Usage of Stem Editor for expressing Arguments for the Serbian Language
Training experts
(read after Stemmer Builder suite Manual)

C:\stemSuiteProgs> java -jar StemEditorV4.jar
Username: test
Password: *none*

ID	Word	Translation	Stem	first	second
1	abara	NA	ab		
2	abari	NA	ab		
3	abarski	NA	abar		
4	abarsko	NA	abars		
5	aboridžina	NA	aboridž	CS,2,aboridžin	
6	aboridžinima	NA	aboridžin	CS,2,aboridžin	
7	administrativne	NA	administrativ	DS,3	CS,1
8	administrativni	NA	administrat	DS,3	CS,1
9	admirala	NA	admira	DS,3	CS,2
10	adolescentski	NA	adolescent		
11	adresa	NA	adr	CS,4,adres	
12	adrese	NA	adres	CS,4,adres	
13	advokat	NA	advok		
14	aerobni	NA	aerob	DS,5	CS,1
15	aerodrom	NA	aerodr	DS,5	CS,2
16	aerodromu	NA	aerodr	DS,5	CS,2
17	afp	NA	afp		
18	afričkim	NA	afrič		
19	agencija	NA	ag	CS,6,agenc	
20	agencije	NA	agenc	CS,6,agenc	
21	agenciji	NA	agenc	CS,6,agenc	
22	agresije	NA	agres	CS,7,agres	
23	agresivnu	NA	agresiv	CS,7,agres	
24	agro	NA	ag	CS,9,agro	
25	agronoma	NA	agronom	CS,9,agro	
26	ah	NA	ah		
27	ajde	NA	ajd		
28	akademije	NA	akadem	CS,10,akadem	
29	akademiju	NA	akadem	CS,10,akadem	
30	akademik	NA	akadem	CS,10,akadem	
31	akademskih	NA	akademsk	CS,10,akadem	
32	akcenta	NA	akcen	DS,11	CS,1
33	akcija	NA	ak	DS,11	CS,2
34	akcijama	NA	akcijam	DS,11	CS,2
35	akcije	NA	akc	DS,11	CS,2
36	akciji	NA	akc	DS,11	CS,2
37	akcijom	NA	akc	DS,11	CS,2
38	akciju	NA	akc	DS,11	CS,2
39	aklamacijom	NA	aklamac		
40	aktera	NA	akte	CS,12,akter	
41	aktere	NA	akter	CS,12,akter	

CS and DS/CS arguments

For the set of words with IDs 5…6 there are two candidates: "aboridž" and "aboridžin". According to the rule defined in article A13 (highlighted text at end of page 320) the user (expert) select the second candidate stem. The same applies for the set of words with IDs 11…12 and the expert select the stem "adres". Once again, the same applies for the set of words with IDs 19…21 and the expert select the stem "agenc".

In case of the set of words with IDs 22…23 there are two candidates: "agres" and "agresiv". Only the first exist in both words and consequently the expert select the first candidate stem "agres". This is because, the sub-rule "exist in every word" has priority against the rule "the longest".

In case of the set of words with IDs 24…25 there are two candidates: "ag" and "agronom". According to the rule defined in article A13 (highlighted text at end of page 320) the user (expert) should select "ag". But it has the possibility to over-conflate (assign the same stem for words with different meaning). Thus, the expert makes a declension (more than the declensions presented in Appendix 3 of paper A13) and suggests the stem "agro" which is no one of the alternative stems provided by the primary stemmer.

ID	Word	Translation	Stem	first	second
42	aktivirati	NA	aktivira	CS,13,aktiv	
43	aktivisti	NA	aktivis	CS,13,aktiv	
44	aktivno	NA	aktiv	CS,13,aktiv	
45	aktivnog	NA	aktivn	CS,13,aktiv	
46	aktivnost	NA	aktiv	CS,13,aktiv	
47	aktivnosti	NA	aktivnos	CS,13,aktiv	
48	aktivnostima	NA	aktivnost	CS,13,aktiv	
49	aktivnošću	NA	aktivnošć	CS,13,aktiv	
50	aktu	NA	akt	DS,14	CS,2
51	aktualnoj	NA	aktualn	DS,14	CS,3
52	akumulacije	NA	akumulac	DS,14	CS,4
53	akustičnoj	NA	akustičn	DS,14	CS,5
54	akutni	NA	akut	DS,14	CS,1
55	akutnih	NA	akutn	DS,14	CS,1
56	al-kaide	NA	al-kaid		
57	al-mulk	NA	al-mul		
58	alarm	NA	alar		
59	alate	NA	alat		
60	albanaca	NA	albana	CS,15,alban	
61	albanci	NA	albanc	CS,15,alban	
62	albanske	NA	albansk	CS,15,alban	
63	albanskih	NA	albansk	CS,15,alban	
64	album	NA	albu	CS,16,album	
65	albumu	NA	album	CS,16,album	
66	aleksandar	NA	aleksand	DS,18	CS,1
67	aleksandra	NA	aleksand	DS,18	CS,1
68	aleksandrom	NA	aleksandr	DS,18	CS,1
69	aleksandru	NA	aleksandr	DS,18	CS,1
70	alekseja	NA	alekse	DS,18	CS,1
71	aleksić	NA	aleksić	DS,18	CS,2
72	alen	NA	al		
73	alfred	NA	alfred	CS,19,alfred	
74	alfreda	NA	alfre	CS,19,alfred	
75	alhemičar	NA	alhemič		
76	alki	NA	alk		
77	alkohol	NA	alkoh	CS,20,alkohol	
78	alkohola	NA	alkoho	CS,20,alkohol	
79	alkoholu	NA	alkohol	CS,20,alkohol	
80	alo	NA	al		
81	aloja	NA	alo	CS,21,aloj	
82	aloje	NA	alo	CS,21,aloj	

CS and DS/CS arguments

For the set of words with IDs 42…49 there are seven candidates: "aktivira", "aktivis", "aktiv", "aktivn", "aktivnos", "aktivnost" and "aktivnošć". Only the candidate "aktiv" exist in every word. It is also the most frequent (2/8). Consequently the expert selects the candidate stem "aktiv".

For the set of words with IDs 60…63 there are three candidates: "albana", "albanc" and "albansk". No one candidate exists in every word. Thus, the expert makes a declension (more than the declensions presented in Appendix 3 of paper A13) and suggests the stem "alban" which is no one of the alternative stems provided by the primary stemmer. This decision of expert is explained because "alban" exist in every word and does not seem to have the possibility to over-conflate (merge with the nearby – neighbor – words).

For the set of words with IDs 81…82 there only one candidate: "alo". However, the expert is afraid that this stem is possible to over-conflate. Consequently the expert makes a declension to the rule (more than the declensions presented in Appendix 3 of paper A13) and suggests the stem "aloj" which exist in every word of the set and it is longer than the one ("alo") provided by the primary stemmer.

ID	Word	Translation	Stem	first	second
83	alpa	NA	alp	CS,22,alp	
84	alpe	NA	alp	CS,22,alp	
85	alpi	NA	alp	CS,22,alp	
86	alpskih	NA	alpsk	CS,22,alp	
87	aluminijum	NA	aluminiju		
88	alžira	NA	alž		
89	ama	NA	am		
90	ambasada	NA	ambas	CS,24,ambasad	
91	ambasade	NA	ambasad	CS,24,ambasad	
92	ambasadi	NA	ambasa	CS,24,ambasad	
93	ambasadu	NA	ambasa	CS,24,ambasad	
94	ambicije	NA	ambic	CS,25,ambic	
95	ambiciozni	NA	ambic	CS,25,ambic	
96	amenofisa	NA	amenof		
97	americi	NA	americ	CS,26,amerik	
98	amerika	NA	amer	CS,26,amerik	
99	amerikanizovani	NA	amerikanizov	CS,26,amerik	
100	amerikanka	NA	amerik	CS,26,amerik	
101	amerike	NA	amerik	CS,26,amerik	
102	američke	NA	američk	CS,26,amerik	
103	američki	NA	američk	CS,26,amerik	
104	američkih	NA	američk	CS,26,amerik	
105	američkim	NA	američk	CS,26,amerik	
106	američkom	NA	američk	CS,26,amerik	
107	amonijak	NA	amoni		
108	amsterdama	NA	amsterdam		
109	anabaptist	NA	anabapt	CS,27,anabaptis	
110	anabaptista	NA	anabapt	CS,27,anabaptis	
111	anabaptisti	NA	anabaptis	CS,27,anabaptis	
112	anabaptistovu	NA	anabaptistov	CS,27,anabaptis	
113	analiza	NA	anal	CS,28,analiz	
114	analizama	NA	analizam	CS,28,analiz	
115	analizira	NA	anal	CS,28,analiz	
116	anbar	NA	anb		
117	andaluziji	NA	andaluz		
118	ane	NA	an		
119	angažman	NA	angažm	CS,29,angaž	
120	angažuje	NA	angažu	CS,29,angaž	
121	angelom	NA	angel		
122	anibal	NA	anib		
123	aničiće	NA	aničić	CS,30,aničić	
124	aničići	NA	aničić	CS,30,aničić	
125	anksioznost	NA	anksioz	CS,31,anksioznos	
126	anksioznosti	NA	anksioznos	CS,31,anksioznos	

Only CS arguments

ID	Word	Translation	Stem	first	second
127	anonimni	NA	anonim		
128	antibioticima	NA	antibiotic		
129	antiboja	NA	antibo		
130	antihrist	NA	antihr		
131	antike	NA	antik		
132	antipušačke	NA	antipušačk		
133	antisemitizmom	NA	antisemitizm	CS,32,antisemit	
134	antisemitski	NA	antisemit	CS,32,antisemit	
135	antisemitskom	NA	antisemitsk	CS,32,antisemit	
136	antiterorističke	NA	antiterorističk		
137	antoneti	NA	anton	CS,33,antonet	
138	antonetija	NA	antonet	CS,33,antonet	
139	antropologije	NA	antropolog		
140	antunović	NA	antunović		
141	aosti	NA	aos		
142	aparat	NA	apar	DS,35	CS,1
143	aparatom	NA	aparat	DS,35	CS,1
144	apartman	NA	apartm	DS,35	CS,2
145	apartmana	NA	apartm	DS,35	CS,2
146	apartmanima	NA	apartman	DS,35	CS,2
147	apelaciono	NA	apelacio		
148	apeluje	NA	apelu		
149	apeninskom	NA	apeninsk		
150	apetita	NA	apeti		
151	apija	NA	ap		
152	aplauze	NA	aplau	CS,36,aplauz	
153	aplauzi	NA	aplau	CS,36,aplauz	
154	aplauzom	NA	aplauz	CS,36,aplauz	
155	apokalipsu	NA	apokalips		
156	apotekara	NA	apote		
157	aprila	NA	apri	CS,37,april	
158	aprilu	NA	april	CS,37,april	
159	apsolutno	NA	apsolut	CS,38,apsolut	
160	apsolutnu	NA	apsolut	CS,38,apsolut	
161	arabeski	NA	arabe		
162	aranžmana	NA	aranžm		
163	arapska	NA	arap	CS,39,arapsk	
164	arapske	NA	arapsk	CS,39,arapsk	
165	arenu	NA	ar		
166	argumente	NA	argument	CS,40,argument	
167	argumenti	NA	argumen	CS,40,argument	
168	argumentima	NA	argument	CS,40,argument	

CS and DS/CS arguments

ID	Word	Translation	Stem	first	second
169	arhitekata	NA	arhitek	CS,41,arhitek	
170	arhitekta	NA	arhitek	CS,41,arhitek	
171	arhitektonskom	NA	arhitektonsk	CS,41,arhitek	
172	arhitekture	NA	arhitektur	CS,41,arhitek	
173	arhitekturi	NA	arhitektur	CS,41,arhitek	
174	arhitekturu	NA	arhitektur	CS,41,arhitek	
175	arhiva	NA	arhi		
176	arijeh	NA	arije		
177	aristotelom	NA	aristotel		
178	armani	NA	arm	CS,42,arman	
179	armanija	NA	arm	CS,42,arman	
180	armanijem	NA	arman	CS,42,arman	
181	armanijeva	NA	arman	CS,42,arman	
182	armanijeve	NA	armanijev	CS,42,arman	
183	armanijevo	NA	armanij	CS,42,arman	
184	arsen	NA	ars		
185	art	NA	art		
186	artikal	NA	artik		
187	asfaltu	NA	asfalt		
188	asistenta	NA	asisten		
189	asocijacija	NA	asoc		
190	atanasković	NA	atanasković		
191	atarima	NA	atar	CS,43,atar	
192	atarski	NA	atar	CS,43,atar	
193	ataru	NA	atar	CS,43,atar	
194	ateh	NA	ate		
195	ateistkinja	NA	ateist		
196	ateljeima	NA	atelje		
197	atenice	NA	aten		
198	atentat	NA	atent	CS,44,atentat	
199	atentata	NA	atent	CS,44,atentat	
200	atentatore	NA	atentator	CS,44,atentat	
201	atentatu	NA	atentat	CS,44,atentat	
202	atini	NA	ati		
203	atlanti	NA	atlan		
204	atmosfera	NA	atmosfe	CS,45,atmosfer	
205	atmosfere	NA	atmosfer	CS,45,atmosfer	
206	atmosferu	NA	atmosfer	CS,45,atmosfer	
207	audi	NA	au		
208	australijanska	NA	australijan	DS,46	CS,1
209	australije	NA	austral	DS,46	CS,1
210	australiji	NA	austral	DS,46	CS,1
211	australiju	NA	austral	DS,46	CS,1
212	austrija	NA	austr	DS,46	CS,2
213	austrije	NA	austr	DS,46	CS,2

CS and DS/CS arguments

ID	Word	Translation	Stem	first	second
214	autentična	NA	autentič		
215	auto	NA	aut	DS,47	CS,3
220	automobil	NA	automobil	DS,47	CS,3
222	automobila	NA	automobi	DS,47	CS,3
223	automobile	NA	automobi	DS,47	CS,3
224	automobili	NA	automobi	DS,47	CS,3
225	automobilom	NA	automobil	DS,47	CS,3
216	autobusa	NA	autob	DS,47	CS,1
217	autobuse	NA	autobus	DS,47	CS,1
226	autoputa	NA	autop	DS,47	CS,2
218	autoimuno	NA	autoimu		
219	automatizovane	NA	automatizova		
221	automobil-bomba	NA	automobil-bom		
227	autor	NA	au		
228	avangarde	NA	avangard	DS,49	CS,1
229	avangardni	NA	avangar	DS,49	CS,1
230	avantura	NA	avan	DS,49	CS,2
231	avanturi	NA	avantur	DS,49	CS,2
232	avanturu	NA	avantur	DS,49	CS,2
233	aveniji	NA	aven		
234	avet	NA	av		
235	avganistan	NA	avganist		
236	avgustu	NA	avgust		
237	avijatičar	NA	avijatič		
238	avio-kompanije	NA	avio-kompan		
239	avion	NA	avi	CS,50,avion	
240	aviona	NA	avi	CS,50,avion	
241	avioni	NA	av	CS,50,avion	
242	avionima	NA	avion	CS,50,avion	
243	avionu	NA	avio	CS,50,avion	
244	azrine	NA	azri		
245	azurno	NA	azur		
246	aćimovića	NA	aćimović		
247	aždahu	NA	ažda		
248	b92	NA	b92		
249	ba	NA	ba		
250	babakov	NA	babak		
251	bacamo	NA	baca	DS,52	CS,1
252	bacan	NA	bac	DS,52	CS,1
253	baci	NA	bac	DS,52	CS,1
256	bacio	NA	baci	DS,52	CS,1
257	bacite	NA	bacit	DS,52	CS,1
254	bacil	NA	bacil	DS,52	CS,2
255	bacili	NA	baci	DS,52	CS,2

CS and DS/CS arguments

Words with IDs in the range 215 … 226, excluding 218, 219 and 221, are rearranged to define CS subsets (CS1, CS2 and CS3) for the dominating DS47.

Words with IDs 218, 219, 221 are moved lower, outside the DS47 arguments.

Words with IDs 251 … 257 are rearranged to define CS subsets (CS1 and CS2) for the dominating DS52.

ID	Word	Translation	Stem	first	second
258	bagdad	NA	bagd	CS,53,bagda	
259	bagdadu	NA	bagda	CS,53,bagda	
260	bagdalom	NA	bagdal	CS,53,bagda	
261	bajagine	NA	bajagi	DS,54	CS,1
262	bajke	NA	bajk	DS,54	CS,2
263	bajki	NA	bajk	DS,54	CS,2
264	bajkoviti	NA	bajkovi	DS,54	CS,2
268	bajčice	NA	bajč	DS,54	CS,2
265	bajman	NA	bajm	DS,54	CS,3
266	bajonet	NA	bajon	DS,54	CS,3
267	bajonete	NA	bajon	DS,54	CS,3
269	bakica	NA	bak		
270	bakterija	NA	bakt	CS,55,bakter	
271	bakterijom	NA	bakter	CS,55,bakter	
272	bala	NA	ba	DS,56	CS,1
273	balans	NA	balans	DS,56	CS,2
274	balasta	NA	balas	DS,56	CS,3
275	balensijaga	NA	balens	DS,56	CS,4
276	balensijage	NA	balensijag	DS,56	CS,4
277	balensijagina	NA	balensijag	DS,56	CS,4
278	balensijaginog	NA	balensijagin	DS,56	CS,4
279	bali	NA	ba	DS,58	CS,1
280	balijama	NA	balijam	DS,58	CS,2
281	balističkim	NA	balističk	DS,58	CS,3
282	balkana	NA	balk	CS,57,balkan	
283	balkanska	NA	balkan	CS,57,balkan	
284	balkanu	NA	balka	CS,57,balkan	
285	balmen	NA	balm	CS,59,balmen	
286	balmenov	NA	balmen	CS,59,balmen	
287	balon	NA	bal	CS,60,balon	
288	balona	NA	bal	CS,60,balon	
289	balone	NA	balo	CS,60,balon	
290	balonom	NA	balon	CS,60,balon	
291	balonu	NA	balo	CS,60,balon	
292	balonče	NA	balonč	CS,60,balon	
293	balova	NA	bal	CS,62,bal	
294	balske	NA	balsk	CS,62,bal	
295	banalno	NA	banal	DS,63	CS,1
296	band	NA	band	DS,63	CS,2
297	banjaluci	NA	banjaluc	DS,63	CS,3
298	banji	NA	ba	DS,63	CS,4

CS and DS/CS arguments

Words with IDs in the range 261 … 267 are rearranged to define CS subsets (CS1, CS2 and CS3) for the dominating DS54.

The arguments for the words with IDs 282 … 294 are simple CS arguments (and not DS with CS subsets) because there is not possibility (or it is very low the possibility) for over conflating (merging the words with a common stem).

January 23, 2020,
Nikitas N. Karanikolas.

--
Professor,
Dept. of Informatics and Computer Engineering,
University of West Attica,
Tel. +30-210-5385737,
Fax +30-210-5910975,
(Uni.W.A. web page) http://www.ice.uniwa.gr/emd_person/17335/
(personal web page) http://users.uniwa.gr/nnk/
(dblp web page) http://dblp.uni-trier.de/db/indices/a-tree/k/Karanikolas:Nikitas_N=.html
(scopus web page) http://www.scopus.com/authid/detail.url?authorId=6507018215