

Telpisko datu digitālā apstrāde Biol2021

Dati, to izcelsme un ieguve

Kārlis Kalviškis, LU Bioloģijas fakultāte

2012. gada 8. februārī



GRABAŽAS IELIKSI – GRABAŽAS DABŪSI

Pasaules aprakstīšanas iespējas

- Kvalitatīvs:
 - * vārdisks (apraksti, mīti);
 - * grafisks (gleznas, fotogrāfijas, shēmas, simboli, kartes);
 - * skaņu pieraksts (putnu balsu ieraksti, mūzika);
 - * ...

Pasaules aprakstīšanas iespējas

- Kvantitatīvs:
 - * skaitlisks (vides parametri izteikti ciparu formā):
 - absolūtie skaitļi (T° Kelvinos, garums metros) var reizināt un saskaitīt;
 - relatīvie skaitļi ($t^{\circ}\text{C}$) var saskaitīt.

Pasaules aprakstīšanas iespējas

- Kvantitatīvs:
 - * simbolisks (nav iespējams veikt matemātiskas darbības):
 - kārtas skaitļi (kādā secībā skrējēji pienāk finišā)
nosaka tikai secību, matemātiskās darbības nav iespējamās;
 - skaitļi kā iezīmes (skrējēju numuri, tālruņa numuri, sadalījuma klases) –
kvantitatīvi nav savstarpēji salīdzināmi,
tikpat sekmīgi var izmantot arī burtus (automašīnu reģistrācijas numuri).

Skaitļu veidi

- Reālie skaitļi:
 - * fiksētā komata notācija;
 - * peldošā komata notācija.
- Veselie skaitļi.
- Binārie skaitļi.
(Biti, parasti izmanto 8 bitus (= 1 baits) neatvēlot vietu zīmei. Šādi iespējams aprakstīt veselos skaitļus robežas [0 – 255].)

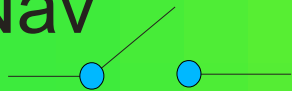
Lietotāja un datora saistība



Datu uzglabāšana datorā

- Biti, baiti.
- Binārā skaitīšanas sistēma.

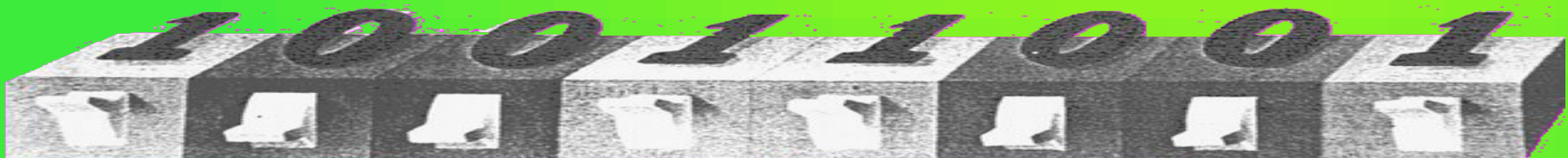
Nav



Ir



Slēdžu skaits	Varianti	Pieraksts
1	○ ●	0 1
2	○○ ○● ●○ ●●	00 01 10 11
n	2^n	



Dažādu skaitīšanas sistēmu piemēri

- Binārā (01)
 - * 101001101
- Oktālā (01234567)
 - * 515
- Decimālā (0123456789)
 - * 333
- Heksadecimālā (0123456789ABCDEF)
 - * 14D

Skaitļu pieraksts datorā

- Skaitļa precizitāti un iespējamo lielāko/mazāko skaitli nosaku skaitļa pierakstam izmantotais bitu daudzums, piemēram:
 - * reāliem skaitļiem:
 - parastu precizitāte (*single*) 32 biti;
 - dubultprecizitāte (*double*) 64 biti;
 - četrkārtīga precizitāte (*quadruple*) 128 biti;
 - * veseliem skaitļiem:
 - parasts (*short integer*) 16 biti;
 - garš (*long integer*) 32 vai 64 biti.

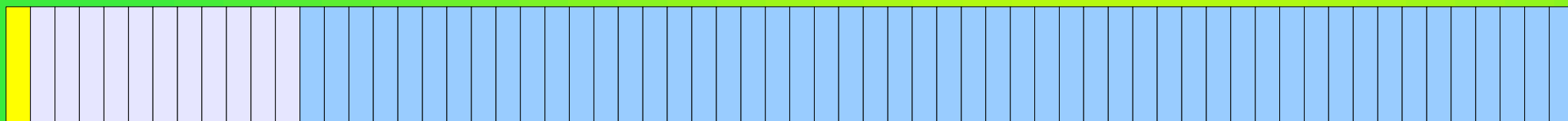
Skaitļu pieraksts datorā

- Dubultprecīzs reāls skaitlis ar peldošā komata pierakstu:

1 bits – zīme

52 biti – skaitlis (mantisa)

$$2^{52} = 4503599627370496$$



11 biti – pakāpe
–1022 līdz 1023

Skaitļu pieraksts

- Kā teksts:
 - * piemēram, 45;
 - * divi baiti, no 00 līdz 99.
- Binārais pieraksts izmantojot divus baitus:
 - * 00000000 00101101;
 - * no – 32767 līdz 32768.

Kodu tabula (8 bitu)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
2		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3	@	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	{		}	~	
8	€	,	/	"	"	...	†	‡		‰		<		..	˘	,
9		,	/	"	"	•	-	-		™		>		-		,
A			ç	£	α		ı	š	ø	©	Ŕ	«	¬		®	Æ
B	°	±	²	³	´	μ	¶	·	ø	¹	ŕ	»	¼	½	¾	æ
C	Ą	Į	Ā	Ć	Ä	Å	Ę	Ē	Č	É	Ž	È	Ğ	Ķ	Ī	Ļ
D	Š	Ņ	Ņ	Ó	Ō	Õ	Ö	×	Ū	Ł	Ś	Ū	Ü	Ž	Ž	ß
E	ą	į	ā	ć	ä	å	ę	ē	č	é	ž	è	ğ	ķ	ī	ļ
F	š	ņ	ņ	ó	ō	õ	ö	÷	ų	ł	ś	ū	ü	ż	ž	·

Simboli

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
2																
3																
4																
5																
6																
7																

Teksts un citi simboli

- 7 biti (ASCII).
- 8 biti (ANSI).
- Unicode:
 - * **UTF-8**
(vienam simbolam lieto 8, 16, 24 vai 32 bitus);
 - * UTF-16
(vienam simbolam lieto 16 vai 32 bitus);
 - * UTF-32
(vienam simbolam tiek lietoti 32 biti).

Jā/Nē lauks

- Loģiskais lauks:
 - * 1 simbols (T/F – .TRUE./FALSE.);
 - * 1 bits (0/1).

Skaitļu pārveide

- Skaitļu pārveide no viena veida citā:
 - * absolūtie skaitļi => klases;
 - * klases => ~~absolūtie skaitļi~~.

Skaitļu noapaļošana

- Skaitli vairākkārtīgi noapaļojot var iegūt citu vērtību:
 - * $\text{ROUND}(n;0) \neq \text{ROUND}(\text{ROUND}(n;1);0)$
 - * Ja $n = 9,49$
 - 9
 - 10

Saproti, ko aprēķini

- Saullēkti dažādās Latvijas pilsētās (2012.06.06):
 - * Daugavpilī 4:33;
 - * Liepājā 4:50.
- Attālums starp pilsētām ir ~ 350 km.
- Vai mēs varam aprēķināt gaismas ātrumu?

Mērvienības

- Dažadas mērīšanas sistēmas:
 - * metriskā (m);
 - * angļu (collas, jardi, jūdzes);
 - * senu mērvienību sistēmas (aršinas, verstis);
 - * ...

Mērvienības

- Līdzīga nosaukuma mērvienības:
 - * (metriskā) tonna (1 000 kg), (īsā) tonna (907,18474 kg), garā tonna (1 016,0469088 kg);
 - * (angļu) jūdze (1 609,344 m), jūras jūdze (1 852 m), metriskā jūdze (1,5 vai 1,6 km), zviedru jūdze (10,687 km),
...

Mērvienības

- Dažādi nosaukumi pakāpēm ($> 10^6$):
 - * *Long scale* ($\times 10^6$):
 - **billion** = 10^{12} – triljons, biljons;
 - **trillion** = 10^{18} – kvintiljons;
 - * *Short scale* ($\times 10^3$):
 - **billion** = 10^9 – miljards;
 - **trillion** = 10^{12} – triljons, biljons.

Laika mērvienības

- Datuma pieraksts:
 - * vislabāk lietot: gggg.mm.dd (2011.02.07);
 - * gg.mm.dd (11.02.07);
 - * dd.mm.gg (07.02.11);
 - * mm.dd.gg (02.07.11).
- Dažādas laika skaitīšanas sistēmas:
 - * laika atskaites sākums;
 - * dienu skaits gadā;
 - * gada pirmā diena;
 - * ...

Laika mērvienības

- Piemēram, Gregora kalendāra mūsu ēras 2000. gads ir kā 5101./5102. gads Kali Jugas kalendārā, 4696./4697. gads Ķīniešu kalendārā, bet Tai saules kalendārā tas ir kā Budas ēras 2543. gads.
- Gregora kalendārs tika ieviests dažādos laikos. Krievijas Pareizticīgo Baznīca vēl joprojām turas pie Jūlija kalendāra.
- Datora izpratne, par Gregora kalendāra ieviešanu:

[~]\$ cal -m 9 1752

septembris 1752

P	O	T	C	Pk	S	Sv
	1	2	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	

Laika mērvienības

- Datuma pieraksts datorā:
 - * kā simbolu virkne;
 - * kā skaitlis (laiks sekundēs no atskaites punkta, dažādām sistēmām var atšķirties).

Laiks datoros

- *Unix* vai *POSIX* sāk skaitīt no 1970. gada 1. janvāra. 32 bitu sistēmām ir iespējams ~136 gadu intervāls. Laiks sākas 1901.12.13 (mīnuss sekundes), savukārt laiks „iztecēs” 2038. gada 19. janvārī 03:14:08. (64 bitu sistēmām beigu gads ir 292 277 026 596.
- *Microsoft Windows* 1601.01.01. – 30 828.
- *Microsoft DOS* 1980.01.01. – 2099.12.31.
- *Apple Mac OS* 1904.01.01 – 2040.02.06 06:28:16,
- *Mac OS X* 2001.01.01 ±10 000 gadi.

Koordinātu pieraksts

- Ģeogrāfiskās koordinātas (grādi):
 - * 27,07939424; 56,18744321
 - * 27° 4' 45,82"; 56° 11' 14,80"
 - * 56°11'14,80"N 27°4'45,82"E
- Taisnleņķa koordinātas (m, km, ...):
 - * 691089; 231210
 - * 691089; 6231210
 - * 006-91-089E; 062-31-210N
- Atceries, ka ne vienmēr pirmais skaitlis nozīmē R-A un otrais D-Z virzienu, varbūt arī otrādi.

Telpisko datu digitālā apstrāde Biol2021

Datu telpiskais dalījums

Kārlis Kalviškis, LU Bioloģijas fakultāte

2013. gada 15. februārī.



Datu telpiskais dalījums

- Pēc administratīvām robežām.
- Pēc karšu lapām
- Pēc dabā eksistējošām robežām (ceļi, ūdensšķirtnes)

Datu glabāšana datorā

- Lai arī fiziski dati varētu būt sadalīti pa vairākiem failiem vai pat serveriem, gala lietotājam dati būtu jāvar lietot kā vienots veselums.

Telpisko datu digitālā apstrāde

Biol2021

Datu izcelsme

Kārlis Kalviškis, LU Bioloģijas fakultāte

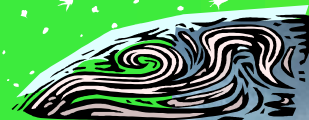


Datu izcelsme

- Dati papīrformātā:
 - * ievada ar klaviatūru, ciparotāju, skeneri, digitālo fotoaparātu;
 - * telpiskā informācija var manīties, piemēram, mitruma ietekmē.
- Dati datorformātā:
 - * dažādu datu un failu formātu savietojamība.

Datu izcelsme

- Reālās pasaules objektu mērījumi:
 - * tieši mērījumi (objekts ir pieejams un ir aptveramā lielumā);
 - * pastarpināti mērījumi (objekts ir ļoti liels, vai ļoti mazs, kad tas ir ļoti tālu, vai ir pārāk bīstams, kad pētāmie procesi ir ļoti ilgstoši vai pārāk ātri) – attālā izpēte.

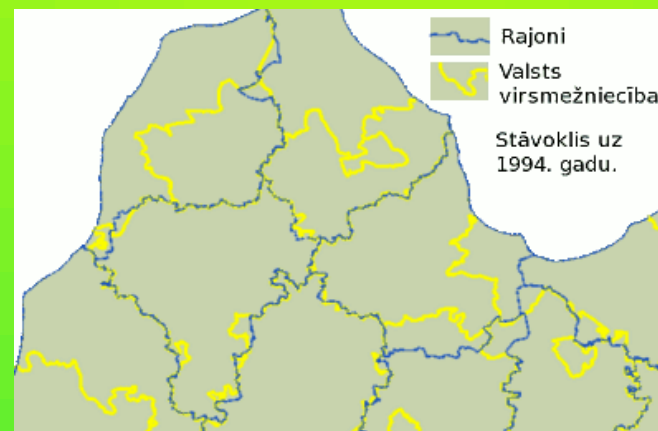
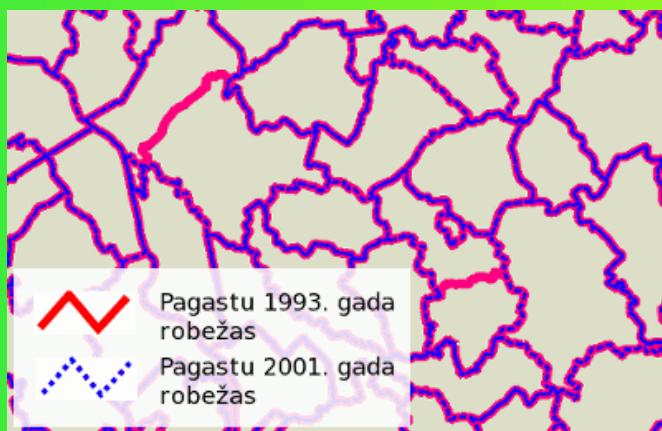


Datu izcelsme

- Reālās pasaules objektu mērījumi:
 - * mehāniskie un elektromehāniskie mērinstrumenti, sensori;
 - * dažādu raidītāju un uztvērēju izmantošana;
 - * ģeogrāfiskās atrašanās vietas noteikšanas sistēma (GPS).

Datu izcelsme

- Citu cilvēku vāktā informācija:
 - * juridiski (autortiesības, īpašumtiesības) un finansiāli šķēršļi;
 - * statistisko datu krātuves, ar to saistītās problēmas:
 - apzināta datu falsifikācija;
 - administratīvo robežu maiņas;
 - dažādi, telpiski atšķirīgi administratīvie iedalījumi.



Datu izcelsme

- Citu cilvēku vāktā informācija:
 - * grafiki, kartes, shēmas:
 - kartes ir novecojušas jau to izdošanas brīdī;
 - * satelītainas, aerofotouzņēmumi:
 - aktuāla informācija;
 - diemžēl, var būt arī sagrozītas, tas ir, „aizkrāsotas” dažas teritorijas.

Datu kvalitāte

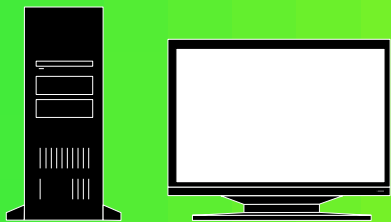
- Datu izmaiņas posmā
reālais objekts => pasaules modeļa objekts.
- Datu avota precizitāte un ticamība.
- Skaitļu mērīgā precizitāte: cik cipari aiz komata ir patiesi.



Datu kvalitāte

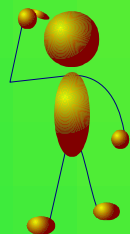
- Mērīšanas / pierakstīšanas sistēmas ierobežojumi.
- Datu veidu ierobežojumi.
- Karšu projekcijas, koordinātu sistēmas, mērogs, ģeneralizācijas pakāpe.

Datu ievadītāji (cilvēks / dators)



Dators ir AAAN:

Ašs, Akurāts, Atminīgs, Nesaprātīgs.



Cilvēks ir NAAA:

Nesteidzīgs, Aptuvens, Aizmāršīgs, Attapīgs.

Telpisko datu digitālā apstrāde

Biol2021

Objekti

Kārlis Kalviškis, LU Bioloģijas fakultāte



Objektu raksturojums

- Piesaiste telpā.
- Piesaiste laikā.
- Jēdzieniskā piesaiste.

Objektu raksturojums

- Objekta veids, klase (mājas, upes utt.).
- Objekta īpašības, atribūtinformācija (ķieģeļu ēka).
- Iespējamās darbības ar objektu (gan datu ievadīšanas brīdī, gan attēlojot objektus, gan veicot kādus aprēķinus).

Objekta *ID*

- *ID* – *identity* – vienuma iezīme. Izmanto, piemēram, datubāžu tabulu savstarpējai sasaistei.
- Vienam *ID* var atbilst viens un tikai viens objekts.
- Parasti automātiski, datorveidots.
- Varbūt arī cilvēka veidots, piemēram, parauglaukumu numuri.
- *UUID* (*Universally unique identifier*) – 128 bitu virkne, ko parasti pieraksta kā 32 heksadecimālos simbolus. Pavisam iespējami 3.4×10^{38} unikālu simbolu.

ID pieraksts

- Piemēram, ja atvēlēti divi baiti:
 - * ja izmanto simbolus (cilvēkam saprotams pieraksts), tad 00 – 99 (100 varianti), iesaistot burtus, var iegūt $(26 + 10)^2 = 1296$ variantus;
 - * ja izmanto bināro pierakstu (tikai datoram saprotams pieraksts), tad ar diviem baitiem var nokodēt $2^{16} = 65536$ variantus.

Objekta klases *ID* (klasifikācija)

- Parasti cilvēka veidots.
- Nosacīti varētu izdalīt divus *ID* veidošanas principus
 - * Pēc kārtas.
 - Nevar iespraust pa vidu jaunas objektu klases *ID*.
 - Pēc *ID* vērtības grūti saprast dotās klases vietu klasifikatorā.
 - * Pēc noteiktas sistēmas (grupējot klases) – hierarhiska klasifikācija.
 - Pēc *ID* var noteikt dotās klases vietu klasifikatorā.

Objektu klases *ID* pieraksts

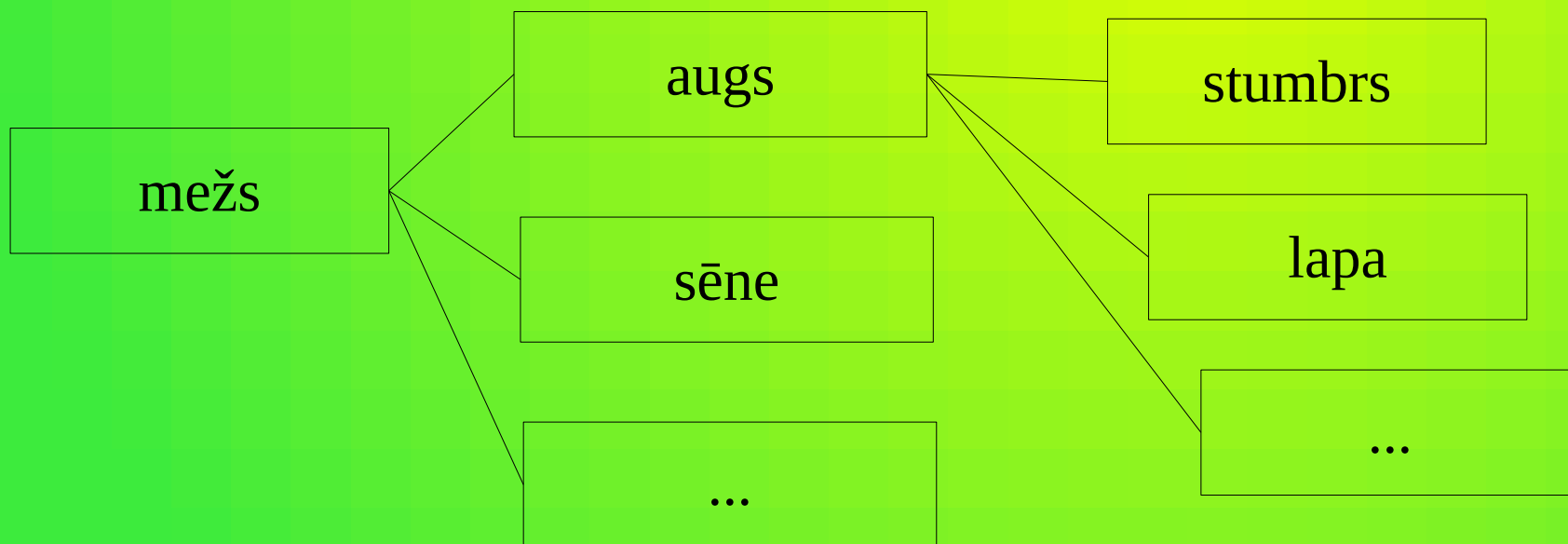
- Katram līmenim atvēl noteiktu simbolu skaitu, piemēram, **AA123C**.
 - * Problēmas, kad dotā līmeņa klašu skaits pārsniedz tam atvēlēto vietu.
- Klašu līmeņi atdalīti ar noteiktu simbolu, piemēram, **AA-123-C**.
 - * Ir vieglāk papildināmi (arī ar apakšklasēm).

Objektu raksturojums

- Ģeometriskais raksturojums, telpiskās dimensijas:
 - * 0D – punkts;
 - * 1D – līnija;
 - * 2D – figūra plaknē;
 - * 3D – figūra, kurai ir augstums, garums un platums.
- Laika dimensija.
- Objektu savstarpējās saiknes (gan telpiskas, gan funkcionālas)

Objektu raksturojums

- Objekts var sastāvēt no citiem objektiem.



Objektu raksturojums

- Ne vienmēr mūs interesē „elementārdaļiņu” telpiskā piesaiste.

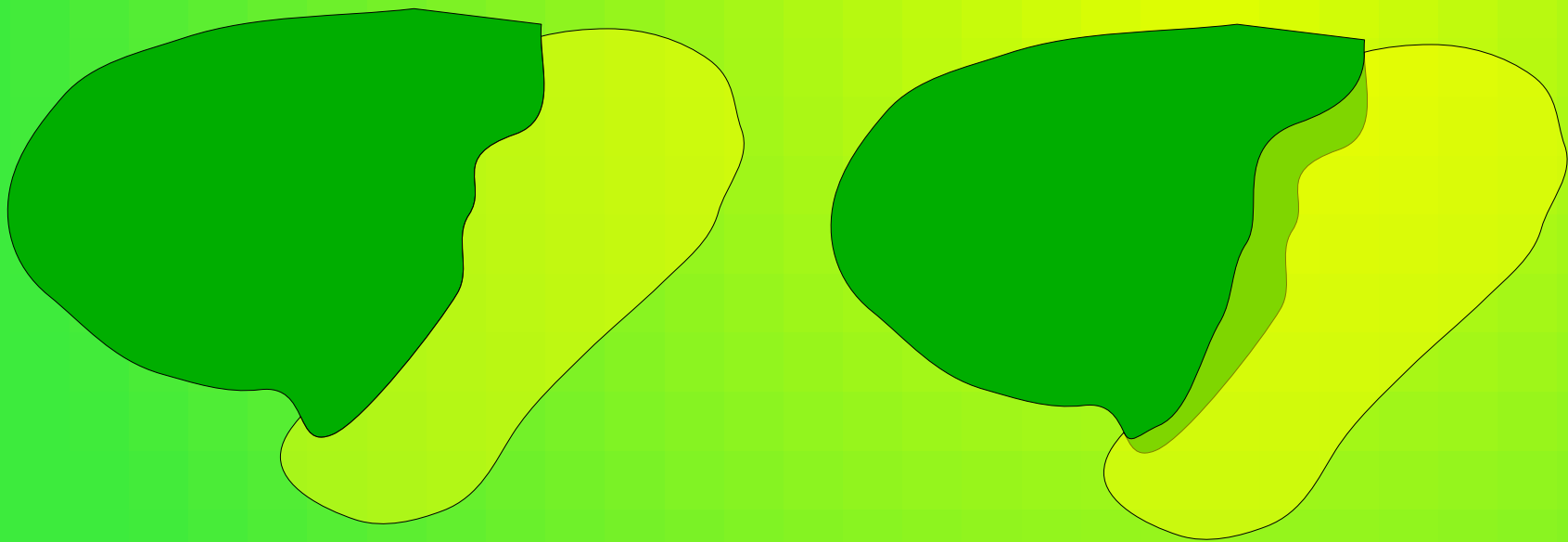


Objektu telpiskie parametri – iespējamās neskaidrības

- Robežas.
- Centrs.

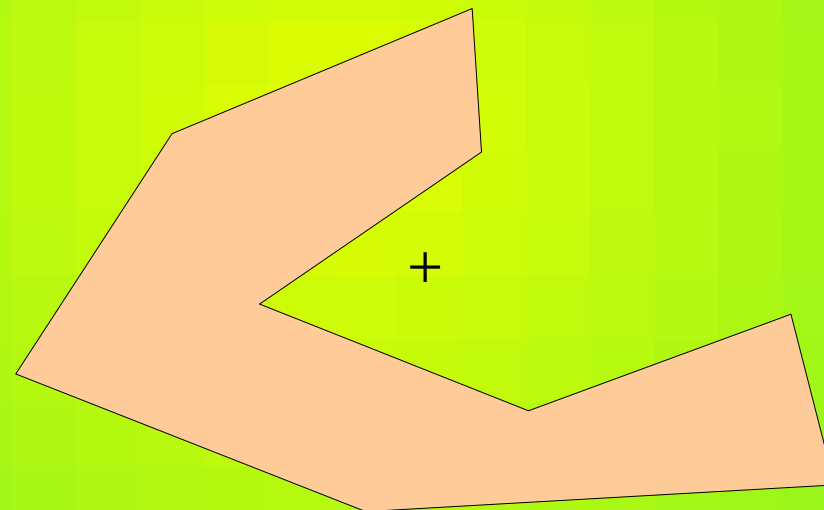
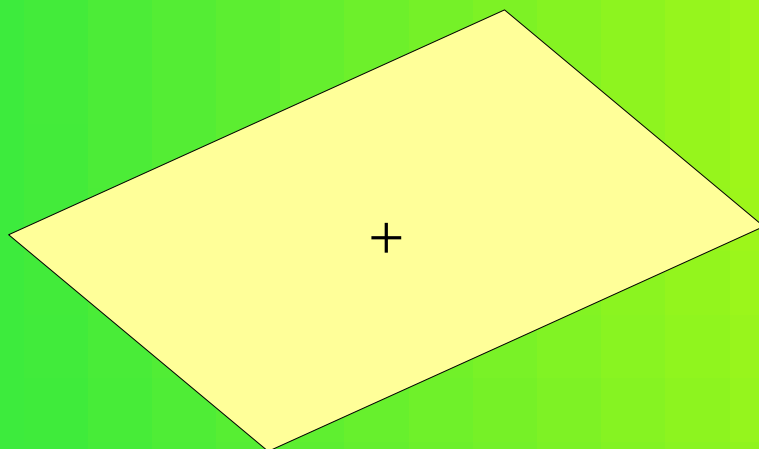
Robežu problēma

- Ne vienmēr starp diviem kaimiņos esošiem vienāda nozīmes objektiem var atrast robežu.



Ne visa programmatūra spēj darboties ar izplūdušām robežām (izplūdušo loģiku, *fuzzy logic*).

Kur ir objekta centrs ?



- Tā kā centrs var atrasties arī ārpus objekta, to ne vienmēr var izmantot par piesaistes punktu.

Kas ir objekta centrs?

- Aritmētiskais vidējais starp galējām koordinātu vērtībām. Var atrasties ārpus objekta.
- Mazākās objektu ietverošas riņķa līnijas centrs. Var atrasties ārpus objekta.
- Lielākās objektā ievilkta riņķa līnijas centrs (nepieejamības centrs). Var būt vairāki.
- Siena kaudzes centrs – vieta ar minimālo vidējo attālumu nešķērsojot robežu līdz jebkuram objekta punktam. Noteikti atradīsies objektā.
- ...

Kuru punktu izvēlēties ?

- Ja objektu piesaista telpai izmantojot tikai vienu punktu, jābūt skaidri dokumentētam algoritmam, pēc kā dotais punkts ticis izvēlēts. Tas:
 - * var būt kāds no objekta robežas punktiem, piemēram, tālākais Ziemeļu punkts;
 - * var tikt izvēlēts vadoties pēc objekta iekšējās struktūras, piemēram, pilsētas galvenā pasta ēkas;
 - * ...

Telpisko datu digitālā apstrāde

Biol2021

Metadati

Kārlis Kalviškis, LU Bioloģijas fakultāte



Datu aprakstīšana

- Dati bez datiem par datiem (**metadati**) nav lielā vērtē.
- Par ko ir dati, kādu teritoriju aptver, kādu laika periodu aptver.
- Kurš, kad, kur un kā ievācis datus.
- Kam pieder dati un kā tos var izplatīt.

Ar metadatiem saistīti standarti

- *ISO-19115 – Geographic information – Metadata.*
Ģeogrāfiskas informācijas metadati.
- *ISO-19139 – Geographic information Metadata XML schema implementation.*
Ģeogrāfiskas informācijas metadatu glabāšana *XML* tipa failos.

Likumdošana

- *INSPIRE* – Eiropas Parlamenta un Padomes direktīva 2007/2/EK par Telpiskās informācijas infrastruktūras izveidei Eiropas Kopienā. Direktīvas mērķis ir radīt vairāk un labākus publiski pieejamus ģeotelpiskos datus.
- 22.03.2011. MK noteikumi Nr. 211 „Noteikumi par ģeotelpisko datu kopu metadatu obligāto saturu” [spēkā no 26.03.2011.] – papildinājums Eiropas Komisijas 2008. gada 3. decembra Regulai (EK) Nr. 1205/2008 par Eiropas Parlamenta un Padomes Direktīvas 2007/2/EK īstenošanu attiecībā uz metadatiem.

Datu aprakstīšana

- Ievākšanas laiks un vieta un kas to vecis.
- Ievākšanas metodika.
- Ticamības un precizitātes novērtējums.
- Formāts, mērvienības.
- Klasifikācijas veids.
 - * Jebkura klasifikācija ir falsifikācija
/prof. Kristaps Rudzītis/.
- Koordinātu sistēma.
- Cita informācija atbilstoši noteikumiem un regulai.

Standartizācija

- Specifiski vākti un apkopoti dati der vienam, bet neder daudziem.
- Tikai pēc noteiktiem standartiem vākti, apkopoti, apstrādāti un saglabāti dati ir salīdzināmi un vairākkārt izmantojami.

Telpisko datu digitālā apstrāde

Biol2021

Izmantotā literatūra

Kārlis Kalviškis, LU Bioloģijas fakultāte



Literatūra

- *Tor Bernhardsen*, 2002.; **Geographic information systems : an introduction**, 3rd edition. John Wiley & Sons, 448 lpp.; ISBN: 0-471-41968-0
- *Peter A. Burrough* and *Rachael A. McDonnell*, 2000. (1998.), **Principles of geographical information systems**. Oxford University Press, 346 lpp.; ISBN13: 978-0-19-823365-7; ISBN10: 0-19-823365-5

Literatūra

- *Tor Bernhardsen*, 2002.; **Geographic information systems : an introduction**, 3rd edition. John Wiley & Sons, 448 lpp.; ISBN: 0-471-41968-0
- *Peter A. Burrough* and *Rachael A. McDonnell*, 2000. (1998.), **Principles of geographical information systems**. Oxford University Press, 346 lpp.; ISBN13: 978-0-19-823365-7; ISBN10: 0-19-823365-5

Literatūra

- *Robert Laurin and Derek Thompson, 1994.;*
Fundamentals of spatial information systems,
Academic Press, 680. lpp., ISBN: 0-12-438380-7
- *Longley Paul A., Goodchild Michael F., Maguire David J., Rhind David W. (red.) , 1999.;*
Geographical information systems (Principles and Applications) Volume 1. & 2., «Longman», London, 1101. lpp, ISBN: 0-471-32182-6.

Tīmekļa lapas

- <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:2008R1205:20081224:LV:HTML>
- <http://jonins.mii.lu.lv/Raksti/>
- <http://www.likumi.lv/doc.php?id=227704>
- <http://www.lgia.gov.lv/INSPIRE.aspx>