

# **Navodila za semantično označevanje korpusa JOS**

različica 1.2

23.04.2009

**Projekt**  
**Jezikoslovno označevanje slovenskega jezika**

<http://nl.ijs.si/jos/>

**Odsek za tehnologije znanja**  
**Institut Jožef Stefan**

**Pripravila: Darja Fišer**

## Kazalo

1	Cilj projekta.....	3
2	Osnovni pojmi.....	3
3	DEBVisDic.....	4
3.1	Navodila za namestitev urejevalnika DEBVisDic.....	4
3.2	Navodila za uporabo urejevalnika DEBVisDic.....	8
3.3	Primerjava sinsetov v obeh jezikih.....	8
3.4	Različni pogledi.....	8
3.4.1	Pogled Edit.....	9
4	Navodila za popravljanje wordneta.....	10
4.1	Viri.....	10
4.2	Večpomenski in večbesedni literali.....	10
4.3	Popravljanje sinseta.....	10
4.4	Dodajanje nadpomenk.....	11
4.5	Dodajanje literalov.....	12
5	Navodila za označevanje korpusa.....	12
5.1	Kaj vpisujemo.....	12
5.2	Kako določimo pravilni pomen.....	14
5.2.1	Večbesedne zveze.....	14
5.2.2	Manjkajoči pomeni v sloWNetu.....	15
5.2.3	Lastna imena.....	18
5.2.4	Nejasni pomeni.....	19
6	Seznam koristnih spletnih virov.....	19

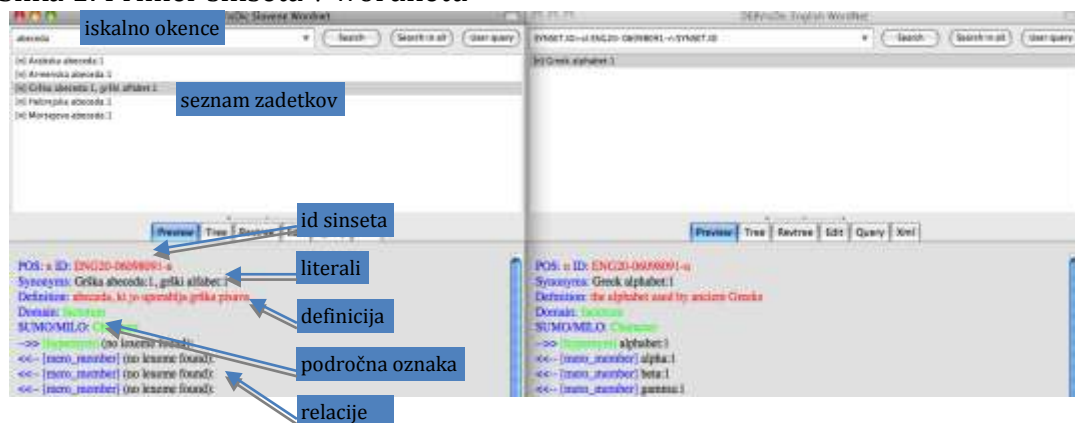
## 1 Cilj projekta

- povezava korpusa jos100k in slovenskega wordneta
- popravljanje wordneta glede na korpusne podatke
- označevanje pomena besed v korpusu s pomočjo wordneta

## 2 Osnovni pojmi

Wordnet je pojmovno zasnovan semantični leksikon, v katerem so pojmi glede na pomen med seboj povezani s semantičnimi relacijami. Pojmi v wordnetu so razdeljeni na samostalniške, glagolske, pridevniške in prislovne. Vse besede, ki označujejo isti pojem in imajo isti pomen, so povezane v **sinset**. Vsak sinset ima svoj **ID**. Sopomenke v sinsetu imenujemo **literali**, ki so lahko eno- ali večbesedni. Poleg sopomenske relacije so v wordnetu označene še nad- in podpomenke, protipomenke ter holo- in meronimi. Sinset vsebuje tudi definicijo, primere rabe, področno oznako (npr. avtomobilizem) in povezavo na ontologijo. Primer sinseta prikazuje Slika 1.

Slika 1: Primer sinseta v wordnetu



Dodatne informacije o slovenskem wordnetu najdete na projektni spletni strani:  
<http://lojze.lugos.si/~darja/slownet.html>.

### 3 DEBVisDic

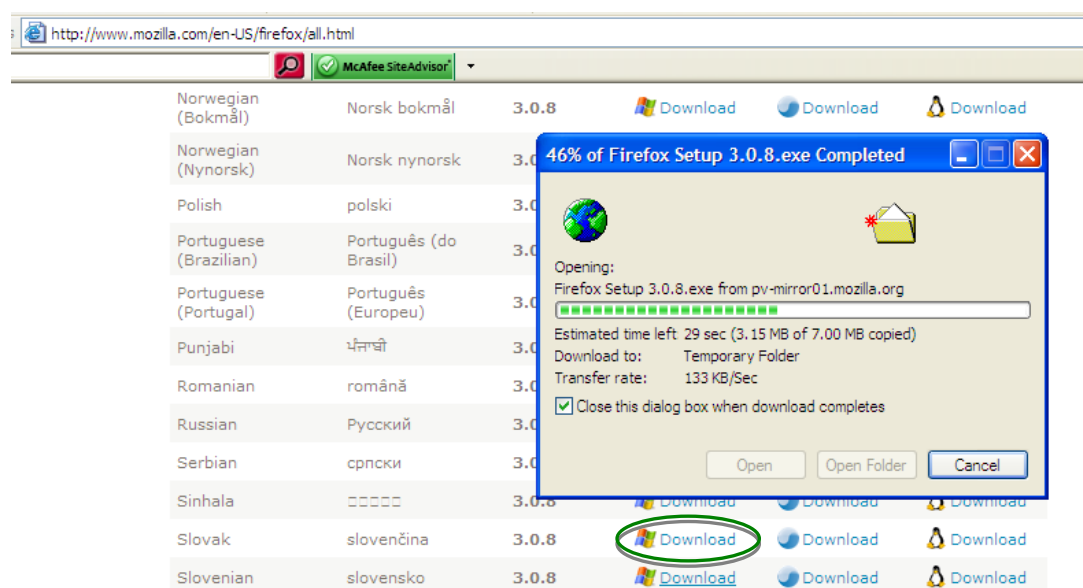
DEBVisDic je orodje za pregledovanje wordneta in ga potrebujete za sodelovanje pri projektu. V njem boste lahko iskali in popravljali sinsete v avtomatsko izdelanem wordnetu za slovenščino in ga hkrati primerjali z angleškim wordnetom. DEBVisDic si morate najprej namestiti na svoj računalnik, nato pa preko interneta dostopate do slovenskega in angleškega wordneta.

#### 3.1 Navodila za namestitev urejevalnika DEBVisDic

DEBVisDic deluje v operacijskem sistemu Windows. Za namestitev DEBVisDica boste potrebovali spletni brskalnik Firefox 3, ki ga dobite tukaj:

<http://www.mozilla.com/en-US/firefox/firefox.html> (glej Sliko 1).

Slika 1: Namestitev urejevalnika Firefox 3.0.8



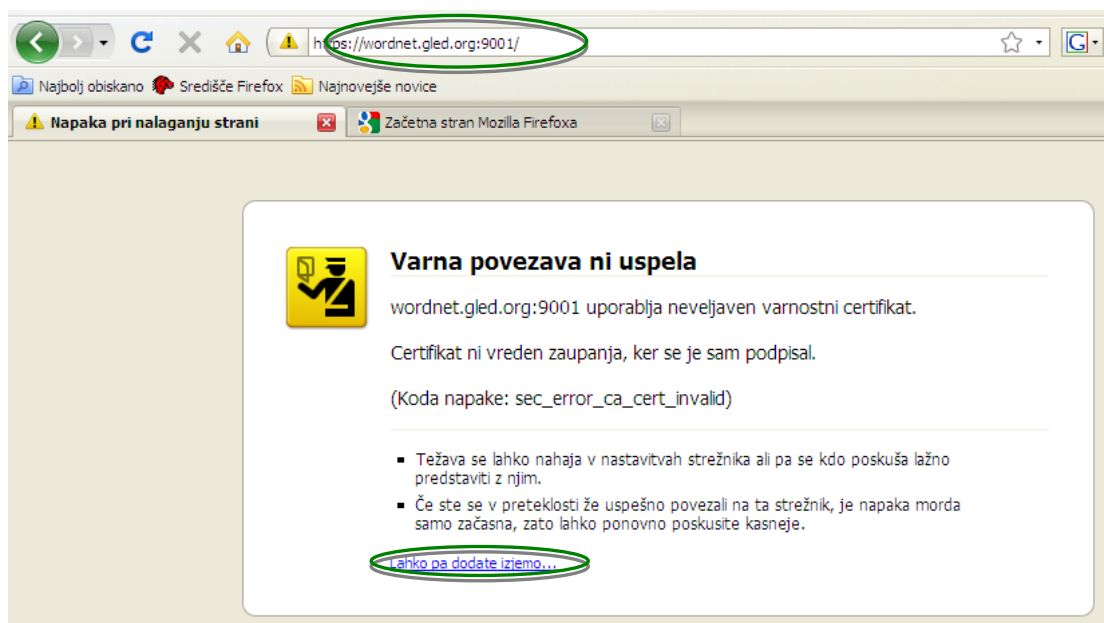
Ko ste uspešno namestili Firefox, si morate namestiti še DEBVisDic klient, ki ga dobite tukaj: <http://deb.fi.muni.cz/install.php> (druga povezava na strani, glej Sliko 2).

Slika 2: Namestitev urejevalnika DEBVisDic



Zdaj ponovno zaženite Firefox. Pojdite na naslov <https://wordnet.gled.org:9001> in potrdite izjemo za certifikat (glej Sliko 3). Nato zaženite DEBVisDic (Tools → DEBVisDic, glej Sliko 4).

Slika 3: Dodajanje izjeme za certifikat

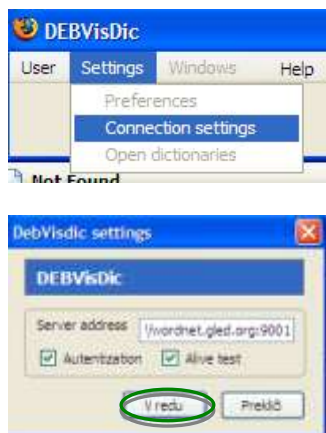


Slika 4: Zagon urejevalnika DEBVisDic



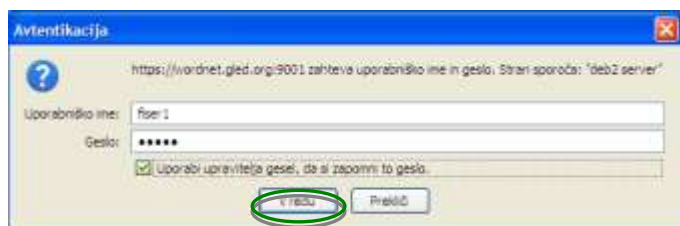
Ko ste DEBVisDic zagnali, se vam odpre prijavno okno, vendar se sprva ne boste mogli prijaviti, zato zaprite prijavno okno in spremenite naslov strežnika za DEBVisDic. Pojdite na Settings → Connection Settings in v okence Server Address vpišite <https://wordnet.gled.org:9001>. Ostalih nastavitev ne spreminjajte in kliknite OK (glej Sliko 5).

Slika 5: Nastavitev strežnika za dostop do wordneta



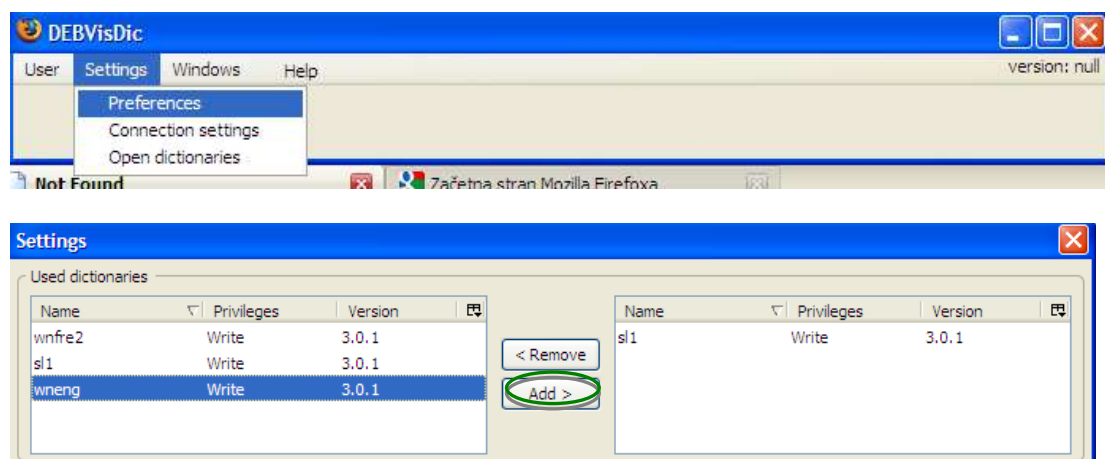
Zdaj v prijavno okno vpišite svoje uporabniško ime in geslo, kot prikazuje Slika 6.

Slika 6: Prijava na strežnik



Ko ste se uspešno prijavili na strežnik, pojdite na Settings → Preferences in izberite slovenski in angleški wordnet ter ju dodajte s klikom na gumb Add. Nato kliknite OK (glej Sliko 7).

Slika 7: Dodajanje slovenskega in angleškega wordneta



### 3.2 Navodila za uporabo urejevalnika DEBVisDic

Podrobnejša navodila za delo z DEBVisDicom najdete tukaj:

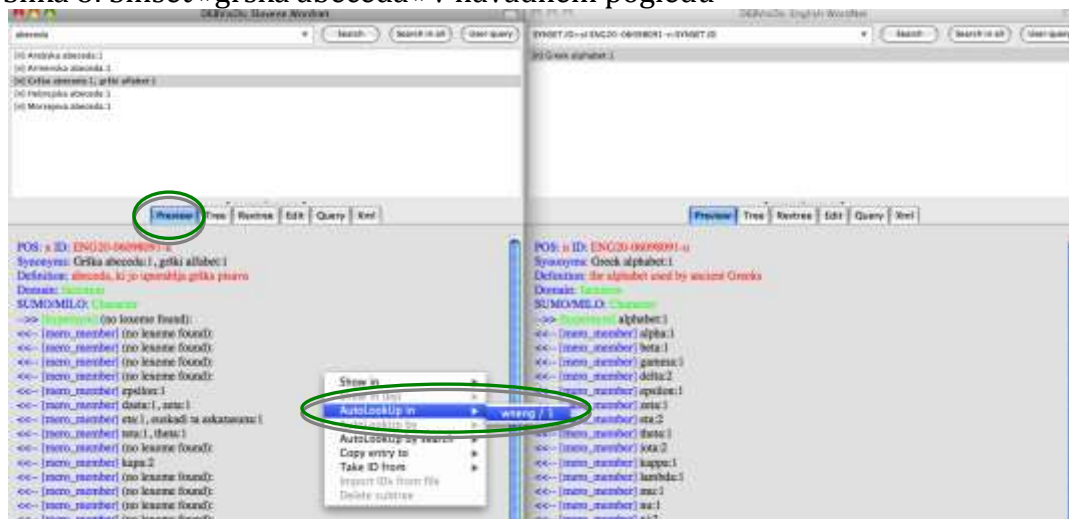
<http://nlp.fi.muni.cz/trac/deb2/wiki/DebVisDicManual> (v angleščini) in jih toplo priporočam. Na hitro pa tole:

### 3.3 Primerjava sinsetov v obeh jezikih

DEBVisDic v enem okencu prikazuje angleški, v drugem pa slovenski wordnet.

Iskano besedo vpišite v iskalno polje in kliknite gumb Search. Če želite, da se vam isti sinset prikaže tudi v angleškem wordnetu, z desno miško kliknite na prikazan sinset v največjem spodnjem oknu, in izberite ukaz »AutoLookUp in« (glej Sliko 8). Angleški wordnet uporabljajte za preverjanje ustreznosti slovenskih prevodov (upoštevajte sinonime, definicijo, primere rabe in semantične relacije).

Slika 8: Sinset »grška abeceda« v navadnem pogledu

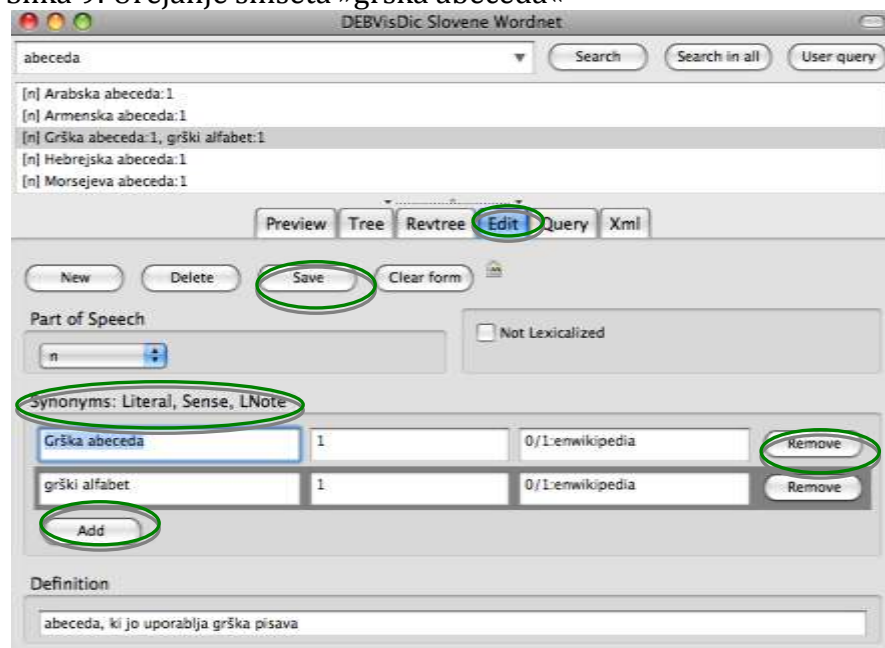


### 3.4 Različni pogledi

V zavihku »Preview« se prikaže celoten sinset. V zavihku »Tree« si lahko ogledate celotno drevo, v katerega sinset sodi. V zavihku »Edit« pa popravljate napake in dodajate manjkajoče sinonime (glej Sliko 9).



Slika 9: Urejanje sinseta »grška abeceda«



### 3.4.1 Pogled Edit

Vse sinonime prikažete s klikom na plus v kategoriji »Synonyms: Literal, Sense, LNote«. V levem okvirčku so posamezni sinonimi, v naslednjem okvirčku št. pomena, v zadnjem pa vir, iz katerega je bil literal pridobljen. Na koncu vrstice je gumb »Remove«, na dnu kategorije pa gumb »Add«.

Gumb »Remove« uporabite, kadar je celoten sinonim neustrezen in ga želite odstraniti. Gumb »Add« uporabite, kadar pomemben sinonim v sinsetu manjka in ga želite dodati. Ko želite spremeniti le obstoječ sinonim, spremembe vnašate v levi okvirček (»Literal«). Spremembe shranite s pritiskom na gumb »Save« na vrhu okna (glej Primer 1).

**Primer 1:** V sinsetu »grška abeceda« je prvi literal napačno zapisan z veliko začetnico, ki jo moramo popraviti v malo. Drugi literal se nam zdi nenavaden, zato ga preverimo v korpusu FidaPLUS, ki vrne 61 zadetkov, pa še v SSKJ ga najdemo, zato je literal v wordnetu ustrezen in ga ne spreminjamo. Definicija sinseta je v slovenščini, vendar je pravilna, zato tudi te ne spreminjamo. Nato kliknemo še gumb »Save«.

## 4 Navodila za popraviljanje wordneta

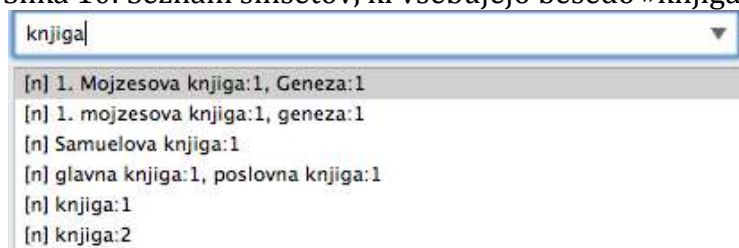
### 4.1 Viri

Pri projektu uporabljajte vse eno- in večjezične slovarske, glosarske in korpusne vire, ki so za slovenščino na voljo: SSKJ, FidaPLUS, Evroterm itd (glej poglavje 6). Svoje predloge za prevode vedno preverite v virih in si jih in njihovih pomenov ne zmišljajte.

### 4.2 Večpomenski in večbesedni literali

Isti literal v wordnetu se lahko pojavi v enem samem sinsetu (enopomenska beseda oz. besedna zveza) ali v več kot enem sinsetu (večpomenska beseda oz. besedna zveza). Vaša naloga je, da pregledate **vse pomene besede**, ki vam je bila dodeljena, pa tudi **vse pomene večbesednih zvez**, ki vašo besedo vsebujejo (glej Sliko 10 in Primer 2).

Slika 10: Seznam sinsetov, ki vsebujejo besedo »knjiga«



*Primer 2: Zanima nas beseda »knjiga«, ki jo vpišemo v iskalno okence. Med zadetki se pojavita dva pomena besede »knjiga«, poleg tega pa tudi štiri večbesedni izrazi, ki vsebujejo besedno »knjiga«, zato preverimo vseh šest sinsetov in v njih odpravimo morebitne napake (glej Primer 3).*

### 4.3 Popraviljanje sinseta

Slovenske sinsete primerjajte z angleškimi. Če v slovenskem wordnetu opazite kakšno napako, pojdite na jeziček Edit in jo popravite oz. izbrišite napačen literal v sinsetu. Pri tem upoštevajte angleške sinonime za ta pomen, definicijo in primere rabe ter semantične relacije sinseta (glej Primer 3).

**Primer 3:** Pri pregledovanju sinsetov, ki vsebujejo besedo »knjiga«, opazimo, da prva dva sinseta vsebujeta enake literale. Ko sinseta preverimo še v angleškem wordnetu, ugotovimo, da sta literala ustrezna za prvi pomen (*the first book of the Old Testament*), za drugega pa je ustrezen samo literal »geneza« (*genesis:1, generation:5, a coming into being*), zato napačen literal iz tega sinseta izbrišemo. Prav tako moramo popraviti sinset »Samuelova knjiga« v »Samuel«. Ostali sinseti so pravilni in jih ne spreminjamo.

V angleškem wordnetu ne popravljajte ničesar. V slovenskem wordnetu pa **popravljajte samo literale** (sinonime), primerov rabe in semantičnih relacij ne spreminjajte, definicije pa samo, če so v slovenščini in v njih opazite kakšno napako. V tem primeru napačne slovenske definicije zamenjajte z angleško definicijo, ki jo najdete pri istem sinsetu v angleškem wordnetu (glej Primer 4).

**Primer 4:** Pri pregledovanju sinseta »Križev pot« opazimo, da vsebuje naslednjo definicijo: *poleg tega tudi molitev, ki jo verniki molijo, ko se spominjajo tega trpljenja. Ker ji manjka prvi del definicije, jo ali izpolnimo, če dela ni veliko, ali zamenjamo z angleško definicijo iz angleškega sinseta.*

#### 4.4 Dodajanje nadpomenk

Pri pregledovanju wordneta pogledite tudi **vse nadpomenke** dodeljene besede do vrha taksonomije. Če nadpomenka še ni prevedena v slovenščino (»no lexeme found«), s klikom na gumb Go To pojdite nanjo in vanjo dodajte ustrezen prevod oz. prevode (glej Sliko 11 in Primer 5).

Slika 11: Nadpomenka sinseta »grška abeceda« je prazna

POS: n ID: ENG20-06098091-n  
Synonyms: Grška abeceda:1, grški alfabet:1  
Definition: abeceda, ki jo uporablja grška pisava  
Domain: factotum  
SUMO/MILO: Character  
--> [hypernym] (no lexeme found):

Relations

[\*] [n] (no lexeme found):    hypernym    Go to    Remove

Add

**Primer 5:** Pri pregledovanju sinseta »grška abeceda« opazimo, da je njegova nadpomenka prazna (»no lexeme found«) Zato se v pogledu za urejanje

*premaknemo na ta sinset s klikom na gumb »Go to«, ga preverimo v angleškem sinsetu (alphabet:1, a character set that includes letters and is used to write a language) in ustrezen prevod »abeceda« vstavimo kot literal za ta sinset v slovenski wordnet.*

#### 4.5 Dodajanje literalov

Če v wordnet dodate kak literal, ga vpišite v okence Literal. Okence Sense v angleškem delu wordneta vsebuje številko pomena večbesednih literalov, v slovenščini pa ga uporabljamo za štetje pojavitev v korpusu pri avtomatski izdelavi wordneta. Vi v to okence vpišite črko »x«, v okence LNote pa vir, v katerem ste slovenski literal našli. Če želite dodati več sopomenk, vsako sopomenko vpišite v novo vrstico kot nov literal, ne v isti okvirček, ločen z vejicami. Spremembe shranite s klikom na gumb Save (glej Sliko 12).

Slika 12: Primer dodajanja literala »abeceda«



The screenshot shows a software interface for adding synonyms. At the top, it says "Synonyms: Literal, Sense, LNote". Below this, there are three input fields: "Literal" containing "abeceda", "Sense" containing "x", and "LNote" containing "sskj". Each of these three fields is circled in green. To the right of the "LNote" field is a "Remove" button. Below the input fields is an "Add" button.

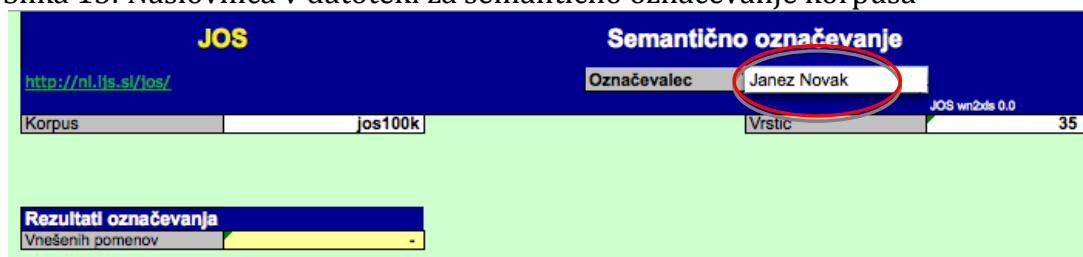
## 5 Navodila za označevanje korpusa

V Excelovih datotekah boste dobili konkordance iz korpusa, v katerih morate izbranim samostalnikom glede na kontekst določiti pomen iz wordneta. Pri določanju pomena upoševajte spodnja navodila. V primeru težav si med seboj pomagajte s forumom, ki je posebej za ta projekt na voljo na spletni učilnici, če tudi to ne zadošča, pa mi čim prej pišite (Subject: slownet - vprašanje).

### 5.1 Kaj vpisujemo

Najprej v list Naslovnica v celico F2 **vpišite svoje ime in priimek** (glej Sliko 13). Ko končate z delom, datoteko shranite, pri čemer v ničemer ne spreminjajte njenega imena. Nato ga oddajte v spletni učilnici.

Slika 13: Naslovnica v datoteki za semantično označevanje korpusa



Nato pojdite na list Anotacije, preberite vsako konkordanco posebej in si izoblikujte mnenje o pomenu besede v tem kontekstu. Potem preverite vse pomene te besede v wordnetu in izberite najboljšega. **Id tega pomena besede skopirajte** iz wordneta v datoteko s konkordancami (v stolpec C – »Pomen«). V ta stolpec vpisujte samo ID-je in nič drugega (glej Sliko 14 in Primer 6). Pri označevanju morate izbrati sinset iste besedne vrste, kot je beseda oz. besedna zveza, ki jo označujete.

Slika 14: Označevanje korpusa z wordnetovimi id-ji

A	C	D	E	F	
n	<b>pomen</b>	Opomba	levi kontekst	beseda	desni kontekst
13			poto Triglava , je za	knjigo	Triglav , Sveta gora Slo
14			i Založbi Mladinska	knjiga	, izbral France Stele . F
15			to iz avtorjeve nove	<b>knjige</b>	Imena nebesnih teles .
16			tu 1986 izšla njena	knjiga	How Institutions Think S

POS: n ID: **ENG20-06013091-n** BCS: 1  
 Synonyms: knjiga:2  
 Definition: a written work or composition that has been published (printed on pages bound together)  
 Usage: I am reading a good book on economics  
 Domain: publishing  
 SUMO/MILO: Book  
 -->> [hypernym] edicija:1, publikacija:2

A	C	D	E	F	
n	<b>pomen</b>	Opomba	levi kontekst	beseda	desni kontekst
13			poto Triglava , je za	knjigo	Triglav , Sveta gora Slo
14			i Založbi Mladinska	knjiga	, izbral France Stele . F
15	<b>ENG20-06013091-n</b>		to iz avtorjeve nove	knjige	Imena nebesnih teles .
16			tu 1986 izšla njena	knjiga	How Institutions Think S

**Primer 6:** Označiti želimo 15. pojavitev besede »knjiga« v korpusu. Preberemo stavek in rabo besede primerjamo z vsemi pomeni »knjige« v wordnetu. Najprej besedo poiščemo in opazimo, da ima dva možna pomena. Prvi se nanaša na knjigo kot objekt, drugi pa na knjigo kot pisno stvaritev. V našem primeru je najustreznejši ta pomen, zato skopiramo njegov id in ga za 15. pojavitev besede vnesemo v stolpec C v datoteki s konkordancami.

## 5.2 Kako določimo pravilni pomen

Izbrani besedi pomen določite tako, da preberete konkordance. Če kontekst, ki ga pokaže konkordanca, ni zadosten za določitev pomena, kliknite na besedo v stolpcu F, da dobite širši kontekst preko mrežnega konkordančnika. Nato v wordnetu med vsemi možnimi pomeni ob upoštevanju definicije, primerov rabe in semantičnih relacij skušajte najti najprimernejšega za pojavitev, ki jo v korpusu označujete.

Sledijo podrobnejša navodila za težje primere, ki jih pozorno preberite in upoštevajte. Če naletite na primer, ki v navodilih ni opisan, mi čim prej pišite.

### 5.2.1 Večbesedne zveze

Če menite, da je **beseda v konkordancah del večbesedne zveze**, ki nosi pomen, in jo tudi najdete v wordnetu, v stolpec C vpišite id te večbesedne zveze iz wordneta. Pri tem si vedno prizadevajte označiti največjo možno večbesedno zvezo, ne samo njenih delov (npr. »beli morski pes«, ne samo »morski pes«). V stolpec D - »Opomba« jo vpišite v isti obliki, kot se pojavi v konkordancah. Pred njo vpišite znak \*. Razen večbesednega termina v celico ne vpisujte ničesar drugega (glej Sliko 15 in Primer 7). Večbesedne zveze naj bodo večbesedni strokovni termini (npr. »veliki metljaj«), večbesedni idiomatski izrazi, katerih pomen ni prosojen, (npr. »panjska končnica«) in zelo pogoste kolokacije (npr. »srednja šola«). Prostih večbesednih zvez (npr. »rdeč avto«) ne označujete kot večbesedne zveze.

Slika 15: Označevanje večbesedne zveze

A	C	D	E	F
n	<b>pomen</b>	Opomba	levi kontekst	beseda d
17			e pojavil rdeč kolobar na telesu a	
18			in predsednike delovnih teles	,
19		( Delno povzeto iz avtorjeve nove knjige Imena <b>nebesnih teles</b>		
20			abimo Foucaultov izraz , telo in	

POS: n ID: **ENG20-08664330-n** BCS: 1  
Synonym: **nebesno telo:1**  
Definition: **natural objects visible in the sky**  
Domain: **astronomy**  
SUMO/MILO: **AstronomicalBody**  
--> [hypernym] naravni objekt:1

A	C	D	E	F
n	<b>pomen</b>	Opomba	levi kontekst	beseda d
17			e pojavil rdeč kolobar na telesu a	
18			in predsednike delovnih teles	,
19	<b>ENG20-08664330-n</b>	<b>*nebesnih teles</b>	e knjige Imena nebesnih	teles

**Primer 7:** Označiti želimo 19. pojavitev besede »telo« v korpusu. Preberemo stavek in ugotovimo, da gre za večbesedno zvezo »nebesno telo«, ki jo tudi najdemo v wordnetu. Zato 19. pojavitvi besede »telo« pripišemo id sinseta »nebesno telo«, v opombo pa vpišemo večbesedno zvezo, kot se pojavi v korpusu, in ji dodamo znak \*.

## 5.2.2 Manjkajoči pomeni v sloWNetu

Če v wordnetu **pomen, ki sta ga identificirali v korpusu, manjka**, ga za to besedo oz. besedno zvezo dodajte tako, da s pomočjo angleškega wordneta najdete primeren sinset in ga dodajte še v slovenskega, kjer zaenkrat piše samo »no lexeme found« (glej Sliko 16 in Primer 8).

Slika 16: Dodajanje literala v slovenski wordnet in označevanje korpusa

A	C	D	E	F	G
n	<b>pomen</b>	Opomba	levi kontekst	beseda	desni kontekst
5			raje lenari in prebira	knjige	, predvsem tiste
6			prebral prvo takšno	knjigo	, za njo so se zv
7			matizacija	<b>zemljiške knjige</b>	
8			ilelfu , da je napisal	knjigo	zgodb , ki so jih

POS: n ID: **ENG20-06100818-n**  
 Synonyms: cadaster:1, cadastre:1  
 Definition: a public register showing the details of ownership and value of land; made for the purpose of taxation  
 Domain: building\_industry  
 SUMO/MILO: Certificate  
 --> [hypernym] register:1, registry:1  
 <<-- [derived] cadastral:1

New Delete Save Clear form

Part of Speech: n  Not Lexicalized

Synonyms: Literal, Sense, LNote

**(no lexeme found)**

Definition: a public register showing the details of ownership and value of land; made for the purpose of taxation

Usage:

Relations: [n] register:2

A	C	D	E	F	G
n	<b>pomen</b>	Opomba	levi kontekst	beseda	desni kontekst
5			raje lenari in prebira	knjige	, predvsem tiste za osebnos
6			prebral prvo takšno	knjigo	, za njo so se zvrstile druge
7	<b>ENG20-06100818-n</b>	<b>*zemljiške knjige</b>	matizacija zemljiške	knjige	
8			ilelfu , da je napisal	knjigo	zgodb , ki so jih opisali kot "





**Primer 8:** Označiti želimo 7. pojavitev besede »knjiga« v korpusu. Preberemo stavek in ugotovimo, da gre za večbesedni termin »zemljiška knjiga«, ki je v slovenskem wordnetu ne najdemo. Poskusimo še poiskati po angleškem wordnetu, v katerem prevod za »zemljiško knjigo«, ki smo ga dobili v glosarju Evroterm, najdemo: »cadaster, cadastre«. Slovenski sinset je prazen, zato namesto opozorila »no lexeme found« vpišemo termin »zemljiška knjiga«, v polje Sense vpišemo »x«, v polje LNote pa vir »evroterm«. Nato še 7. pojavitvi besede »telo« v datoteki s konkordancami v stolpec C - Pomen pripišemo id sinseta »nebesno telo«, v stolpec D - Opomba pa vpišemo večbesedno zvezo, kot se pojavi v korpusu, in ji dodamo znak \*. Podobno bi storili z izrazoma »kuharska knjiga« (ang. »cookbook«) in »ovitek knjige« (ang. »book jacket«).

Če **pojem v angleškem wordnetu ne obstaja**, za pojavitev besede oz. besedne zveze v korpusu izberite tisti pomen, ki v wordnetu obstaja in je za vašo pojavitev v korpusu od vseh, ki so na voljo, najbližji in najbolj primeren, v stolpec D - »Opomba« pa vpišite besedo oz. besedno zvezo, za katero v wordnetu niste našli ustreznega pomena. Vpišite jo točno tako, kot se pojavi v konkordancah, pred njo pa vpišite znak #. Razen tega v to celico ne vpisujte ničesar (glej Sliko 17 in Primer 9).

Slika 17: Označevanje pojma, ki ga v nobenem wordnetu ne najdemo

A	C	D	E	F	G
n	<b>pomen</b>	Opomba	levi kontekst	beseda	desni kontekst
33			to tega bere preveč	knjig	. «
34			rodilni pisci otroških	knjig	veliko časa namenijo rednen
35			om in podpisovanju	knjig	. Navzočih je več sto otrok .

POS: n ID: **ENG20-06013091-n** BCS: 1  
 Synonyms: knjiga:2  
 Definition: a written work or composition that has been published (printed on pages bound together)  
 Usage: I am reading a good book on economics  
 Domain: publishing  
 SUMO/MILO: Book  
 -->> [hypernym] edicija:1, publikacija:2

A	C	D	E	F	G
n	<b>pomen</b>	Opomba	levi kontekst	beseda	desni kontekst
33			to tega bere preveč	knjig	. «
34	<b>ENG20-06013091-n</b>	<b>#otroških knjig</b>	rodilni pisci otroških	knjig	veliko časa namenijo rednen
35			om in podpisovanju	knjig	. Navzočih je več sto otrok .

**Primer 9:** Označiti želimo 34. pojavitev besede »knjiga« v korpusu. Iz stavka razberemo, da gre morda za večbesedni termin »otroška knjiga«, ki ga pa niti v slovenskem niti v angleškem wordnetu ne najdemo. Zato označimo samo pomen besede »knjiga«. Najprimernejši je drugi pomen v slovenskem wordnetu (pisna stvaritev), tako da izberemo njegov id in ga skopiramo v stolpec C, pod opombo pa vpišemo večbesedni izraz, natanko tako kot se pojavi v korpusu in ki ga v wordnetu nismo našli, in mu dodamo znak #.

### 5.2.3 Lastna imena

Če je beseda ali večbesedna zveza, ki jo označujete, **del lastnega imena**, pa se ji lepo da določiti pomen, to besedo ali besedno zvezo označite kot vse ostale pojavitve. Če je lastno ime enobesedno, ji pomen prav tako skušajte določiti in vpišite ustrezen id. Če pomena vendarle ne morete določiti, ker gre za poimenovanje v prenesenem pomenu ali ga ne morete razbrati iz sobesedila, ga ne označujte z nobenim id-jem, v stolpec C vpišite znak !, v stolpec D pa vpišite opombo »!lastno« (glej Sliko 18 in Primer 10).

Slika 18: Označevanje lastnih imen

A	C	D	E	F	G
n	<b>pomen</b>	Opomba	levi kontekst	beseda	desni kontekst
11			Nekatere	stranke	se zavzemajo tudi za kazenske ovadb
12			imenoma . Da bi v ljudski	stranki	nastala zaradi spornih finančnih poslov
13			Tako je kmečka Ljudska	stranka	( PSL ) lani ustvarila največji dohodek
14			ov . Predstavniki Ljudske	stranke	so namreč iz državnega proračuna dol

POS: n ID: ENG20-07758173-n BCS: 1  
 Synonyms: partija:1, stranka:1  
 Definition: organizacija, ki si prizadeva priti na oblast ali vsaj dobiti čim večji vpliv na izvajanje državne oblasti  
 Usage: in 1992 Perot tried to organize a third party at the national level  
 Domain: anthropology  
 SUMO/MILO: PoliticalOrganization  
 -->> [hypernym] kolektiv:1, organizacija:4  
 -->> [holo\_member] oblika vladavine:1, politični sistem:1

A	C	D	E	F	G
n	<b>pomen</b>	Opomba	levi kontekst	beseda	desni kontekst
11			Nekatere	stranke	se zavzemajo tudi za kazens
12			imenoma . Da bi v ljudski	stranki	nastala zaradi spornih finanč
13	ENG20-07758173-n		Tako je kmečka Ljudska	stranka	( PSL ) lani ustvarila največji
14			ov . Predstavniki Ljudske	stranke	so namreč iz državnega prof
15			zametke novih političnih	strank	. Darko je do konca verjel v

A	C	D	E	F	G
n	<b>pomen</b>	Opomba	levi kontekst	beseda	desni kontekst
10			o in umrejo , ostane samo njihova	hiša	iz kalcijevega karbonata in protein
11			tternichovem stanovanju v Zoisovi	hiši	na Bregu in drugod .
12			tudi Damjana Golavšek , skupin	Hiša	Božo A. Kolerič , Irena Vrčkovnik
13			i dva delavca , ki sta živela v njeni	hiši	. Povedala sta , da ni jedla svinine

A	C	D	E	F	G
n	<b>pomen</b>	Opomba	levi kontekst	beseda	desni kontekst
10			o in umrejo , ostane samo njihova	hiša	iz kalcijevega karbonata in protein
11			tternichovem stanovanju v Zoisovi	hiši	na Bregu in drugod .
12	!	!lastno	tudi Damjana Golavšek , skupina	Hiša	, Božo A. Kolerič , Irena Vrčkovnik
13			i dva delavca , ki sta živela v njeni	hiši	. Povedala sta , da ni jedla svinine

*Primer 10: Označiti želimo 13. pojavitev besede »stranka«. Med prebiranjem konkordanc ugotovimo, ga je beseda del večbesedne zveze, ki ji hkrati tudi lastno ime. Vendar pomena besede »stranka« v njej ni težko določiti (gre za politično stranko), zato ji pripišemo ta id. Po drugi strani pa pri označevanju 12. pojavitve vesede »hiša« ugotovimo, da gre za glasbeno skupino s tem imenom, ne moremo pa razbrati, v kakšnem smislu je uporabljena. Zato v tem primeru besedi »hiša« ne pripišemo nobenega pomena, temveč dodamo opombo, da gre za lastno ime.*

#### 5.2.4 Nejasni pomeni

Če se med več pomeni iz wordneta **ne morete odločiti, kateri je pravi**, se potrudite izbrati najsplošnejšega (v angleškem wordnetu so splošnejši in pogostejši pomeni označeni z najnižjimi številkami). Če se vam vendarle zdi, da gre za več kot en možni pomen, v stolpec C vpišite vse id-je, ki po vašem ustrezajo pomenu iz konteksta. Id-je med seboj ločite s presledkom in med njiju ne vpisujte nobenih drugih znakov (npr. vejice).

Če za neko pojavitev besede v korpusu ne najdete nobenega primerne sinseta v wordnetu, je ne označujte z nobenim id-jem, v stolpec C vpišite znak !, v stolpec D pa po potrebi vpišite opombo, kateri na začetku prav tako dodajte znak !

## 6 Seznam koristnih spletnih virov

Wordnet:

- Princeton WordNet: <http://wordnet.princeton.edu/>
- SloWNet: <http://lojze.lugos.si/~darja/slownet.html>

Programska oprema:

- Firefox 3.0: <http://www.mozilla.com/en-US/firefox/firefox.html>
- DEBVisDic: <http://deb.fi.muni.cz/install.php>

Slovarji in glosarji:

- SSKJ: <http://bos.zrc-sazu.si/sskj.html>
- Pravopis
- Evroterm: <http://evroterm.gov.si/index.php?jezik=angl>
- Islovar : [http://www.islovar.org/iskanje\\_enostavno.asp](http://www.islovar.org/iskanje_enostavno.asp)

Korpusi:

- FidaPlus: <http://www.fidaplus.net/>
- Sketch Engine: <http://www.sketchengine.co.uk/>
- korpusi na IJS: <http://nl2.ijs.si/>