

**Primož Jakopin**

Inštitut za slovenski jezik Frana Ramovša ZRC SAZU

## **Slovar sodobnega slovenskega jezika in Wikislovar**

V prispevku je predstavljenih več vidikov naloge, kako povezati bodoči Slovar sodobnega slovenskega jezika (SSSJ) tako z angleškim wikipedijinim slovarjem (530.000 angleških gesel, 2.700.000 vseh gesel) Wiktionaryjem kot tudi s slovenskim Wikislovarjem (3830 slovenskih gesel). V zaključku je predlagana vključitev spletne zbirke *Besede slovenskega jezika* ter dela SSSJ s pogosteje rabljenimi gesli v oba omenjena slovarja.

### **The New Dictionary of Modern Slovenian Language and the Wikipedia's Wiktionary**

The modalities of how to connect the proposed new Dictionary of Modern Slovenian Language (DMSL) to the Wikipedia's main Wiktionary of English language and to its Slovenian sibling, Wikislovar, are briefly discussed. The inclusion of the web list *Slovenian words* and the more frequent part of DMSL into both Wiktionary and Wikislovar is proposed in the conclusion.

**Ključne besede:** Slovar sodobnega slovenskega jezika, Wikislovar, struktura slovarja, enojezični slovarji, večjezični slovarji

**Keywords:** Dictionary of Modern Slovenian Language, Wiktionary, dictionary entry structure, monolingual dictionaries, multilingual dictionaries

## **1 Uvod**

Z nastopom interneta pred dvajsetimi leti, še posebej pa z njegovo širšo uveljavitvijo pred desetimi leti, se je način uporabe slovarjev, tako večjezičnih kot enojezičnih, močno spremenil. Strojno poizvedovanje po slovarjih, po njihovi celotni vsebini, ne le po geslih, pa še ob hkratni možnosti preverjanja rabe alternativnih možnosti s svetovnim spletnim iskalnikom, postaja vse bolj samoumevno. In ne le na klasičnih podmiznih računalnikih na delovnem mestu in doma, ampak tudi na prenosnikih, ki se s kar nekaj kilogrami skupne teže nikoli niso mogli prav znebiti

težkoatletskega pridiha, ampak v zadnjem letu že tudi na pol kilogramskih tablicah z brezžično wifi in mobilnotelefonsko povezavo. Pri slednjih šele lahko mirno rečemo: slovarji in splet, vedno in povsod.

Skupaj s širjenjem teh še pred kratkim znanstvenofantastičnih možnosti do prav vsakega pišočega so se zelo povečala pričakovanja uporabnikov po brezplačnih rešitvah. Če brezplačne strojne opreme in tiskanih slovarjev ni nikoli nihče pričakoval, veljajo druga merila za programsko opremo in za elektronske zbirke. S širjenjem dostopnosti spletnih izobraževalnih vsebin, v navezavi s svetovnim iskalnikom, kakršen je Google, se je pojavila potreba po sicer skromnejših, a še vseeno uporabnih enciklopedičnih in slovarskih rešitvah.

## 2 Wiktionary in Wikislovar

Kot odgovor na ta vprašanja je prišlo leta 2001 do izvedbe wikipedije, splošne enciklopedije brez uredniškega odbora, ki jo lahko ureja vsakdo, tudi če ni registriran kot uporabnik, a hrani zgodovino vseh sprememb in je tako mogoča sledljivost vseh urejanj posameznih člankov. Tako je mogoče hitro vrniti v prejšnje stanje članke, ki so bili vandalizirani, to je deležni neprimernih popravkov. Avtorji, ki so se izkazali in potrdili s svojim delom, lahko dobijo status administratorjev. Ti imajo večje pravice od navadnih avtorjev, lahko med drugim zaščitijo gesla pred urejanjem, onemogočijo urejanje uporabnikom ali internetnim naslovom, članke lahko odstranijo (in imajo edini dostop do odstranjenih člankov) ali jim spremenijo naslov. Angleška wikipedija ima 4,500.000 člankov in 1.419 administratorjev, slovenska, ki nastaja od leta 2002, pa 140.000 člankov in 28 administratorjev.

Kmalu za wikipedijo so se v istem okviru pojavili še sestrski projekti in med njimi je za nas najzanimivejši wikislovar (angleško wiktionary), prostodostopen in prosto urejevalen slovar.

1. latinski	646,560	26. galicijski	19,561	51. albanski	4,729	76. retoromanski	2,116
2. angleški	528,622	27. kantonski	18,615	52. luksemburški	4,341	77. škotski (Skots)	2,110
3. italijanski	502,952	28. turški	17,925	53. staroangleški	4,289	78. slovaški	2,109
4. francoski	283,800	29. korejski	16,288	54. ferski	4,276	79. starosaksonski	2,028
5. španski	257,622	30. škotsko gelski	13,885	55. venetski	4,131	80. urdu	1,971
6. nemški	155,646	31. gruzijski	13,857	56. arabski	3,950	81. malteški	1,921
7. portugalski	151,038	32. irski	13,616	57. estonski	3,872	82. jidiš	1,913
8. finski	113,034	33. ido	13,014	<b>58. slovenski</b>	<b>3,830</b>	83. interlingva	1,877
9. esperanto	106,762	34. romunski	12,197	59. volapük	3,552	84. bengalski	1,860
10. latvijski	100,249	35. islandski	11,135	60. hindi	3,496	85. indonezijski	1,859

11. švedski	90,960	36. armenski	10,361	61. navajski	3,442	86. klasično azteški	1,819
12. mandarinski	67,646	37. gotski	10,026	62. malgaški	3,327	87. sanskrt	1,755
13. nizozemski	65,189	38. asturijski	9,704	63. lojban	3,179	88. aramejski	1,684
14. japonski	60,826	39. teluški	9,257	64. malajski	3,176	89. azerski	1,614
15. katalonski	60,741	40. vietnamski	8,461	65. srednjefrancoski	2,999	90. srednjeangleški	1,544
16. srbohrvaški	46,911	41. perzijski	6,809	66. min nan (kitajski)	2,611	91. staro irski	1,544
17. translingual	45,800	42. manski	6,616	67. svahili	2,604	92. ukrajinski	1,540
18. bolgarski	37,101	43. novonorveški	6,403	68. tajski	2,546	93. okcitanski	1,489
19. poljski	33,598	44. starogrški	6,142	69. afrikanski	2,508	94. hiligajnonski	1,319
20. madžarski	31,586	45. staroarmenski	6,064	70. valižanski	2,323	95. staronordijski	1,292
21. grški	28,764	46. norveški	6,012	71. klasični sirski	2,310	96. aromunski	1,269
22. danski	24,944	47. hebrejski	5,713	72. starocerkenoslovanski	2,277	97. vilamovski	1,195
23. ruski	22,421	48. starofrancoski	5,697	73. adigejski	2,265	98. sicilijanski	1,193
24. češki	21,584	49. ladinski	5,576	74. protogermanski	2,257	99. baskovski	1,186
25. litovski	20,817	50. kurdski	4,914	75. krimskotatarski	2,145	100. tagaloški	1,179

Tabela 1: Najmočnejši wikislovarji po jezikih, s številom gesel, februar 2014

Njegov namen je vsem internetnim uporabnikom ponuditi prost, vsebinsko odprt slovar angleškega jezika. Slovar je napisan v angleščini, njegov končni cilj pa je zajeti vse besede vseh jezikov (Desroches in Roth 2011), tudi slovenskega. Angleškemu wikislovarju so kot pri wikipediji sledili taki slovarji še v drugih jezikih in do danes nastali za 1.214 jezikov (tabela 1). Resnejši so trije, za latinski, angleški in italijanski jezik, še sedem jih ima več kot 100.000 gesel. V angleškem je približno 220.000 gesel s preprostimi definicijami tipa “form of”, npr. za pridevnik koščičast:

**stone-fruit**, adj., 1. attributive form of **stone fruit**, noun

*It had a **stone-fruit** taste to it: it tasted something like a nectarine, actually.*

Tako **je** še precej skromnejši od glavnih tiskanih enojezičnih slovarjev angleškega jezika, kakršna sta npr. *Oxford English Dictionary* s 615.000 gesli in *Merriam-Webster's Third New International Dictionary of the English Language, Unabridged* s 475.000 gesli in še mnogo podgesli. Angleški wiktioary je bil deležen tudi precej kritičnih besed; primer je ocena, ki mu jo je namenila profesorica zgodovine s harvardske univerze Jill Lepore:

*Pri angleškem wikislovarju niso razvidni avtorji. Celo uredniškega odbora ni. “Bodi sam svoj slovaropisec” je geslo dneva. Kdo potrebuje strokovnjake? Zakaj bi pošteno plačali slovar, ki so ga napisali leksikografi, če ga lahko sami zmečemo skupaj? Za wikislovar bi težko rekli, da je republikanski ali demokratski, bolj je maoističen. In dober je prav toliko, kot so dobri stari slovarji s potečenimi avtorskimi pravicami, iz katerih je napaberkovan.*

(V izvirniku: *There's no show of hands at Wiktionary. There's not even an editorial staff. "Be your own lexicographer!", might be Wiktionary's motto. Who needs experts? Why pay good money for a dictionary written by lexicographers when we could cobble one together ourselves? Wiktionary isn't so much republican or democratic as Maoist. And it's only as good as the copyright-expired books from which it pilfers.*) (Lepore 2006: 78, prev. P. J.)

Članek avtorice je sicer pel hvalo Noahu Websterju in njegovemu prvemu slovarju iz leta 1825. Oba omenjena tiskana slovarja sta v celoti tudi na spletu, s plačljivim dostopom, po izboru 50.000 gesel iz Merriam-Websterjevega slovarja pa je mogoče poizvedovati tudi brezplačno.

Vendar se angleški wikislovar počasi, a vztrajno postavlja na noge in v dogledni prihodnosti se zna močno približati svojim tiskanim vzornikom in jih celo preseči.

Slovenski wikislovar je seveda veliko skromnejši in primerjave s tiskanim slovarjem, kakršen je *Slovar slovenskega knjižnega jezika*, še niti ne zasluži. Kaj se je dogajalo z njim, lepo vidimo iz tabele 2; števila v stolpcih se tu ne nanašajo le na gesla opazovanega jezika, kot v tabeli 1, ampak na vsa gesla in so zato številke višje, podatki pa so zajeti vedno na koncu januarja tekočega leta. Opazimo, da se je slovenski wikislovar močno okreplil v letih 2005 in 2006, potem pa nekako ni šlo več naprej.

Leto	SI	SH	EN
2002	-	-	2
2003	-	-	864
2004	1	1	31.000
2005	3	418	51.000
2006	4.900	516	112.000
2007	7.000	688	320.000
2008	7.000	2.700	648.000
2009	7.100	2.900	1.100.000
2010	7.100	6.500	1.600.000
2011	7.100	17.000	2.300.000
2012	7.100	21.000	2.100.000
2013	7.100	25.000	2.600.000
2014	7.300	26.000	2.700.000

Tabela 2: Razvoj wikislovarjev za slovenski, srbohrvaški in angleški jezik po letih

To dejstvo postane še bolj očitno, ko pogledamo, kdo so glavni avtorji in kdaj so delovali.

Avtor	Začetek dela	Konec dela	Št. urejanj
1. Erreka	18. avg. 2005	25. apr. 2006	22.049
2. Interwicket	28. jan. 2009	8. nov. 2010	12.727
3. Klemen Kocjancic	23. jan. 2006	18. nov. 2006	7.519
4. Andrejj	18. avg. 2005	29. jun. 2013	531
5. FJJ	11. feb. 2006	10. sep. 2007	206
6. Zverec	24. jul. 2013	9. avg. 2013	157
7. Opiaterein	31. okt. 2007	5. dec. 2009	122
8. Rumpel77	1. mar. 2013	15. apr. 2013	68
9. Ciciban	4. jul. 2007	28. jun. 2012	67
10. Eleassar	3. avg. 2005	2. dec. 2007	58
11. Zuiarra	12. dec. 2010	6. dec. 2013	44
12. MZaplotnik	24. jan. 2009	15. apr. 2011	43
13. Merlin G.	16. feb. 2008	5. jan. 2010	39
14. Škilavost	17. feb. 2007	17. feb. 2007	36
15. Zoziuuz	22. apr. 2011	12. dec. 2011	36
17. Wikipedius Rex	20. apr. 2013	23. apr. 2013	33
18. Janezgrabrovsek	18. dec. 2012	26. dec. 2012	28
19. VPliousnine	19. dec. 2007	21. nov. 2009	27
20. Archangel	07. avg. 2007	31. jul. 2009	23

Tabela 3: Avtorji slovenskega wikislovarja, do konca januarja 2014

Erreka (reka v baskovskem jeziku) je psevdonim (v letih 2005 in 2006) študentke umetnostne zgodovine z znanjem večjega števila jezikov, predvsem romanskih, Interwicket je robotski urednik, za katerega skrbi angleško govoreči Robert Ullmann iz Nairobijskega; Klemen Kocjančič, Andrej Jakobčič in FJJ na naslednjih treh mestih pa so edini administratorji slovenskega wikislovarja (angleški jih ima 68). Erreka in Klemen sta opravila veliko delo, škoda le, da sta, kot bi se lahko reklo, plesala eno samo poletje in da nista imela pravih naslednikov.

# miška

Vsebina [\[prikaži\]](#)

## Samostalnik [\[uredi\]](#)

**miška** *f* (dv. miški, mn. miške)

## Izgovorjava [\[uredi\]](#)

[miˈʃka]

## Pomeni [\[uredi\]](#)

- majhna **miš**  
*Mačka je ujela miško in zdaj se igra z njo.*
- rač. majhna naprava za premikanje kazalca na zaslonu **računalnika**  
*Prenosnik lahko uporabljaš tudi brez miške.*
- eksp.* navadno mlajša, prikupna **ženska**
- ocvrto **pecivo** iz kvašenega testa

## Etimologija [\[uredi\]](#)

miš + ka

## Besedne zveze [\[uredi\]](#)

optična miška, brezžična miška

### Sopomenke:

### Protipomenke:

### Nadpomenke [\[uredi\]](#)

- [1] žival(i)ca  
[2] strojna oprema

### Podpomenke:

### Podobne in povezane besede:

### Rime [\[uredi\]](#)

hiška, piška, Šiška

### Anagrami [\[uredi\]](#)

mišak

### Prevodi [\[uredi\]](#)

- Angleško: [1, 2] **mouse** [\(en\)](#)
- Nemško: [1] **Mäuschen** *n* [\(de\)](#), [2] **Maus** *f* [\(de\)](#), **Computermaus** *f* [\(de\)](#)
- Rusko: **мышь** [\(ru\)](#) *f*

Kategorija: Slovenski samostalniki

Članek **miška** najdete tudi v [Wikipediji](#).



[2] Računalniška miška

**1. ženska sklanjatev**

Sklon/število	Ednina	Dvojina	Množina
Imenovalnik	miška	miški	miške
Rodilnik	miške	mišk	mišk
Dajalnik	miški	miškama	miškam
Tožilnik	miško	miški	miške
Mestnik	o miški	o miškah	o miškah
Orodnik	z miško	z miškama	z miškami

Tabela 4: Spletna podoba gesla **miška** v slovenskem wikislovarju (Miška-sl 2014)

Zanimiva je tudi leksikalna struktura gesel v wikislovarju – oglejmo si jo na slovenskem primeru gesla **miška**. V tabeli 4 vidimo za malenkost skrajšan zapis. Da tabela ne bi bila predolga, smo na podatkovno polje *vsebina* tik pod iztočnico kliknili na *skrij*. Odkrili bi ga seveda s klikom na *prikaži*. To polje je pomembno pri zelo dolgih geslih s podatki za več zaslonov in z njim lažje

najdemo tisti del zapisa, ki nas zanima. Opazimo tudi, da geslo ni dokončano – polja sopomenke, protipomenke, podpomenke ter podobne in povezane besede imajo samo naslove, vsebine pa še ne. V desnem stolpcu najdemo najprej podatek, da članek s to iztočnico obstaja tudi na wikipediji, s povezavo nanj, pod njim pa sliko drugega pomena iztočnice in vse sklanjatve. Veliko besed je zapisanih z modro barvo, kar pomeni, da skrivajo povezave, z rdečo pa, kjer bi povezave želeli, pa ustreznih člankov še ni. Vsi podatki gesla so zapisani v posebnem wikipedijinem urejevalniškem formatu:

```
==Samostalnik==
''miška'' ''f'' ([[dvojina|dv.]] miški, [[množina|mn.]] miške)

{{Samostalniska tabela s sliko (wikipedia)|
slika=3-Tastenmaus_Microsoft.jpg|velikost=250px|napis=[2] Računalniška miška
|sklanjatev=[[prva ženska sklanjatev|1. ženska]]
|im-ed=miška | im-dv=miški | im-mn=miške
|rod-ed=miške | rod-dv=mišk | rod-mn=mišk
|daj-ed=miški | daj-dv=miškama | daj-mn=miškam
|toz-ed=miško | toz-dv=miški | toz-mn=miške
|mest-ed=o miški | mest-dv=o miškah | mest-mn=o miškah
|or-ed=z miško | or-dv=z miškama | or-mn=z miškami}}

===Izgovorjava===
:[mɨʃka]

===Pomeni===
# majhna [[miš]]
#: ''Mačka je ujela miško in zdaj se igra z njo.''
# ''rač.'' majhna naprava za premikanje kazalca na zaslonu [[računalnik]]a
#: ''Prenosnik lahko uporabljaš tudi brez miške.''
# ''ekspr.'' navadno mlajša, prikupna [[ženska]]
# ocvrto [[pecivo]] iz kvašenega testa

===Etimologija===
[[miš]] + ''ka''

===Besedne zveze===
: optična miška, brezžična miška

{{Sopomenke}}

{{Protipomenke}}

===Nadpomenke===
:[1] žival(i)ca
:[2] strojna oprema

{{Podpomenke}}

{{Podobne besede}}

===Rime===
:[hiška], [[piška]], [[šiška]]

===Anagrami===
:[mišak]

===Prevodi===
* Angleško: [1, 2] {{Ustreznica|jezik=en|geslo=mouse}}
```

\* Nemško: [1] {{Ustreznica|jezik=de|geslo=Mäuschen|spol=n}}, [2] {{Ustreznica|jezik=de|geslo=Maus|spol=f}},  
 {{Ustreznica|jezik=de|geslo=Computermaus|spol=f}}  
 \* Rusko: {{prev|ru|мышь|f}}

[[Kategorija:Slovenski samostalniki]]  
 [[chr:miška]] [[en:miška]] [[eu:miška]] [[hu:miška]] [[lt:miška]] [[mg:miška]]

Tabela 5: Urejevalniška podoba gesla *miška* v slovenskem wikislovarju

Kot vidimo, zapis ni preveč zapleten, desni stolpec recimo ureja predloga *Samostalniska tabela s sliko (wikipedia)*, in bi lahko njegovo strukturo uporabili tudi pri pisanju novih gesel, dopolnjevanju starih ter pri izdelavi vmesnika za prenos gesel iz kakega drugače, npr. v formatu XML zapisanega slovarja.

Omenjeno je bilo že, da angleški Wiktionary ne vsebuje samo angleških gesel, ampak še veliko več tujih. Ker bi neslovenski govorec, ki bi naletel na besedo *miška*, zelo verjetno znal angleško, bi si skušal najbrž najprej pomagati prav z Wiktionaryjem. Izkaže se, da ne bi ostal praznih rok.

## miška

Slovene [\[edit\]](#)

**Noun** [\[edit\]](#)

**miška** *f* (genitive **miške**, nominative plural **miške**)

1. diminutive of mouse (animal)
2. mouse (computing: input device)
3. bird (usually young, pretty woman)
4. fried-dough fritter made of sweet yeast dough

Categories: [Slovene nouns](#) | [Slovene feminine nouns](#)

Tabela 6: Spletna podoba gesla *miška* v angleškem wikislovarju (Miška-en 2014)

Geslo je kratko, ima samo slovensko geslo, in tudi njegova urejevalniška podoba je taka:

```
==Slovene==
===Noun===
{{sl-noun|miška|f|miške|miške}}
# diminutive of [[mouse]] (animal)
# [[mouse]] (computing: input device)
# [[bird]] (usually young, pretty woman)
# fried-dough fritter made of sweet yeast dough
[[eu:miška]] [[lt:miška]] [[hu:miška]] [[mg:miška]] [[sl:miška]] [[chr:miška]]
```

Tabela 7: Urejevalniška podoba gesla *miška* v angleškem wikislovarju



Bolj zanimiva je zgodovina urejanj, kaj se je z geslom dogajalo. Desno zgoraj nad besedo *miška* kliknemo na ploščico *History* (Zgodovina) in dobimo tabelo 8.

- (cur | prev) 09:59, 1 March 2014 JakopinP (Talk | contribs) . . (307 bytes) (+14) . . (genitive and nominative plural have been added) (undo)
- (cur | prev) 11:44, 5 January 2014 Rukhabot (Talk | contribs) m . . (293 bytes) (+28) . . (interwikis: minor spacing; +chr:miška +mg:miška) (undo)
- (cur | prev) 13:34, 22 September 2013 MewBot (Talk | contribs) m . . (265 bytes) (-5) . . (sl-noun: head= to 1=) (undo)
- (cur | prev) 01:58, 22 September 2013 MewBot (Talk | contribs) m . . (270 bytes) (+12) . . (sl-noun: add redundant head=(temporary)) (undo)
- (cur | prev) 23:29, 21 January 2013 MewBot (Talk | contribs) m . . (258 bytes) (-7) . . (add sl-noun) (undo)
- (cur | prev) 13:31, 24 December 2012 Janez gabrovsek (Talk | contribs) . . (265 bytes) (+78) . . (→Noun) (undo | thank)
- (cur | prev) 06:20, 9 August 2012 Rukhabot (Talk | contribs) m . . (187 bytes) (+28) . . (interwikis: +eu:miška +lt:miška) (undo)
- (cur | prev) 07:56, 30 December 2011 MglovesfunBot (Talk | contribs) m . . (159 bytes) (0) . . (→Noun: bot: bypassing template redirects, minor spacing) (undo)
- (cur | prev) 14:59, 25 February 2011 Luckas-bot (Talk | contribs) m . . (159 bytes) (+14) . . (r2.7.1) (robot Adding: hu:miška) (undo)
- (cur | prev) 18:49, 14 December 2007 Tbot (Talk | contribs) . . (145 bytes) (+145) . . (created from translation table at mouse) (thank)

Tabela 8: Zgodovina urejanj gesla *miška* v angleškem wikislovarju

Iz tabele razberemo, da je geslo strojno generiral eden izmed wikislovarskih robotov, programov za opravljanje ponavljajočih se slovarskih postopkov, z enim samim (drugim) pomenom, 14. 12. 2007, na podlagi slovenskega prevoda drugega pomena v geslu *mouse*, dodanem 11. 11. 2007. Z geslom se ni potem nič kaj omembe vrednega dogajalo, predvsem so le roboti, prepoznamo jih iz niza *bot* na koncih uporabniških imen, dodajali podatke, da geslo *miška* obstaja oz. se je pojavilo še v šestih drugih wikislovarjih, in sicer baskovskem (eu), litvanskem (lt), madžarskem (hu), malgaškem (mg), slovenskem (sl) in wikislovarju jezika čeroki (chr). Prvi avtorski popravek se je zgodil 24. 12. 2012, ko je geslu še preostale tri pomene, prvega, tretjega in četrtega, dodal Janez Gabrovšek, in 1. 3. 2014, ko mu je pisec teh vrstic dodal obliki za roditelj ednine in imenovalnik množine. Zadnja sprememba je bila narejena predvsem iz estetskih razlogov, in ker pisanje gesel, kot že omenjeno, urejajo predloge (angl. *templates*), dela ni bilo veliko, le `{{sl-noun|miška|f}}` je bilo treba spremeniti v `{{sl-noun|miška|f|miške|miške}}`.

### 3 Zaključek

Kot je razvidno iz vsega navedenega, delo na wikislovarjih za izurjenega slovaropisca ni prav nič težko, ne zahteva posebne programske opreme, kar napišemo v urejevalniškem načinu, se da takoj preizkusiti, videti v lepi spletni obliki. Če bi imeli na razpolago še vmesnik med formatom XML, v katerem bo v vsakem primeru napisan novi Slovar sodobnega slovenskega jezika, in omenjenim urejevalniškim formatom – delu so v pomoč wikislovarske predloge (angl. templates) za posamezne besedne vrste –, bi se dalo brez prav veliko truda hitro pomagati wikislovarju do prave vrednosti. Odnos med wikislovarjem in siceršnjim slovaropisnim delom pa bi moral biti seveda zastavljen na daljši, ne le večletni, ampak večdesetletni rok. Slovaropisci, ki bi prevzeli to nalogo, bi seveda morali postati tudi administratorji, z vsemi pravicami in dolžnostmi, ki bi jih najlažje opravljali v okviru svoje ustanove, na svojem delovnem mestu.

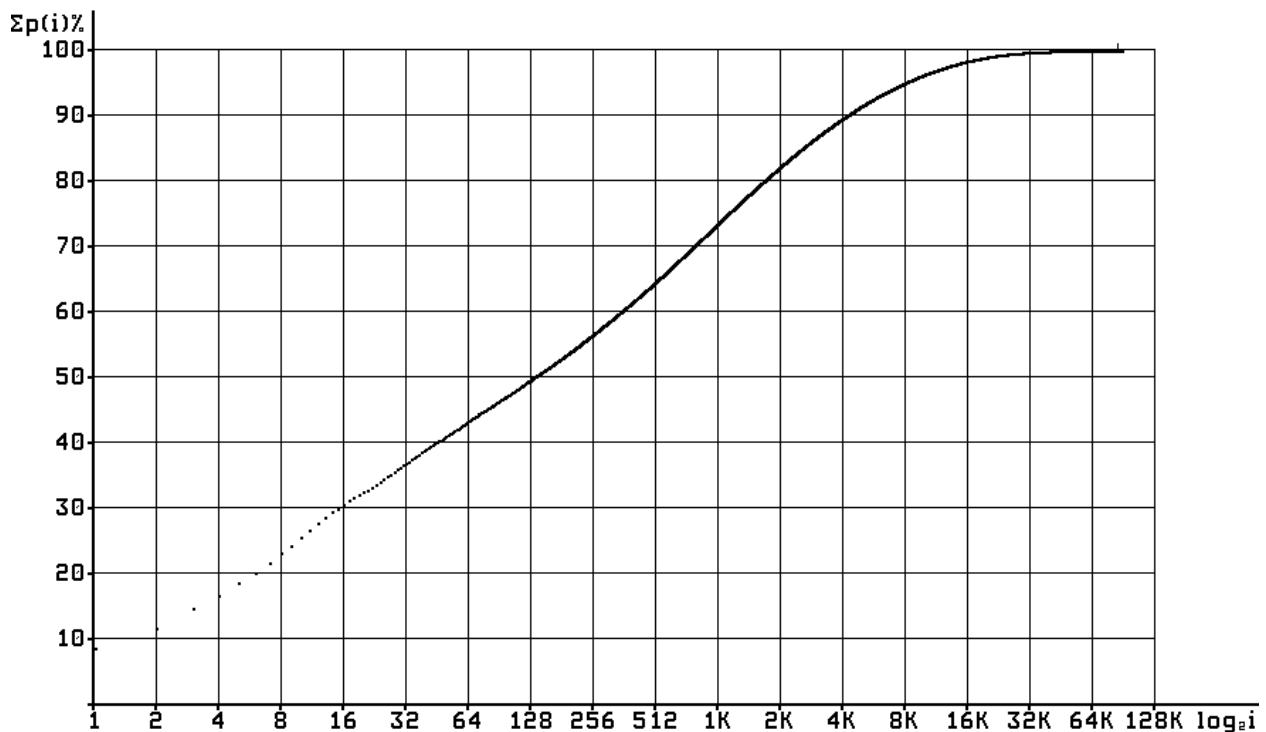
Kaj in koliko vključiti v wikislovar (in Wiktionary), je seveda odprto vprašanje. Kot predhodni korak bi kazalo pomisliti na kak obsežnejši vir, pri katerem je vprašanje avtorskih pravic lažje rešljivo. Primeren kandidat utegne biti spletna zbirka *Besede slovenskega jezika* (SBSJ 2014), ki obsega 357.000 besed s podatki o besedni vrsti in pregibanju, nima pa razlag, ponekod le primere. Viri zanj so bili *Slovar slovenskega knjižnega jezika*, *Besedišče slovenskega jezika*, besedilni korpus *Nova beseda* in spletni iskalnik *NAJDI.SI*. S primernim wikipedijinim robotom bi lahko strojno napravili unijo podatkov iz tega slovarja in iz obstoječega wikislovarja ter s tem postavili solidno osnovo za nastanek vira, primerljivega s tistim za največje svetovne jezike iz tabele 1. Povezava z Wiktionaryjem bi bila tudi sicer zelo pomembna, ker je za veliko iztočnic iz tega seznama (SBSJ) težko dobiti povezave na spletne strani z njihovo rabo. Kakor že se zdi komu obseg sedanjih korpusov velik, se pri slovaropisnem delu pogosto izkaže, da so nezadostni. Seznam lem največjega korpusa Gigafida obsega 3.659.058 enot, to pa seveda še ne pomeni, da ima slovenski jezik tudi toliko besed. Če pogledamo v njem recimo vse leme na miška-:

miška (18,417)	miškaicati	miškaprogramski	miškarstvo
miška.katja	miškalno	miškar	miškas
miškaaa (2)	miškanaslov	miškarati	miškast (7)
miškaaaaa	miškanje (5)	miškarjenje (8)	miškat
miškadorr	miškao (2)	miškarski (9)	
miškaedina	miškapiška	miškarskiti	

Tabela 9: Korpus Gigafida: Besede na *miška-* s pogostnostmi

takoj vidimo, da je med njimi veliko šuma, da pa tudi vse besede s frekvenco 1 niso slabe, taka sta npr. *miškar* in *miškarstvo*. V zbirki *Besede slovenskega jezika* je 150.000 takih, ki jih med lemami Gigafide ne najdemo.

Si pa z Gigafidinim seznamom lem s frekvencami lahko pomagamo pri odgovoru na vprašanje, koliko besed iz novega SSSJ bi kazalo vključiti v wikislovar. Če besede iz npr. *Slovarja slovenskega knjižnega jezika* (SSKJ) opremimo s pogostnostmi iz Gigafide in iz tega seznama napravimo krivuljo rasti, dobimo podobo na sliki 1 (za 10.067 besed iz SSKJ, ki jih tudi ne najdemo med lemami Gigafide, so bile pogostnosti med 1 in 20 naključno izbrane). Prva pika na sliki levo spodaj je pogostnost najpogostejše besede, glagola *biti*, druga pika je skupna pogostnost prvih dveh najpogostejših besed, glagola *biti* in predloga *v*, pri tretji piki se je pridružil še veznik *in* in tako naprej. Opazimo, da polovico besednega zaklada besedil korpusa Gigafida pokrijemo z nekaj manj kot 150-imi najpogostejšimi lemami, tri četrtine z dobrimi 1000 lemami, 90 % s približno 4000, 95 % z 8000, z 32.000 najpogostejšimi lemami pa smo že zelo blizu resnice.



Slika 1: Krivulja rasti za iztočnice iz SSKJ s pogostnostmi iz seznama lem korpusa Gigafida

Z vključitvijo vseh naših gesel v Wiktionary bi bilo vsem jasno, da je slovenski jezik oskrbovan in cvetoč, da se njim ukvarjajo ljudje, ki ga imajo radi, so pa tudi resni, odgovorni in sposobni, da ne gre za jezik, ki ga drži pokonci ena ali morda dve ameriški univerzi.

Če bi nam ta podvig uspel, bi se slovenski jezik v še lepši, sodobnejši in vsem dostopni obliki s svojim besednim zakladom predstavil svetu, kot se mu je prvič pred 120-imi leti s Pleteršnikovim Slovensko-nemškim slovarjem.

#### **4 Viri**

DESROCHES, Joanne, ROTH Melissa 2011: Wiktionary Tutorial, Level I.

[http://linguistech.ca/Wiktionary\\_E\\_TUTCERTT\\_I](http://linguistech.ca/Wiktionary_E_TUTCERTT_I).

GIGAFIDA, 2014. Besedilni korpus: <http://www.gigafida.net>.

LEPORE, Jill, 2006. Life and Letters, "Noah's Mark", *The New Yorker*, November 6, 2006, 78.

MIŠKA-EN, 2014. Geslo v angleškem slovarju Wiktionary:

<https://en.wiktionary.org/wiki/miška>.

MIŠKA-SL, 2014. Geslo v slovenskem slovarju Wikislovar:

<https://sl.wiktionary.org/wiki/miška>.

NOVA BESEDA, 2014. Besedilni korpus: [http://bos.zrc-sazu.si/s\\_beseda.html](http://bos.zrc-sazu.si/s_beseda.html).

SBSJ, 2014. Seznam Besede slovenskega jezika: <http://bos.zrc-sazu.si/besede.html>.