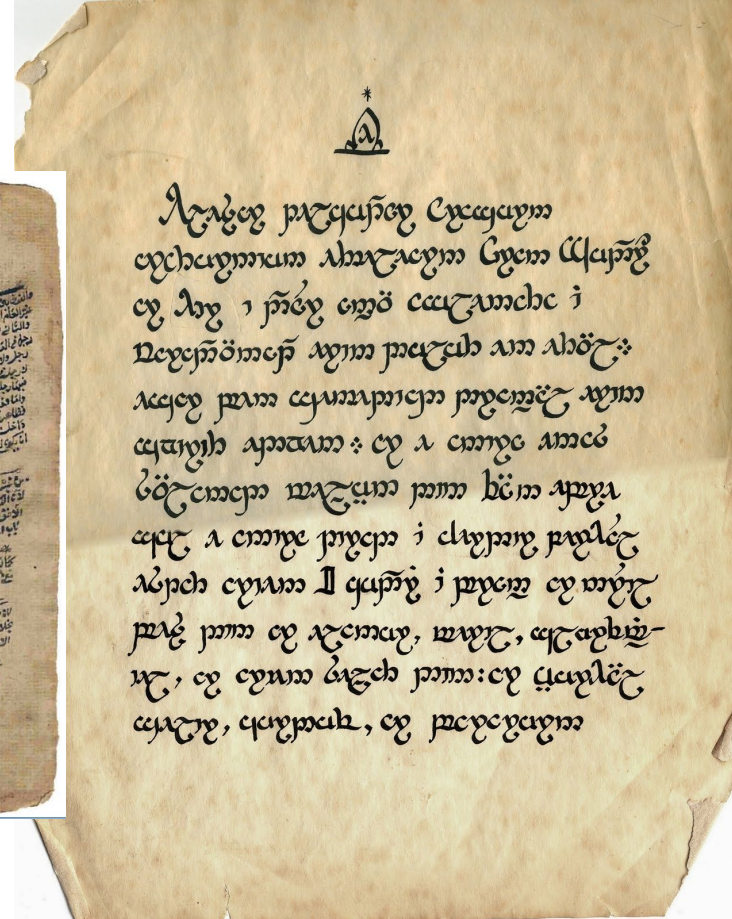
*Подготовка карт и схем для научных публикаций при помощи пакета QGIS
и среды R с использованием открытых картографических ресурсов*

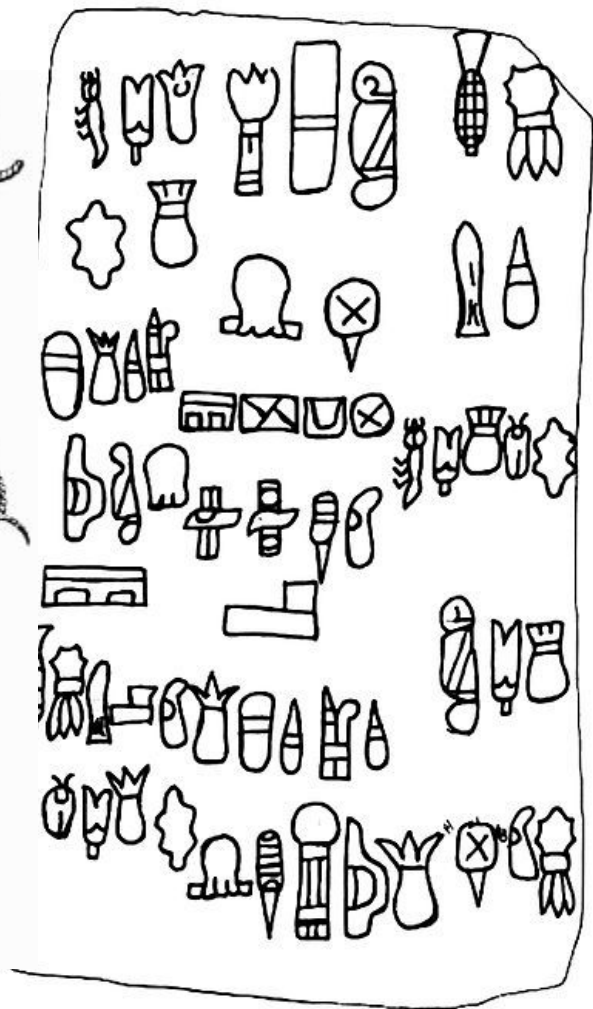


Стандарты и форматы файлов

М.П. Шашков
исследовательский парк
биотехнологий и экомониторинга


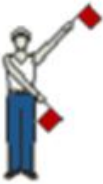


Караганда, 2024

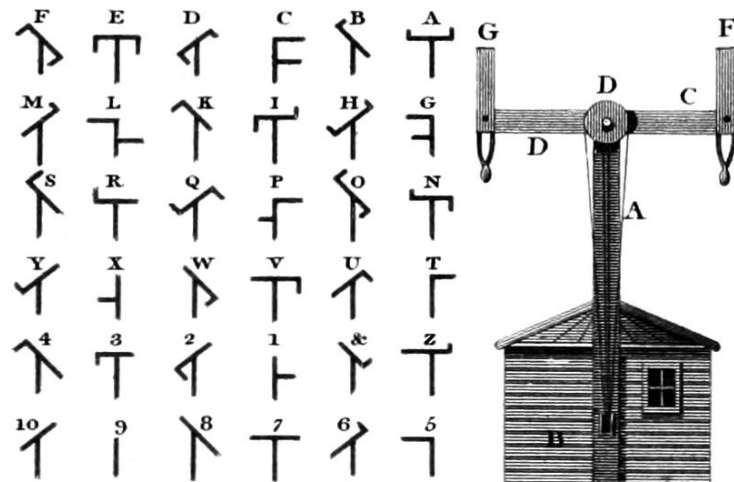




Кодировка текста

Aa	•—	Jj	•— — —	Ss	•••
Bb	—•••	Kk	—•—	Tt	—
Cc	—••—•	Ll	•—••	Uu	••—
Dd	—••	Mm	— —	Vv	••• —
Ee	•	Nn	—•	Ww	• — —
Ff	•• — •	Oo	— — —	Xx	—•• —
Gg	— — •	Pp	• — — •	Yy	—• — —
Hh	••••	Qq	— — • —	Zz	— — ••
Ii	••	Rr	• — •		

			
A	K	T	Y



Кодировка текста

ASCII 1961

American Standard Code
for Information Interchange

ASCII Code Chart

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL

S - 01010011

пробел - 00100000

~ - 01111110

0 - 00110000



W
i
k
i
p
e
d
i
a

Carriage Return
Line Feed

7 бит - $2^7 = 128$ СИМВОЛОВ

<div style="text-align: right;"> b7 b6 b5 Bits </div>						0 0 0	0 0 1	0 1 0	0 1 1	1 0 0	1 0 1	1 1 0	1 1 1
b4 ↓	b3 ↓	b2 ↓	b1 ↓	Column Row ↓	0	1	2	3	4	5	6	7	
0	0	0	0	0	NUL	DLE	SP	0	@	P	\	p	
0	0	0	1	1	SOH	DC1	!	1	A	Q	a	q	
0	0	1	0	2	STX	DC2	"	2	B	R	b	r	
0	0	1	1	3	ETX	DC3	#	3	C	S	c	s	
0	1	0	0	4	EOT	DC4	\$	4	D	T	d	t	
0	1	0	1	5	ENQ	NAK	%	5	E	U	e	u	
0	1	1	0	6	ACK	SYN	&	6	F	V	f	v	
0	1	1	1	7	BEL	ETB	'	7	G	W	g	w	
1	0	0	0	8	BS	CAN	(8	H	X	h	x	
1	0	0	1	9	HT	EM)	9	I	Y	i	y	
1	0	1	0	10	LF	SUB	*	:	J	Z	j	z	
1	0	1	1	11	VT	ESC	+	;	K	[k	{	
1	1	0	0	12	FF	FS	,	<	L	\	l		
1	1	0	1	13	CR	GS	-	=	M]	m	}	
1	1	1	0	14	SO	RS	.	>	N	^	n	~	
1	1	1	1	15	SI	US	/	?	O	_	o	DEL	

<div><div>b7b6b5</div><div>Bits</div><div>→</div></div>					<div>000</div>	<div>001</div>	<div>010</div>	<div>011</div>	<div>100</div>	<div>101</div>	<div>110</div>	<div>111</div>
<div>b4</div>	<div>b3</div>	<div>b2</div>	<div>b1</div>	<div>Column Row ↓</div>	0	1	2	3	4	5	6	7
0	0	0	0	0	NUL	DLE	SP	0	@	P	`	p
0	0	0	1	1	SOH	DC1	!	1	A	Q	a	q
0	0	1	0	2	STX	DC2	"	2	B	R	b	r
0	0	1	1	3	ETX	DC3	#	3	C	S	c	s
0	1	0	0	4	EOT	DC4	\$	4	D	T	d	t
0	1	0	1	5	ENQ	NAK	%	5	E	U	e	u
0	1	1	0	6	ACK	SYN	&	6	F	V	f	v
0	1	1	1	7	BEL	ETB	'	7	G	W	g	w
1	0	0	0	8	BS	CAN	(8	H	X	h	x
1	0	0	1	9	HT	EM)	9	I	Y	i	y
1	0	1	0	10	LF	SUB	*	:	J	Z	j	z
1	0	1	1	11	VT	ESC	+	;	K	[k	{
1	1	0	0	12	FF	FS	,	<	L	\	l	
1	1	0	1	13	CR	GS	-	=	M]	m	}
1	1	1	0	14	SO	RS	.	>	N	^	n	~
1	1	1	1	15	SI	US	/	?	O	_	o	DEL

Кодировка текста

α	β	γ	δ	ε	ζ	η	θ
альфа	бета	гамма	дельта	эпсилон	дзета	эта	тета
ι	κ	λ	μ	ν	ξ	ο	π
йота	каппа	лямбда	мю	ню	кси	омикрон	пи
ρ	σ	τ	υ	φ	χ	ψ	ω
ро	сигма	тау	ипсилон	фи	хи	пси	омега

А	Ә	Б	В
Г	Ғ	Д	Е
Ё	Ж	З	И
Й	К	Қ	Л
М	Н	Ң	О
Ө	П	Р	С
Т	У	Ұ	Ү
Ф	Х	Һ	Ц
Ч	Ш	Щ	Ъ
Ы	І	Ь	Э
Ю	Я		

Ա	Բ	Գ	Դ	Ե	Զ	Է	Ը	Թ	Ժ
a	b	g	d	je	z	e	ë	t'	ž
Ի	Լ	Խ	Ծ	Կ	Հ	Ձ	Ղ	Ճ	Ս
i	l	x	ts	k	h	dz	f	tš	m
Յ	Ն	Շ	Ո	Ր	Պ	Ջ	Ռ	Ս	Վ
j	n	š	vo	tš'	p	dž	r	s	v
Տ	Ր	Ց	Ի	Փ	Ք	Օ	Ֆ		
t	r'	ts'	w	p'	k'	o	f		

қзақша
qazaqşa
قازاقشا

.	.	•		=	l
⊙	⊙	⊙	⊙	⊙	m
Г	∨	Λ	⊥	⊥	n
Π	⊥	⊥	⊥	⊥	s
		h	⊥	⊥	s ²
=		g	≡	÷	
—	—	z	⊥	⊥	p (//)
⊥	⊥	z		≡	q
⊥	⊥	z			
⊥	⊥	h	○	○	r
⊥	⊥	i, d	≡	≡	z
⊥	⊥	z	+	+	t
⊥	⊥	k	⊥	⊥	t'

алфавит ротокас (о-в Бугенвиль) - 12 букв
кхмерский - 74

ASCII - основа всех остальных кодировок

cp866

Windows-1251

KOI-8R

UTF-8

—		Г	г	Л	л	Т	т	Т	т	■	■	■	■	■	
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒
144	145	146	147	148	149	150	151	152	153	nbsp	155	156	157	158	159
=		F	ё	е	Г	г	й	й	й	й	й	й	й	й	й
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
Й	й	Й	й	Е	е	Г	г	Й	й	Й	й	Й	й	Й	й
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
Ю	А	Б	Ц	Д	Е	Ф	Г	Х	И	Й	К	Л	М	Н	О
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
П	Я	Р	С	Т	У	Ж	В	Ь	Ы	З	Ш	Э	Щ	Ч	Б
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
Ю	А	Б	Ц	Д	Е	Ф	Г	Х	И	Й	К	Л	М	Н	О
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
П	Я	Р	С	Т	У	Ж	В	Ь	Ы	З	Ш	Э	Щ	Ч	Б
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

Кодовая таблица Windows-1251

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0		1	2	Ⓔ	Е	Ѕ	Є	·		°						
1	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
2		!	"	#	\$	%	&	()	^	+					/
3	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
4	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
5	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
6	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
7	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
8	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
9	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95
10	‘	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
11	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111
12	p	q	r	s	t	u	v	w	x	y	z	{		}	~	□
13	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127
14	Ъ	Г	,	г	„	...	†	‡		%	Љ	«	Њ	к	ћ	џ
15	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
16	ђ	‘	’	”	”	—	—			™	љ	»	њ	к	ћ	џ
17	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
18	У	у	Ј	ј	Г	г	Ѕ	Е	Є	«		»	Љ	Њ	Ћ	Ї
19	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
20	°	±	І	і	Г	г	μ	·	ё	№	є	»	Ј	Ѕ	ѕ	Ї
21	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
22	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
23	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
24	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
25	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
26	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
27	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
28	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
29	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

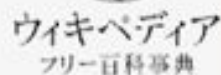
Кодировки казахского алфавита

CP 1251k.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
80	402 Ъ	403 Г	201A ,	453 ġ	201E „	2026 …	2020 †	2021 ‡	20AC €	2030 ‰	409 Љ	2039 Ѝ	40A Њ	49A Қ	4BA Һ	40F Ы
90	452 ъ	2018 ‘	2019 ’	201C “	201D ”	2022 •	2013 –	2014 —	□	2122 ™	459 Ў	203A ›	45A Ў	49B Қ	4BB Һ	45F Ы
A0	A0	4B0 Ұ	4B1 Ү	4D8 Ә	A4 #	4E8 Ә	A6 ¡	A7 §	401 Ё	A9 ©	492 Ғ	AB «	AC ¬	AD -	AE ®	4AE Ү
B0	B0 °	B1 ±	406 І	456 і	4E9 ө	B5 μ	B6 ¶	B7 ·	451 ё	2116 №	493 ғ	BB »	4D9 ә	4A2 Ң	4A3 Ң	4AF ү
C0	410 А	411 Б	412 В	413 Г	414 Д	415 Е	416 Ж	417 З	418 И	419 Й	41A К	41B Л	41C М	41D Н	41E О	41F П
D0	420 Р	421 С	422 Т	423 У	424 Ф	425 Х	426 Ц	427 Ч	428 Ш	429 Щ	42A Ъ	42B Ы	42C Ь	42D Э	42E Ю	42F Я
E0	430 а	431 б	432 в	433 г	434 д	435 е	436 ж	437 з	438 и	439 й	43A к	43B л	43C м	43D н	43E о	43F п
F0	440 р	441 с	442 т	443 у	444 ф	445 х	446 ц	447 ч	448 ш	449 щ	44A ъ	44B ы	44C ь	44D э	44E ю	44F я

Standard RK 1048-2002 encoding

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
80	4B0 Ұ	492 Ғ	201A ,	493 ғ	201E „	2026 …	2020 †	2021 ‡	20AC €	2030 ‰	4E8 Ә	2039 Ё	4A2 Ң	49A Қ	4BA Һ	4AE Ү
90	4B1 Ү	2018 ‘	2019 ’	201C “	201D ”	2022 •	2013 –	2014 —	□	2122 ™	4E9 ө	203A ›	4A3 Ң	49B Қ	4BB Һ	4AF ү
A0	A0	40E Ү	45E Ү	496 Ж	A4 #	4B2 Х	A6 ¡	A7 §	401 Ё	A9 ©	404 Ё	AB «	AC ¬	AD -	AE ®	407 İ
B0	B0 °	B1 ±	406 І	456 і	4B3 Ғ	B5 μ	B6 ¶	B7 ·	451 ё	2116 №	454 Ғ	BB »	497 Ж	4D8 Ә	4D9 ә	457 İ
C0	410 А	411 Б	412 В	413 Г	414 Д	415 Е	416 Ж	417 З	418 И	419 Й	41A К	41B Л	41C М	41D Н	41E О	41F П
D0	420 Р	421 С	422 Т	423 У	424 Ф	425 Х	426 Ц	427 Ч	428 Ш	429 Щ	42A Ъ	42B Ы	42C Ь	42D Э	42E Ю	42F Я
E0	430 а	431 б	432 в	433 г	434 д	435 е	436 ж	437 з	438 и	439 й	43A к	43B л	43C м	43D н	43E о	43F п
F0	440 р	441 с	442 т	443 у	444 ф	445 х	446 ц	447 ч	448 ш	449 щ	44A ъ	44B ы	44C ь	44D э	44E ю	44F я

[illegible][illegible]
$$\vdash \bar{a}f.\bar{a}f\% \bar{a}f \leftarrow \bar{a}f \text{ce} \bar{a}f f \bar{a}.\bar{a}.$$
[illegible]

8f58f%8,

§ 2874

 $\bar{e}^2 \bar{e}_1^2 \bar{e}_2^2$

ç'est

 $\Delta \pm 4\sigma \approx 0.5 \text{ cm}^2$ 

æ-þā—āœ-ā



À,â,ä,å,f,g,h,i,j,k,l,m,n,o,p,q,r,s,t,u,v,w,x,y,z,0-9,-.,_!@#\$%^&*~`{|}~
 à,â,ä,å,f,g,h,i,j,k,l,m,n,o,p,q,r,s,t,u,v,w,x,y,z,0-9,-.,_!@#\$%^&*~`{|}~

æ-ɬá-áCE-áð ʔʔá,áð ʔð ʔð ʔʔ%áð ʔð ʔð á,áʔáʔáʔáʔáʔáʔá,áð §æ-ɬá-á,éʔçʔáð ʔʔá,éʔáð «áð æáð — áð ʔ éʔçʔáð ʔá,CEáð ʔð ,çð ʔéáðð ʔð ʔð ʔá,

- [illegible]

சுருஷா

1 ä »ä! *äžŸä)

[illegible]

1.2 æð è¼%ãf·ã,©ãf³ãf³ã,»ãffãf³ãð ©éí·ãð ,ãð «ã,ã,ãf³ãf©ãf-ãf«

1.3 æ-tá-ā, āf³ā, āf%āf¹ā.fā³ā,*āī @ān%āāī ,āī «é-čāī mā,āf³āf@āf-āf.

1.4 āf—āfā, *āf@āf āl @æ—ʔææə-è'zā~%ā_iœāl @ç~āl •āl «ā, ā, āf~āf@āf-āf»

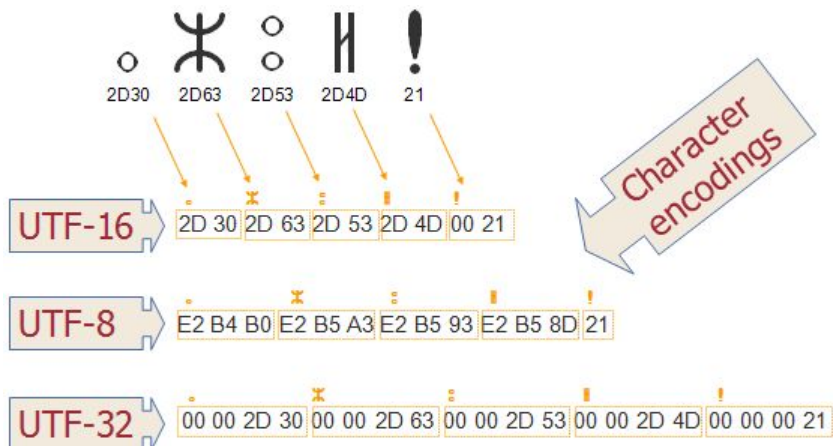
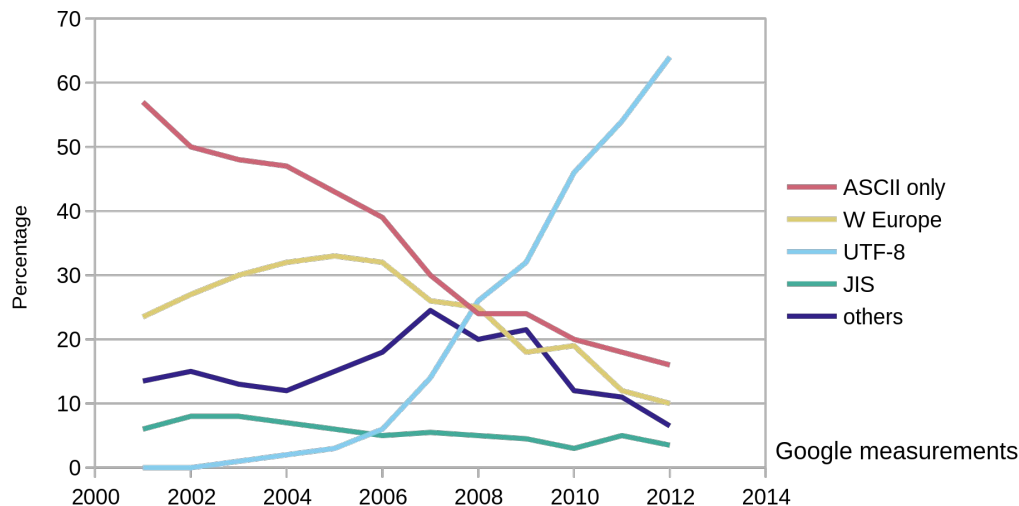
1.5 é€šăŁıçıœè~ăă şăă ăăf~ăfăăf-ăf«

UTF-8 - Unicode Transformation Format

Character		Binary code point	Binary UTF-8	Hex UTF-8
\$	U+0024	010 0100	00100100	24
£	U+00A3	000 1010 0011	11000010 10100011	C2 A3
И	U+0418	100 0001 1000	11010000 10011000	D0 98
₹	U+0939	0000 1001 0011 1001	11100000 10100100 10111001	E0 A4 B9
€	U+20AC	0010 0000 1010 1100	11100010 10000010 10101100	E2 82 AC
한	U+D55C	1101 0101 0101 1100	11101101 10010101 10011100	ED 95 9C
о	U+10348	0 0001 0000 0011 0100 1000	11110000 10010000 10001101 10001000	F0 90 8D 88

ёмкость
1,112,064
СИМВОЛОВ

Share of web pages with different encodings



Что мы используем (стандарты)

формат даты: **YYYY-MM-DD**

формат для хранения первичных данных: **CSV**
(разделители - запятые)

кодировка текста: **UTF-8**
(при сохранении и открытии файла)

для названий файлов, директорий (папок) использовать только **латиницу (без пробелов)**

система координат: **WGS 84**

для стандартизации (формализации) данных о биоразнообразии в соответствие с данными, хранящимися в GBIF используется стандарт **Darwin Core**

G	H	I
6/11/1990		1990-06-11
8/11/2002		2002-08-11
3/6/2008		2008-03-06
12/6/2003		2003-12-06
8/1/2010		2010-08-01
4/6/1997		1997-04-06
9/8/1993		1993-09-08

International Organization for Standardization

Стандартизация во всех областях,
кроме электротехники и электроники



<https://www.iso.org/home.html>

ISO 3166-1: коды стран

ISO 3166-1 alpha-2	ISO 3166-1 alpha-3	ISO 3166-1 numeric
KZ	KAZ	398
Домены первого уровня https://buketov.edu.kz Номера на автомобилях	Машиночитаемые заграничные паспорта	Не зависит от алфавита. Используется в ООН

Дата сбора (встречи, описания)

dwc: eventDate ISO 8601-1:2019 ГГГГ-ММ-ДД

2022-11-07	7 ноября 2022 г.
2022-11-01/30	1-30 ноября 2022 г.
2022-10-01/11-30	с 1 октября по 30 ноября 2022 г.
2021-01-01/2022-01-01	с 1 января 2021 г. по 1 января 2022 г.
2022	в 2022 г.
2022-11	в ноябре 2022 г.

ISO 6709

geographic coordinates

Human Interface

50° 40' 46.5" N 95° 48' 26.5" W 123.45m

50° 03' 46.5" S 125° 48' 26.5" E 978.90m

Machine readable

50.67957 -95.80737

-50.06291 125.80737

CRS: WGS 84

WHAT THE NUMBER OF DIGITS IN YOUR COORDINATES MEANS

LAT/LON PRECISION	MEANING
28°N, 80°W	YOU'RE PROBABLY DOING SOMETHING SPACE-RELATED
28.5°N, 80.6°W	YOU'RE POINTING OUT A SPECIFIC CITY
28.52°N, 80.68°W	YOU'RE POINTING OUT A NEIGHBORHOOD
28.523°N, 80.683°W	YOU'RE POINTING OUT A SPECIFIC SUBURBAN CUL-DE-SAC
28.5234°N, 80.6830°W	YOU'RE POINTING TO A PARTICULAR CORNER OF A HOUSE
28.52345°N, 80.68309°W	YOU'RE POINTING TO A SPECIFIC PERSON IN A ROOM, BUT SINCE YOU DIDN'T INCLUDE DATUM INFORMATION, WE CAN'T TELL WHO
28.5234571°N, 80.6830941°W	YOU'RE POINTING TO WALDO ON A PAGE
28.523457182°N, 80.683094159°W	"HEY, CHECK OUT THIS SPECIFIC SAND GRAIN!"
28.523457182818284°N, 80.683094159265358°W	EITHER YOU'RE HANDING OUT RAW FLOATING POINT VARIABLES, OR YOU'VE BUILT A DATABASE TO TRACK INDIVIDUAL ATOMS. IN EITHER CASE, PLEASE STOP.

Разные инструменты измерения - разная точность



1 mm



0.1 mm



0.01 mm

Точность определения координат



10 м

54.9634

37.4580

1 м

54.96342

37.45795



1 см

54.9634224

37.4579473

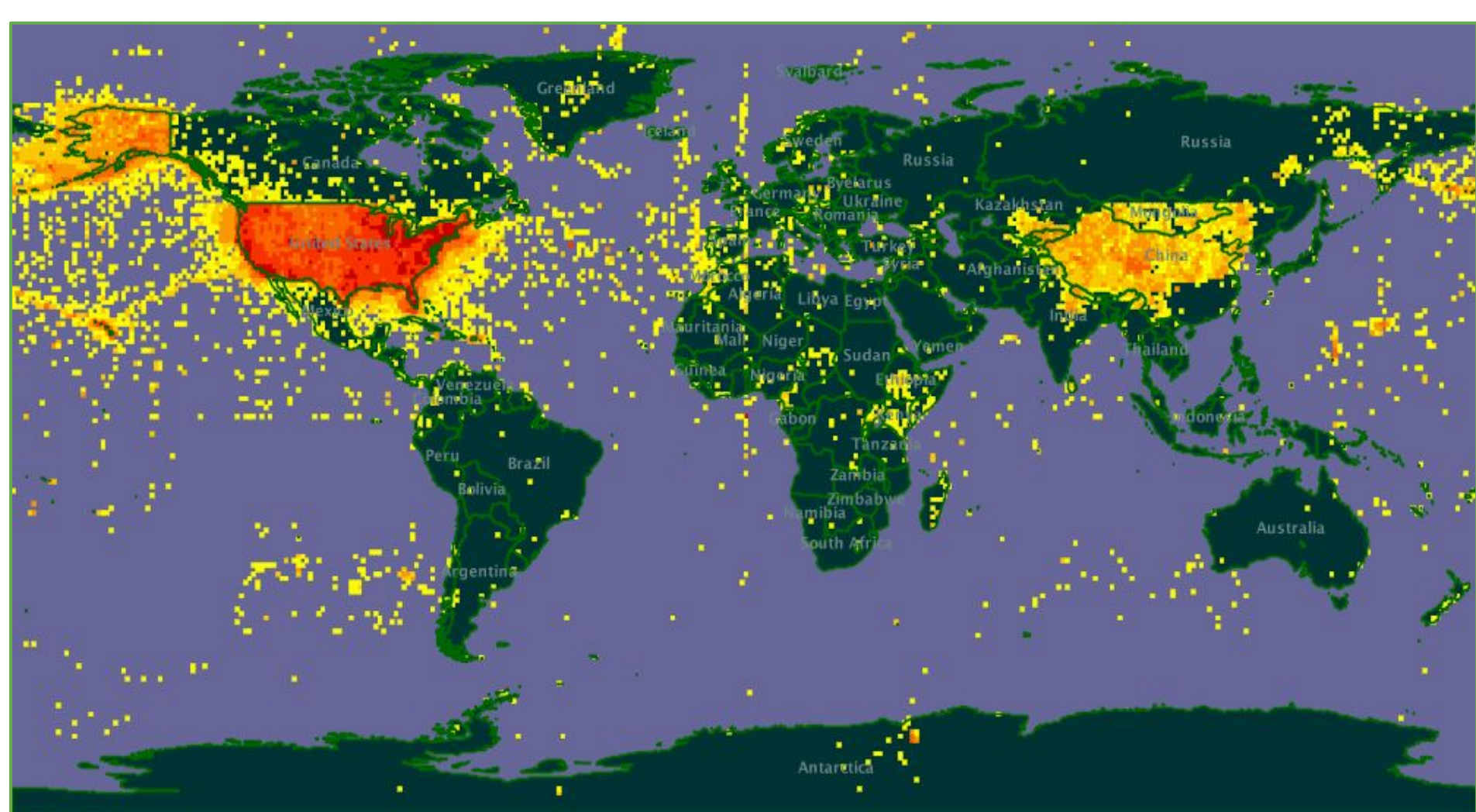
Ошибки представления координат

- Широта и долгота перепутаны местами
- Неправильно указано полушарие
- Нулевые значения
- Неизвестная система координат
- Ошибки преобразования координат из одной системы в другую или из одной формы представления в другую
- десятичные доли градусов записаны как минуты, десятичные записаны как секунды

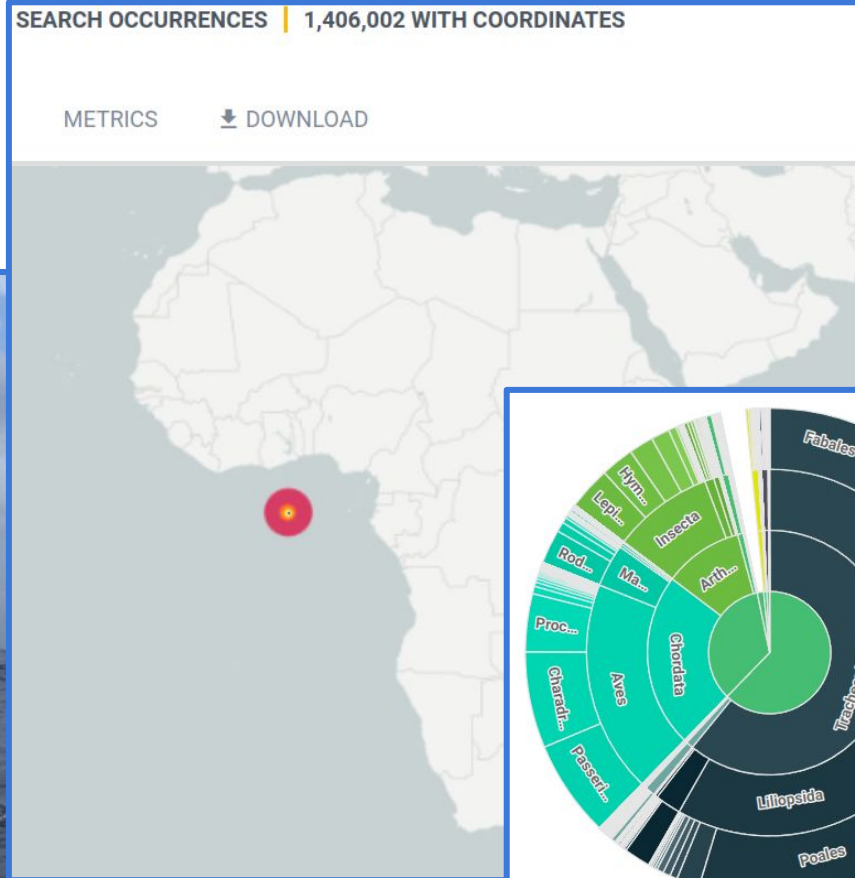
$$\text{ГГГ.ГГГГГ} = \text{ГГГ} + \frac{\text{ММ}}{60} + \frac{\text{СС.С}}{3600}$$

- Избыточная точность
- Отсутствие оценки неточности

Координаты	
В.Д.	С.Ш.
59°03'911"	62°03'604"
58°57'450"	62°04'616"
58°94'176"	62°07'776"
59°01'379"	62°04'488"
58°96'519"	62°05'083"
58°39'602"	62°09'166"
58°02'953"	61°98'325"



Null Island POINT (0,0)



Форматы файлов

Текстовые (простой без форматирования) *.txt *.csv *.tsv

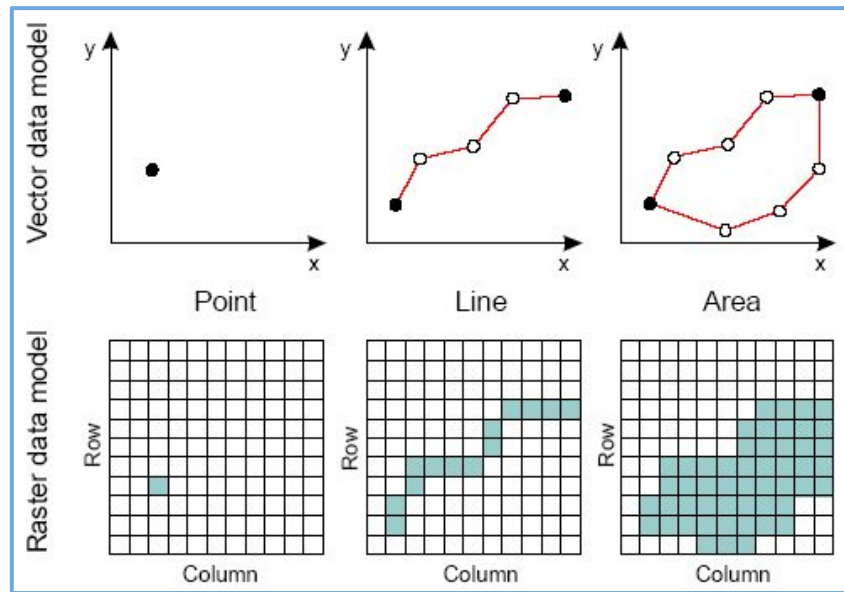
Текстовый (с форматированием) *.doc *.docx *.rtf *.odt

Табличные данные *.xls *.xlsx *.ods

Графические: растровые и векторные

*.jpg *.tiff *.bmp *.png *.gif *.pcx

*.shp *.kml *.json *.csv



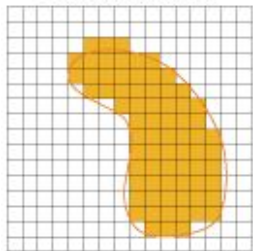
Форматы хранения растровых данных

geoTIFF - графический формат TIFF с пространственной привязкой
основной формат

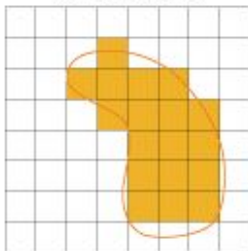
IMG, GIS - формат файла ERDAS IMAGE

GIF, JPEG, PNG, BMP, ...

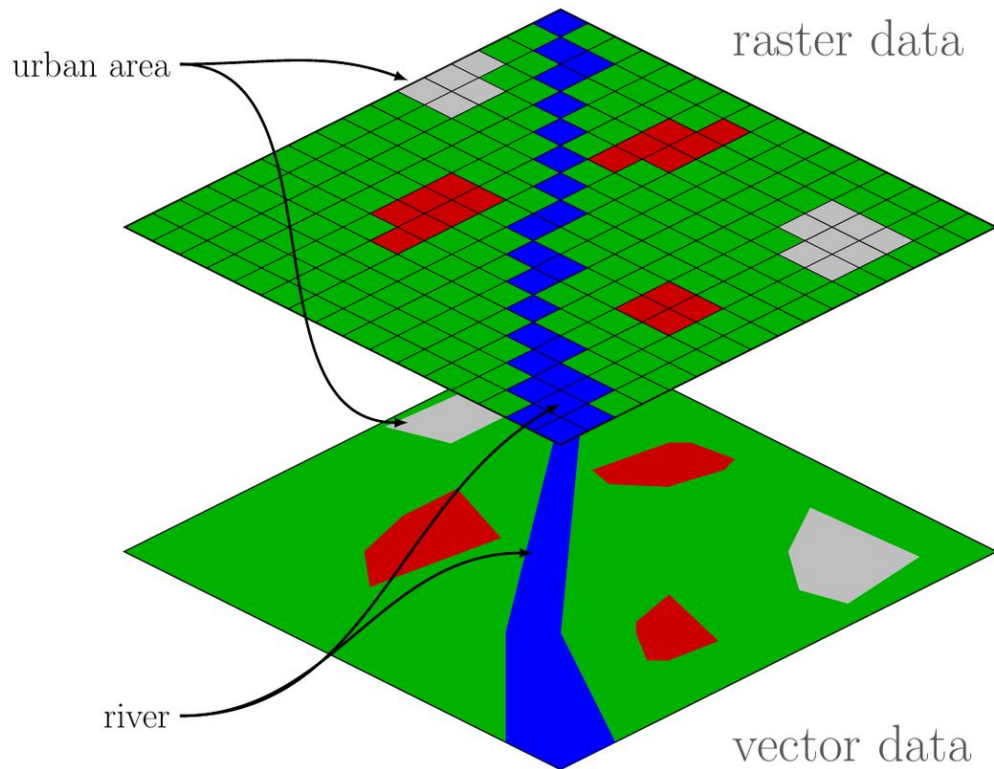
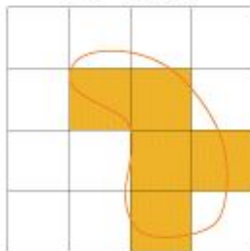
73 m²
1 m cell
16 x 16 cells



72 m²
2 m cell
8 x 8 cells



80 m²
4 m cell
4 x 4 cells



World file - файл привязки растрового слоя

map128.png

map128.pgw

32.0

0.0

0.0

-32.0

691200.0

4576000.0

масштаб по оси X

сдвиг по оси X

сдвиг по оси Y

масштаб по оси Y

X координата левого верхнего угла

Y координата левого верхнего угла

} поворот

[GIS-lab: Формат файла географической регистрации \(world-файл\)](#)

Векторный слой - ESRI Shapefile

набор файлов с одинаковым названием: Топо10km, Топо2km, gadm41_KAZ_1, ..

.shp - основной файл, геометрия

.shx - пространственный индекс

.dbf - база данных в формате dBASE

.prj - сведения о системе координат

.cpq - указание текстовой кодировки

.qml - стиль оформления слоя в формате XML

.qmd - метаданные слоя (название, краткое описание, версия, автор, ...)

...



обязательные файлы



рекомендуемые