sl

Enaindvajseta mednarodna olimpijada iz jezikoslovja

Brasilia (Brazilija), 23.–31. julij 2024

Naloga skupinskega tekmovanja

Leksikostatistika je skupina metod, s katerimi lahko na podlagi besedišča ocenimo, kako tesno so si jeziki med seboj sorodni. Te metode se običajno uporabljajo skupaj z dolgimi seznami besed, ki jih strokovnjaki ročno označijo in navedejo, ali naj bi določen par besed imel skupen izvor. Včasih pa jezikoslovci uporabljajo leksikostatistične metode za sezname besed, ki so označeni z avtomatskimi postopki. Eden takih postopkov temelji na konceptu *konzonantnih razredov*, ki ga je uvedel sovjetsko-izraelski jezikoslovec Aron Dolgopoljski leta 1964.

P.	p b ɓ ɸ β f v	K.	k g x ɣ q ɢ χ ɯ	Y.	j ç (na začetku korena)	M.	m ɱ
T.	t d ɗ θ ð ʈ ɖ	R.	r ɾ ɽ ɹ l ɬ ɮ ɭ ʎ ɬ	W.	w ʍ (na začetku korena)	N.	n ɳ ɲ ŋ
S.	s z ʃ ʒ ʂ ʐ ɕ ʑ ɕ ʝ					Q.	t͡ɬ d͡ɮ
H.	ħ ʕ ʜ ʢ ʔ h ɦ ʔ, samoglasniki in j ç w ʍ (razen na začetku korena)						

Konzonantni razredi Dolgopoljskega

Spodaj so navedeni označeni deli seznamov besed za več jezikovnih družin s celega sveta. Oznake so podane s podpisanimi številkami. Na podlagi teh seznamov so bila s pomočjo dveh poenostavljenih različic t. i. algoritma *StarlingNJ* zgrajena drevesa jezikovnih družin, vsaki besedi pa je bil dodeljen *indeks stabilnosti*. Drevesa in indeksi stabilnosti na vrhu temeljijo na ročno označenih seznamih besed, tisti na dnu pa temeljijo na avtomatsko označenih seznamih. Za vsak seznam besed sta izdelani dve drevesi: eno z algoritmom A in drugo z algoritmom B. V nekaterih primerih obstaja več možnih dreves, ki ustrezajo posameznemu seznamu besed; v takih primerih je bilo naključno izbrano samo eno drevo. Vsakemu vozlišču na vsakem drevesu je pripisana leksikostatistična razdalja. Večja kot je razdalja, tesnejše je razmerje med jezikoma. Natančneje bi bilo torej uporabiti izraz „obrnjena leksikostatistična razdalja“ namesto „leksikostatistična razdalja“. Zaradi poenostavitve je v tej nalogi uporabljen izraz „leksikostatistična razdalja“.

Indeksi stabilnosti in leksikostatistične razdalje so zaokroženi na dve decimalni mesti. Če je tretja števka za decimalno vejico manjša od 5, zaokrožite navzdol, sicer zaokrožite navzgor. Npr. 2,836 se zaokroži na 2,84; 0,705 se zaokroži na 0,71; 0,703 pa se zaokroži na 0,70. Zaokroževanje velja samo za vrednosti, ki so prikazane ljudem. Z drugimi besedami, računalnik, ki poganja algoritme, „vidi“ nezaokrožene vrednosti.

Upoštevajte, da so nekatere besede gotovo ali domnevno izposojene iz drugih jezikov. Tako je beseda **jokːi** ‘sol’ kadivejščine izposojena iz gvaranske besede **juki**, beseda **ʔaːnʲ** ‘leto’ ipajščine (mesa-grandsko narečje) pa je izposojena iz španske besede **ˈaɲo**.

V nekaterih primerih je na seznamih besed za posamezni pomen navedenih več sinonimov, ki so ločeni z vejico. Tak primer je ‘noga’ v vehoščini.

V spodnjih podatkih so vse predpone ločene z znakom „=“, vse pripone pa z znakom „-“. Nekatere besede se vedno uporabljajo s predponami. Začenjajo se z znakom „=“.

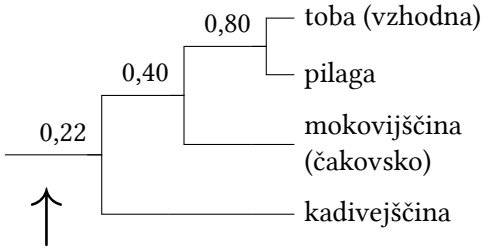
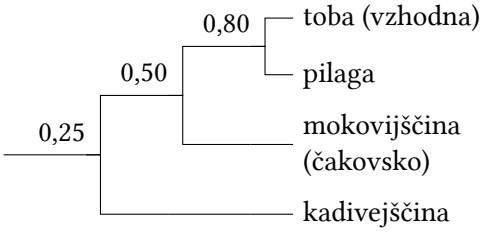
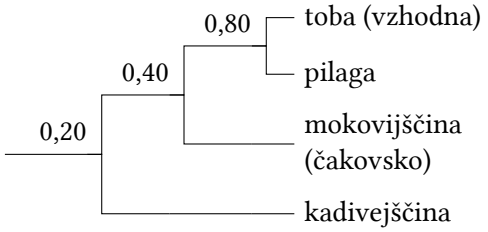
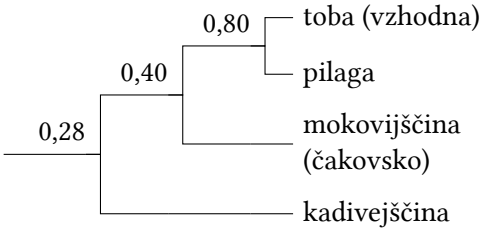
Podatki so zapisani z mednarodno fonetično abecedo. ˈ = glavni naglas, ˌ = stranski naglas (šibkejši kot glavni naglas), ◌ː = dolg glas, ◌̆ = zelo kratek glas, X͡Y = X in Y se izgovarjata kot en glas, ◌́ = visok ton, ◌̀ = nizek ton, ◌̂ = padajoč ton, ◌ˀ = predglotaliziran glas (s predhodno kratko prekinitvijo zračnega toka v grlu), ◌’ = ejektivni soglasnik (izgovorjen s kratko prekinitvijo zračnega toka v grlu), ◌̥ = nezveneč soglasnik, ◌̃ = nosni glas (izgovorjen skozi nos), ◌̝ = škripajoč glas (nizek, prasketajoč

zvok), ⁿ◌ označuje zračni tok skozi nos pred soglasnikom, ◌ʰ = pridihnjeni soglasnik (izgovori se s pridihom zraka), ◌ʷ = labializiran soglasnik (izgovorjen z zaobljenimi ustnicami), ◌ʲ = mehčan glas (izgovori se z delom jezika pri trdem nebu). **ɑ**, **æ**, **ɛ**, **ɪ**, **ɨ**, **ɔ**, **ʊ**, **ʉ**, **ə**, **ʌ**, **ɒ**, **ɘ**, **ʏ**, **ɵ**, **ø** so samoglasniki. Drugi posebni znaki so soglasniki.

⚠ Znanje jezikov, omenjenih v nalogi, ne predstavlja prednosti pri reševanju naloge.

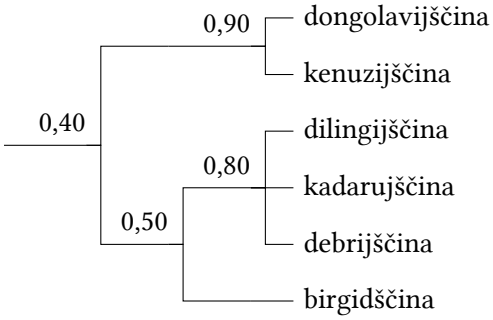
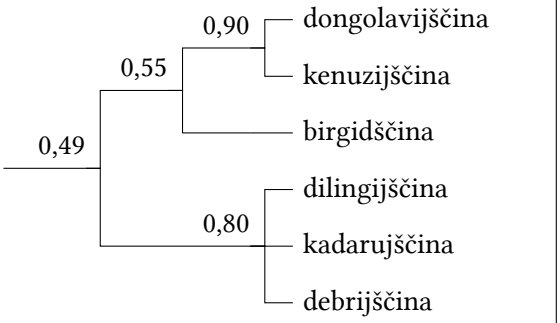
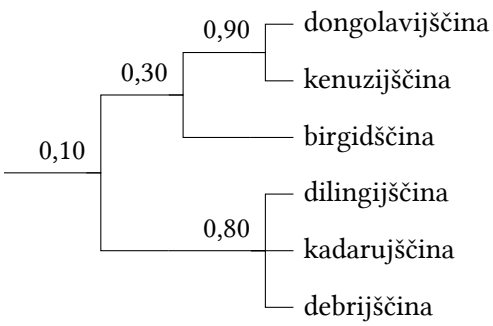
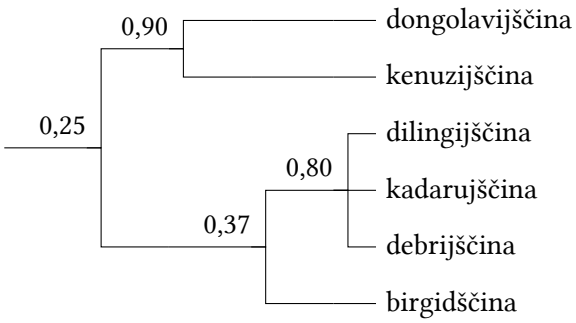
I. sklop. Gvajkurujska družina (Argentina, Brazilija, Paragvaj)

	toba (vzhodna)	pilaga	mokovijščina (čakovsko)	kadivejščina
oblak	l=ʔok$_1$	ˈlo=ʔok$_1$	naweɣelek$_2$	lol:adi$_3$
ogenj	nodek$_1$	ˈd=oleʔ$_2$	noɾek$_1$	n=ol:edi$_2$
riba	njaq$_1$	ˈnijaq$_1$	naʌin$_2$	nij:oɢo-d͡ʒegi$_3$
glava	=qajk$_1$	=ˈqajk$_1$	=qaik$_1$	=ak:ilo$_2$
ubiti	=alawat$_1$	=aˈla:t$_1$	=alawat$_1$	=el:owadi$_1$
luna	ʔawoʁojk$_1$	ʔaˈwoʕojk$_1$	ʃɪrajɣo$_2$	ep:enaj$_3$
nos	=mik$_1$	=ˈmik$_1$	=mik$_1$	=m:iq:o$_1$
sol	towe$_1$	olˈɣek$_2$	ʔwe$_1$	jok:i$_{-1}$
kamen	qaʔ$_1$	ˈqaʔ$_1$	qaʔ$_1$	wet:iɢa$_2$
jezik	=at͡ʃ-aʁat$_1$	=aˈt͡ʃ-aʕat$_1$	=oʔleɣ-aʁan-aʁat$_2$	=ok:el:i$_3$

	algoritem A	algoritem B	
ročno			Indeksi stabilnosti: oblak 0,50 ogenj 0,50 riba 0,50 glava 0,75 ubiti 1,00 luna 0,50 nos 1,00 sol 0,67 kamen 0,75 jezik 0,50
avtomatsko			Indeksi stabilnosti: oblak 0,50 ogenj 0,50 riba 0,75 glava 0,75 ubiti 1,00 luna 0,50 nos 1,00 sol 0,25 kamen 0,75 jezik 0,50

II. sklop. Nubijska družina (Egipt, Sudan)

	dongolavijščina	kenuzijščina	dilingijščina	kadarujščina	debrijščina	birgidščina
ubiti	ˈbɛː$_1$	beː$_1$	hur$_2$	wur-i$_2$	wur-i$_2$	fila:l-e$_1$
luna	uˈn-at-t$_1$	an-at-ti$_1$	nɔn-ti$_1$	nɔn-tu$_1$	nɔn-to$_1$	ma:l$_2$
voda	ˈɛss$_1$	essi$_1$	ɔti$_1$	ɔto$_1$	ɔtu$_1$	eɟi$_1$
dati	ˈtɪr$_1$	tir$_1$	ti$_1$	ti$_1$	ti$_1$	teː-n$_1$
dober	ˈsɛrɛː$_1$	sereː$_1$	ken$_2$	kɛn$_2$	kɛŋ$_2$	azze-n$_3$
veter	ˈturug$_1$	turug$_1$	irʃ-i$_2$	irʃ-o$_2$	irʃ-o$_2$	kurr-i$_3$
lasje	ˈdɪl-tɪ$_1$	siːr$_2$	tel-ti$_1$	til-tu$_1$	til-tu$_1$	ur=dill-e$_1$
trebuh	ˈtuː$_1$	tuː$_1$	te-te$_2$	to$_1$	to$_1$	tuː$_1$
spati	ˈnɛːr$_1$	neːr$_1$	ɟer$_1$	dwallɛli$_2$	ɟer-i$_1$	neːr-i$_1$
sonce	ˈmasɪl$_1$	masɪl$_1$	ɛɟ$_2$	aɟu$_2$	eŋgal-to$_3$	ʔiːzi$_2$

	algoritem A	algoritem B	
ročno			Indeksi stabilnosti: ubiti 0,50 luna 0,83 voda 1,00 dati 1,00 dober 0,50 veter 0,50 lasje 0,83 trebuh 0,83 spati 0,83 sonce 0,50
avtomatsko			Indeksi stabilnosti: ubiti 0,33 luna 0,50 voda 0,50 dati 0,67 dober 0,50 veter 0,50 lasje 0,83 trebuh 1,00 spati 0,50 sonce 0,50

(A) **(2 točki)** Soglasnik **ʁ** se izgovarja kot francoski *r*, torej z zadnjim delom jezika. V kateri razred Dolgopoljskega spada in kako ste to ugotovili?

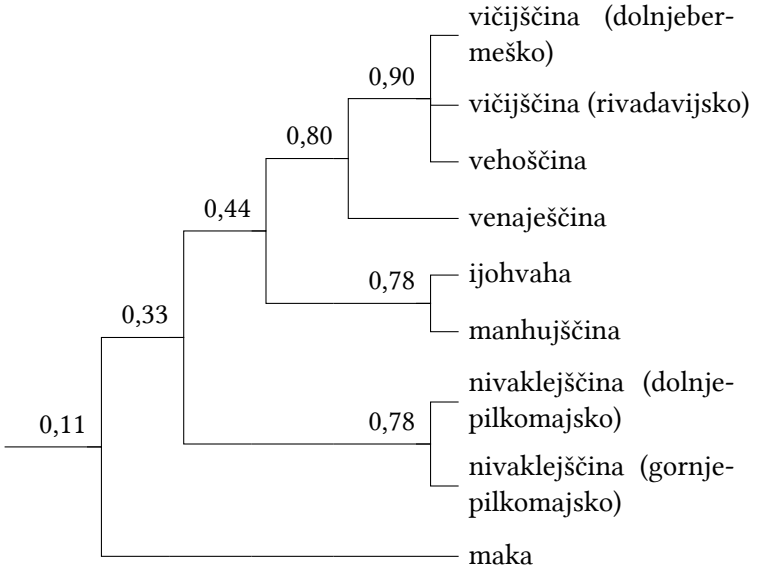
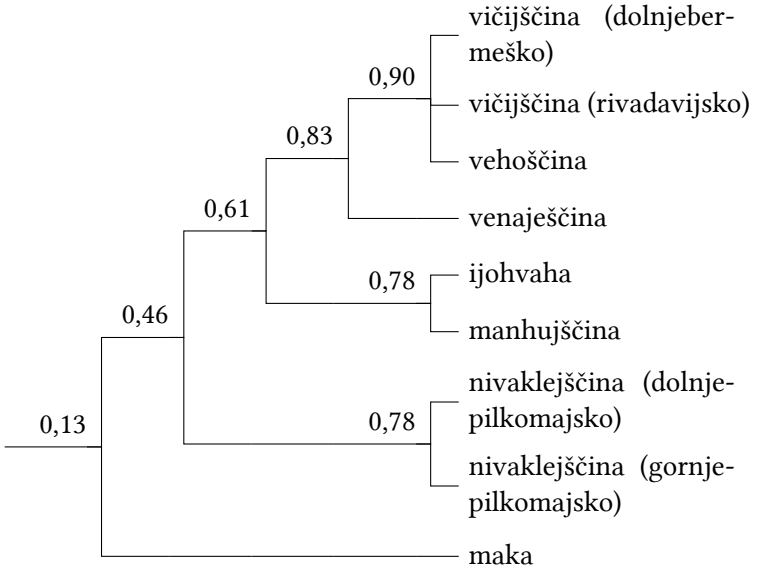
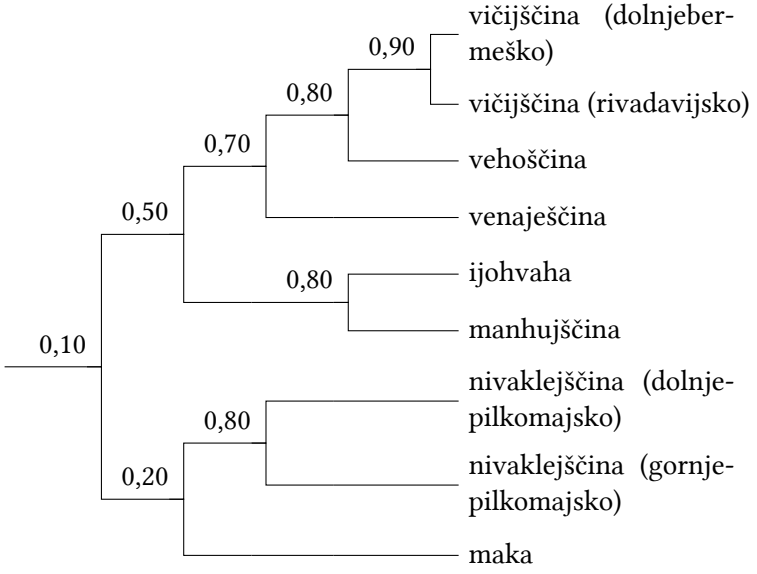
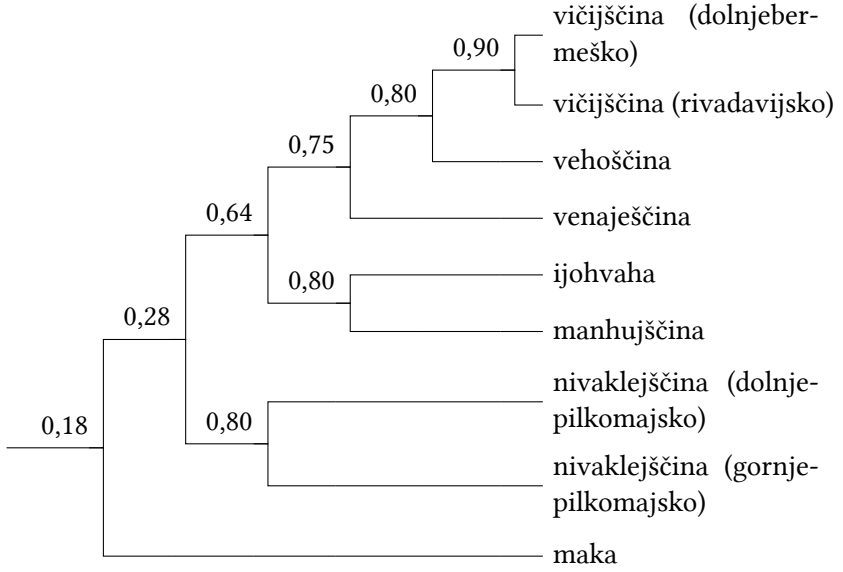
(B) **(2 točki)** Nubijsko drevo zgoraj levo je eno od dveh možnih dreves za to kombinacijo algoritma in označevalnega načina. Narišite drugo možno drevo.

(C) **(2 točki)** Nubijsko drevo spodaj levo je eno od dveh možnih dreves za to kombinacijo algoritma in označevalnega načina. Narišite drugo možno drevo.

(D) **(2 točki)** Leksikostatistična razdalja 0,49, dodeljena korenu nubijskega drevesa zgoraj desno, je zaokrožena na dve decimalni mesti, tako kot nekatere druge razdalje v tej nalogi. Kakšna je točna razdalja?

III. sklop. Matagvajska družina (Argentina, Bolivija, Paragvaj)

	vičijščina (dolnjeber-meško)	vičijščina (rivada-vijsko)	vehoščina	venaješčina	ijohvaha	manhujščina	nivaklejščina (dolnjepil-komajsko)	nivaklejščina (gornjepil-komajsko)	maka
ogenj	ʔitoχ$_{1}$	ʔitɔx$_{1}$	ʔitɑh$_{1}$	ʔiːtɑχ$_{1}$	ˈh͡wat$_{2}$	ˈʔe͡itʲe$_{1}$	ʔitɑx$_{1}$	ʔitɑx$_{1}$	feʔt$_{2}$
riba	ʔwahat$_{1}$	wahat$_{1}$	wahat$_{1}$	ʔwaːhat$_{1}$	siˈʔjus$_{-1}$	ʃiˈʔjus$_{-1}$	saxet͡ʃ$_{-1}$	saxet͡ʃ$_{-1}$	seheːt͡s$_{-1}$
noga	=pat͡ʃu$_{1}$	=qɔlɔ$_{2}$	=pat͡ʃo$_{1}$, =kɑla$_{2}$	=paːkʲʼoʔ$_{1}$	=ˈsat$_{3}$	=kaˈʔlaʔ$_{2}$	=ɸoʔ$_{4}$	=ɸoʔ$_{4}$	=fʼiʔ$_{5}$
voda	ʔinot$_{1}$	ʔinɔt$_{1}$	wah$_{2}$	ʔinɑːt$_{1}$	ʔiˈnʲat$_{1}$	ʔaˈʔnat$_{1}$	jinɑʔt$_{1}$	jinɑʔt$_{1}$	iweliʔ$_{3}$
dati	=ʔweˌn-u$_{1}$	=weˌn-u$_{1}$	=ʔweˌn-o$_{1}$	=ʔweˌn-oʔ$_{1}$	=ˈwɛh͡n-aʔm$_{2}$	=ˈhajʔ$_{3}$, =ˈwɛn$_{2}$	=xut$_{4}$	=xut-ej$_{4}$	tis-ix$_{5}$
dober	ʔis$_{1}$	ʔis$_{1}$	ʔis$_{1}$	ʔis$_{1}$	ˈʔes$_{1}$	ˈʔe͡is$_{1}$	ʔis$_{1}$	ʔis$_{1}$	t=ejkʼun-ej$_{2}$
veter	ʔinwokʷ$_{1}$	ʔinwɔk$_{1}$	ʔihwokʷ$_{1}$	=jaːɬ$_{2}$, =xʷoxʷ$_{3}$	ˈh͡lah͡wuʔ$_{4}$	ˈh͡lah͡wuuʔ$_{4}$	ɬaβiʔm$_{5}$	ɬaβiʔm$_{5}$	tʼunikʼi$_{6}$
drevo	haʔlo$_{1}$	halɔ$_{1}$	haʔlɑ$_{1}$	haʔlɑʔ$_{1}$	ʔaˈʔlaʔ$_{1}$	ʔaˈʔla-k$_{1}$	ʔaʔkxi-juk$_{2}$	jiʔk͡lɑʔ$_{1}$	naxka-k$_{3}$
lasje	=ʔwule-j$_{1}$	=wule-j$_{1}$	=ʔwole-j$_{1}$	=ʔwoːle-ç$_{1}$, hiːlenaχ$_{2}$	=ˈʔwole$_{1}$	=ˈʔwole-j$_{1}$	=ʃateʔt͡ʃ$_{3}$	=jeʔs$_{4}$	=ʔewkux-it͡s$_{5}$
ubiti	=lon$_{1}$	=lɔn$_{1}$	=lɑn$_{1}$	=lɑːn̥$_{1}$	=ˈlaʔan$_{1}$	=ˈlan$_{1}$	=k͡lɑn$_{1}$	=k͡lɑn$_{1}$	=lan$_{1}$

<table>
<tr><th></th><th>algoritem A</th><th>algoritem B</th><th></th></tr>
<tr><td>ročno</td><td> 
</td><td> 
</td><td>Indeksi stabilnosti:
ogenj 0,78
riba 1,00
noga 0,33
voda 0,78
dati 0,44
dober 0,89
veter 0,33
drevo 0,78
lasje 0,67
ubiti 1,00</td></tr>
<tr><td>avtomatsko</td><td> 
</td><td> 
</td><td>Indeksi stabilnosti:
ogenj 0,78
riba 0,44
noga 0,33
voda 0,56
dati 0,67
dober 0,89
veter 0,22
drevo 0,67
lasje 0,67
ubiti 1,00</td></tr>
</table>

IV. sklop. Mongolska družina (Ljudska republika Kitajska, Mongolija, Rusija)

(E) (10 točk) Preglejte naslednji seznam besed. Izračunajte indekse stabilnosti, ki ustrezajo ročnim in avtomatskim oznakam.

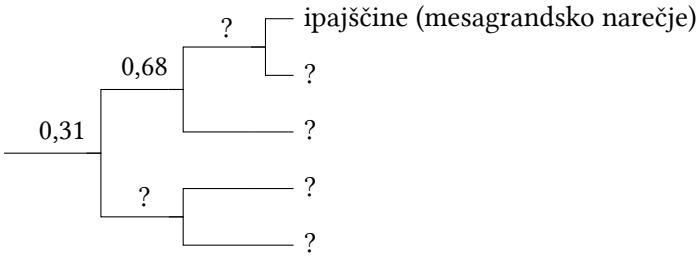
Za lažje reševanje smo za besedo ‘vse’ že izračunali obadva indeksa stabilnosti. V naključnem vrstnem redu sta to: 0,36 in 0,40.

	dagurščina (hajlarsko)	hamniganščina (mandžursko)	burjatščina (horijsko)	novobargutščina	oldščina	hošutščina	kalmiščina	halhaščina	ordoščina	širojugurščina	bonanščina
vse	hɔː$_{1}$	bʊlt$_{2}$	bʊxɨː$_{3}$	bygd$_{4}$	t͡sug$_{5}$	lug$_{5}$	t͡suk$_{5}$, xamak$_{-1}$	pux$_{3}$, pugt$_{4}$, xamăɢ$_{-1}$	pyɣyte$_{4}$, xamuk^{h}$_{-1}$	t͡ʃ^{h}uq$_{5}$	hanə$_{-2}$
lubje	hai̯l͡s$_{1}$	qalihʊn$_{1}$	χɔltəhɔn$_{2}$	xal^{j}huː$_{1}$	xolts$_{2}$	xalis$_{1}$	dursn$_{3}$	xɔɮt^{h}ɔ̈s$_{2}$	turusu$_{3}$	χalsən$_{1}$	arasun$_{4}$
trebuh	keːli$_{1}$	gətəhʉn$_{2}$	gedehen$_{2}$	gedyː$_{2}$	geːs$_{2}$	gets$_{2}$	gesn$_{2}$	gitĭs$_{2}$, xiwɮĭj$_{-1}$	ketysy$_{2}$	ketesən$_{2}$	kele$_{1}$
ptica	dəgi$_{-1}$	ɕiwan$_{1}$	ʃubuːn$_{1}$	ʃuwuː$_{1}$	ʃuvuː$_{1}$	ʃuwuː$_{1}$	ʃowun$_{1}$	ʃuwu$_{1}$	ʃuβuː$_{1}$	ʃuːn$_{1}$, pelt͡ʃɔr$_{2}$	bend͡zer$_{2}$
ogenj	ɢali$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	qal$_{1}$	qal$_{1}$	χal$_{1}$
pot	terg-uːl$_{1}$	qarɢʊ̈i̯$_{2}$	χarɢɨ$_{2}$, zam$_{-1}$	zam$_{-1}$	d͡zam$_{-1}$	d͡zam$_{-1}$	xaː-lʁə$_{3}$	t͡sam$_{-1}$	t͡ʃam$_{-1}$	mør$_{4}$	mor$_{4}$
sol	hataː$_{1}$	dawhʊn$_{2}$	dabhan$_{2}$	dawuhuː$_{2}$	daws$_{2}$	daws$_{2}$	dawsn$_{2}$	tawsă$_{2}$	taβusu$_{2}$	taːpsən$_{2}$	dabsuŋ$_{2}$
plavati	unpa-du$_{1}$	ʊmba-$_{1}$	t^{h}amar-$_{2}$	umb-$_{1}$	sele-$_{3}$	umba-$_{1}$	us-t͡ɕi-$_{4}$, øːm-$_{5}$	siɮĭ-$_{3}$	usu-t͡ʃ^{h}i-la-$_{4}$	umpa-$_{1}$	mba-$_{1}$
voda	ɔsɔ$_{1}$	ʊxʊn$_{1}$	uhan$_{1}$	uːha$_{1}$	usn$_{1}$	us$_{1}$	usn$_{1}$	ʊsŏ$_{1}$	usun$_{1}$	q^{h}usun$_{1}$	sə$_{1}$
veter	kei̯n$_{1}$	halkin$_{2}$	halxin$_{2}$	halxi$_{2}$	sal^{j}xin$_{2}$	salk^{j}i$_{2}$	sal^{j}kn$_{2}$	saɮxĭ$_{2}$	k^{h}iː$_{1}$	k^{h}iː$_{1}$	ki$_{1}$

V. sklop. Jumska družina (Mehika, ZDA)

(F) **(8 točk)** Preglejte naslednji seznam besed. Spodaj si lahko ogledate drevo, ki je bilo sestavljeno na podlagi istega seznama. Manjkajo nekateri podatki (imena jezikov in leksikostatistične razdalje). Zapolnite prazna mesta. Opredelite, ali je drevo izdelano ročno ali avtomatsko in ali je bilo generirano z algoritmom A ali B.

	mohavejščina	kokopa	javapajščina	tipajščina (hamulsko)	ipajščine (mesagrandsko narečje)
kratek	wena=wen-a_1	ˈxʌ=ʔut$_2$	ˈt͡ʃkr=ot-i_2	lə=ˈʔuɲ$_1$	mə=put-k_3
ptica	ʔit͡ʃ=i=jer$_1$	ˈʃa$_2$	ˈʔ=t͡ʃ=sa$_2$	aʔ=ˈʃa$_2$	ʔaː=ʂaː$_2$
kost	ɲ=a=s=ak$_1$	ˈɲ=j=aːk$_1$	ˈt͡ʃ=j=aːk-a_1	ˈak$_1$	aq$_1$
suh	i=roː-v-k$_1$	ˈʂ=ʔar$_2$	ˈru-β-i$_1$	ˈs=ʔaːj$_3$	saːj$_3$
meso	kʷiːkʷay$_1$	ʔi=ˈmaːt͡ʃ$_2$	ˌkʷeː=ˈθo-β-a_3	ˈkʷak$_4$	kukʷaːj-p$_1$
vrat	maʌaqe$_1$	ˈm=puk$_2$	ˈmlqí$_1$	iː=ˈpuk$_2$	iː=puk$_2$
videti	i=juː-k$_1$	ˈwiː$_2$	ˈʔuː$_1$	ˈwiːw$_2$	ə=wuːw$_2$
rep	iː=ʔar$_1$	ˈʃ=juʌ̊$_2$	ˈβ=hé$_3$	ʃə=ˈjuɬ$_2$	xə=juɬ$_2$
dva	havik-k$_1$	ˈx=wak$_1$	ˈh͡wâk-i_1	xə=ˈwak$_1$	xə=wak$_1$
leto	huːðe$_1$	ˌmat-ˈkaːm$_2$	ˈʔ=t͡ʃʰur-a_3	mat-ˈwam$_2$	ʔaːnʲ$_{-1}$



(G) **(20 točk)** Za jumske jezike so bila sestavljena še nekatera druga drevesa z naslednjimi leksikostatističnimi razdaljami v korenu drevesa (na skrajni levi strani vsakega drevesa):

1. 0,20
2. 0,23
3. 0,24

Narišite vsako od teh dreves. Za vsako drevo opredelite, ali je izdelano ročno ali avtomatsko, in ali je bilo generirano z algoritmom A ali B.

(H) **(3 točke)** Dve razdalji, navedeni v izzivu **(G)**, sta zaokroženi na dve decimalni mesti: 0,23 je nastalo z zaokrožitvijo 0,225. Katera druga razdalja je bila zaokrožena in kakšna je njena točna vrednost?

(I) **(4 točke)** Pojasnite, kako se izračunajo indeksi stabilnosti.

(J) **(5 točk)** Pojasnite, kako se izračunajo leksikostatistične razdalje.

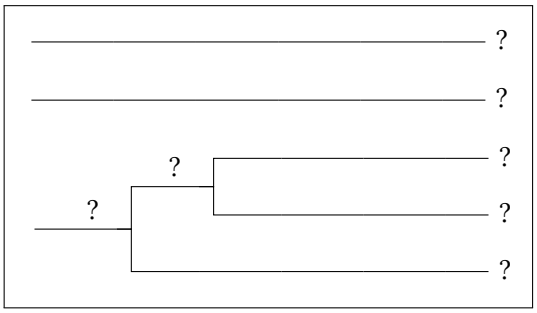
(K) **(4 točke)** Pojasnite razliko med algoritmoma A in B.

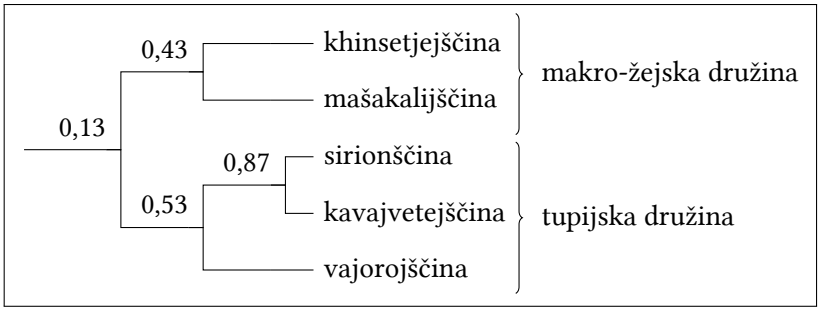
VI. sklop. Makro-žejska družina in tupijska družina (Brazilija, Bolivija)

(L) (28 točk) Makro-žejska in tupijska družina sta pomembni jezikovni družini v Južni Ameriki. Nekateri jezikoslovci menijo, da sta v daljnem sorodstvu. Preglejte naslednje sezname besed.

	A	B	Γ	Δ	E
lubje	e=ˈe-ke	hʷĩ=ˈkʰʌ	kʉp=ˈpɛ	mĩβ͡m=ˈt͡ɕaj	=ˈpɛ
trebuh	ˈe=ɾje	=ˈtʲʰigi	=ãˈɯ̃n	=ˈtæj	=ɾɛˈwɛk
kri	e=ˈɾuki	=kaˈⁿbɭo	=d͡ʒ=aˈʉ	=ˈhɛβ͡p	=ɾuˈɨ
žgati	=ˈɾa͡i	=ɾɔ=ˈkʰɹɜ̃	=poˈkʷa	mũ=...=ˈha͡βp	=ɾaˈpɨ
maščoba	e=ˈkiɾa	=ˈtʲʰwəmɨ	=ˈd͡ʒ=ap	=ˈtua͡βp	=ˈkap
noga	ˈe=i	=ˈhʷaji	=ˈβi	=pɒˈta	=ˈpɨ
roka	ˈe=o	=ɲɨ̃ˈkʰɹa	=ˈβo	=ˈɲĩ͡βm	=ˈpɔ
težek	e=ˈusi	=wɨˈtʲʰĩ	=poˈti	=β͡pˈtəj	=pɔˈij
jetra	ˈe=ja	=ˈⁿba	=piˈa	=t͡ɕɨβ͡pkɨ̃ˈnãj	=pɨˈʔa
nov	e=ˈjasu	=ˈⁿdɨwɨ	=paˈgop	=ˈtɨβ͡p	=pɨaˈu
korenina	e=ˈɾao	=jaˈɭe	kʉp=kʉjoˈpɛ	mĩβ͡m=ɲĩβ͡m=t͡ɕaˈti͡ə	=ɾaˈpɔ
koža	ˈe=ɨ	=ˈkʰʌ	=ˈpɛ	=ˈt͡ɕaj	=ˈpit
rep	e=ˈɾoko͡i	=ˈⁿbɨ	=d͡ʒ=oˈkʷaj	=nãː=ˈkɨβ͡p	=ˈɾaj
bel	ˈe=ʃɨ̃	=jaˈkʰa	=d͡ʒiˈɾa	=β͡pˈdoɯ̨	=ˈsĩŋ
krilo	e=ˈheo	=jaˈɭa	=pɛˈo	=ɲĩˈmãɯ̨	=pɛˈpɔ, =jɨˈwa

Spodaj si lahko ogledate dve drevesi, ki sta bili sestavljeni na podlagi istih seznamov. Manjkajo nekateri podatki (imena jezikov in leksikostatistične razdalje). Zapolnite prazna mesta. Za vsako drevo opredelite, ali je izdelano ročno ali avtomatsko, in ali je bilo generirano z algoritmom A ali B.





A	B	Γ	Δ	E
?	?	?	?	?

⚠ Pri tem izzivu so bile namerno izpuščene ročno pripravljene oznake in indeksi stabilnosti.

(M) (10 točk) Avtomatizirani postopki, ki temeljijo na razredih Dolgopoljskega, lahko dajo napačne rezultate. V tem primeru avtomatski postopek odkrije več podobnosti med sirionščino in določenim makro-žejskim jezikom (khinsetjejščino) kot med sirionščino in drugimi tupijskimi jeziki. Predlagajte spremenjen avtomatiziran postopek, ki bo dal pravilno klasifikacijo, če se uporabi za zgornja makro-žejska in tupijska seznama besed, in ga *na kratko* opišite.

⚠ Ta izziv bo ocenjen le v primeru enakega števila točk med ekipami, ki so dosegle najvišje število točk.

Avtorji se zahvaljujejo Alejandri Vidal, Mariji Konošenko, Ilji Gruntovu in Jamthu Suyu za odgovore na vprašanja o določenih jezikih. —*Andrej Nikulin, Milena Veneva*

Srečno!