



SPECIFIKACIJE ZA IZDELAVO
LEKSIKALNE BAZE ZA SLOVENŠČINO

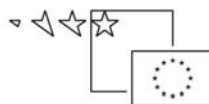
OPIS ANALIZE REFERENČNEGA KORPUSA
(KAZALNIK 5)

PROJEKT "SPORAZUMEVANJE V SLOVENSKEM JEZIKU"

Operacijo delno financira Evropska unija iz Evropskega socialnega sklada ter Ministrstvo za šolstvo in šport. Operacija se izvaja v okviru Operativnega programa razvoja človeških virov za obdobje 2007-2013, razvojne prioritete: razvoj človeških virov in vseživljenjskega učenja; prednostne usmeritve: izboljšanje kakovosti in učinkovitosti sistemov izobraževanja in usposabljanja 2007-2013.

Avtorji: Polona Gantar, Katja Grabnar, Polonca Kocjančič, Simon Krek, Olga Pobirk, Rok Rejc, Mojca Šorli, Simon Šuster in Petra Zaranšek

Kamnik, 30. junij 2009

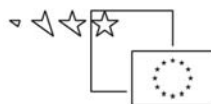


Kazalo vsebine:

1 Uvod: programska oprema za analizo referenčnega korpusa in izdelavo leksikalne baze za slovenščino	7
2 Programska oprema za analizo referenčnega korpusa: <i>Sketch Engine</i>	9
2.1 Uvod.....	10
2.2 Osnovne funkcije orodja SkE: Konkordance, Besedne skice, Primerjalne skice	13
2.2.1 Konkordance (<i>Concordance</i>)	13
2.2.1.1 Izdelava seznama konkordanc	13
2.2.1.2 Opis možnosti v iskalnem oknu za izdelavo seznama konkordanc.....	14
2.2.1.2.1 Iskanje s podatki o ključnih besedah (<i>Keywords</i>).....	14
2.2.1.2.2 Iskanje s podatki o sobesedilu (<i>Context</i>)	16
2.2.1.2.3 Iskanje s podatki o besedilu (<i>Text Types</i>)	17
2.2.1.3 Prikaz in obdelava konkordanc	17
2.2.1.3.1 Možnosti prikaza (<i>View options</i>).....	18
2.2.1.3.2 Vzorec (<i>Sample</i>)	19
2.2.1.3.3 Filter (<i>Filter</i>).....	19
2.2.1.3.4 Razvrščanje (<i>Sort</i>).....	20
2.2.1.3.5 Pogostost (<i>Frequency</i>).....	23
2.2.1.3.6 Kolokacija (<i>Collocation</i>).....	31
2.2.1.3.7 Shrani (<i>Save</i>).....	33
2.2.2 Besedna skica (<i>Word Sketch</i>).....	33
2.2.2.1 Opis	33
2.2.2.2 Izdelava Besedne skice	33
2.2.3 Primerjalna skica (<i>Sketch-Diff</i>).....	36
2.2.3.1 Opis	36
2.2.3.2 Izdelava Primerjalne skice	36
2.3 Dodatne funkcije v SkE beta: Kliksikografija (TBL), Izbor dobrih zgledov (GDEX) 39	
2.3.1 Posebne programske nastavitve	39
2.3.2 Vsebinske nastavitve pri izdelavi leksikalne baze za slovenščino	40
2.3.3 Kliksikografija (<i>TBL, Tick-box lexicography</i>)	40
2.3.4 Izbor dobrih zgledov (<i>GDEX, Good examples</i>).....	41
2.3.5 Prenos zgledov v odložišče	42
2.4 Napredne funkcije orodja SkE: <i>Corpus Builder</i>	44
2.4.1 Uvod	44
2.4.2 Predstavitev modula za razvoj slovnicih relacij v vzorčnem korpusu (<i>Corpus Builder</i>).....	44
2.5 Slovnične relacije v slovnici Besednih skic (SBS), v.15.....	46
2.5.1 Osnove sintakse slovnicih relacij	46
2.5.2 Primer slovnične relacije	47
2.5.3 Slovnične relacije v SBS v.15	51
2.5.4 Slovnične relacije po besednih vrstah s primeri iz korpusa (v.15)	62
2.5.5 Dokument s slovnicih relacijami: FidaPLUS sketch grammar ver. 15.....	64
2.6 Izrazje, povezano z orodjem SkE	72
3 Pretvornik za konkordance in kolokacije.....	73
3.1 Uvod.....	74
3.2 Namestitvev	74
3.3 Opis in uporaba.....	75
3.3.1 Prvi postopek: pretvorba konkordanc iz orodja SkE beta, funkcija Konkordance (<i>Concordance</i>)	75
3.3.2 Drugi postopek: pretvorba kolokacij iz orodja SkE beta, funkcija Besedna skica (<i>Word Sketch</i>)	77
3.3.3 Tretji postopek: pretvorba konkordanc iz konkordančnika ASP32 v Excel.....	78
4 Programska oprema za sestavljanje slovarjev: <i>Dictionary Production System</i>	80

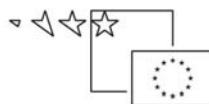


4.1	Uvod	81
4.1.1	Uvodno o programski opremi DPS	81
4.1.2	Profil proizvajalca DPS: podjetje IDM	82
4.2	Opis programske opreme DPS	83
4.2.1	Uvod	83
4.2.2	Celostno delovanje programa	83
4.2.3	Vrste uporabnikov	84
4.2.4	Delovni postopek	85
4.2.5	Projektna konfiguracija	85
4.2.6	Uporabniške pravice	85
4.3	DPS vmesnik	86
4.3.1	Namestitev DPS vmesnika, namestitev konfiguracije, zagon DPS vmesnika in dostop do gradiva	86
4.3.2	Programsko okno DPS vmesnika	87
4.3.3	Gesla v paketih in zunaj njih	89
4.3.4	Sprotno shranjevanje gesel, vračanje gesel na strežnik in status gesel	90
4.3.5	Glavni meniji in podmeniji v DPS vmesniku	93
4.3.6	Gumbi v drevesni strukturi (3. okno) ter priročni meni	95
4.3.7	Urejanje gesla	96
4.3.7.1	Osnovna navigacija po 3. oknu (<i>Editing an Entry</i>)	96
4.3.7.2	Izbor šablone (<i>Template Picker</i>)	100
4.3.7.3	Išči in zamenjaj (<i>Search and Replace</i>)	101
4.3.7.4	Validacija (<i>Highlight DTD errors, Report DTD errors</i>)	102
4.3.7.5	Opombe in atributi (<i>Annotations, Attributes</i>)	103
4.3.7.6	Modul za primerjanje različic gesla (<i>Version manager</i>)	104
4.4	DPS strežnik	107
4.4.1	Projekti (<i>Projects</i>)	108
4.4.2	Profili (<i>Profiles</i>)	108
4.4.3	Uporabniki (<i>Users</i>)	109
4.4.4	Konfiguracija (<i>Configuration</i>)	110
4.4.5	Iskanje (<i>Search</i>)	110
4.4.6	Delovni postopek (<i>Workflow</i>)	111
4.4.7	Tehnični postopki (<i>Processes</i>)	114
4.4.8	Zahtevno iskanje (<i>SkXml Box</i>)	115
4.4.9	Statistika (<i>Statistics</i>)	116
4.5	Izrazje, povezano s programom DPS	118
5	Format zapisa in DTD pri leksikalni bazi za slovenščino	120
5.1	XML: razširljiv označevalni jezik (<i>Extensible Markup Language</i>)	121
5.2	Osnovni pojmi in interpretiranje DTD-ja	121
5.3	DTD pri leksikalni bazi za slovenščino	123
5.4	Primer gesla	124
5.4.1	Izvoz iz programa DPS v obliki XML	124
5.4.2	Izvoz iz programa DPS v obliki PDF oziroma predogled tiskanja	126
6	Viri in spletne povezave	129
6.1	Viri	130
6.2	Spletne povezave	131

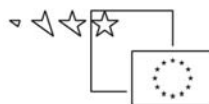


Kazalo slik:

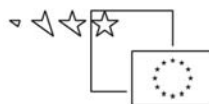
Slika 1: osnovno okno orodja SkE beta s povezavami na besedilne korpuse, module, nastavitve in pomoč ..	11
Slika 2: osnovne funkcije orodja SkE.....	11
Slika 3: funkcija za iskanje konkordanc v orodju SkE z možnostmi iskanja po podatkih o ključnih besedah, sobesedilu in besedilu (Keywords, Context, Text Types)	13
Slika 4: prikaz rezultatov iskanja konkordanc v orodju SkE, lema igrati; na desni so s počrnjenimi znaki označene konkordance, po tem ko jih izberemo za prenos v odložišče (SkE beta)	14
Slika 5: opis atributov za iskanja v orodju SkE	15
Slika 6: primer iskanja v poizvedbenem jeziku CQL v orodju SkE	15
Slika 7: rezultat iskanja v poizvedbenem jeziku CQL v orodju SkE	15
Slika 8: iskanje s podatki o sobesedilu v orodju SkE.....	16
Slika 9: rezultat iskanja s podatki o sobesedilu v orodju SkE.....	16
Slika 10: možnost iskanja s podatki o besedilu iz glave korpusnih dokumentov v orodju SkE	17
Slika 11: možnosti urejanja konkordanc v orodju SkE.....	17
Slika 12: okno z možnostmi prikaza (View options) v orodju SkE.....	18
Slika 13: različne možnosti prikaza konkordanc v orodju SkE.....	19
Slika 14: okno za izdelavo vzorca v SkE	19
Slika 15: možnosti pri nastavitvi filtrov konkordanc v orodju SkE	20
Slika 16: rezultat filtriranja v orodju SkE, vključujoči filter	20
Slika 17: rezultat filtriranja v orodju SkE, izključujoči filter.....	20
Slika 18: rezultat osnovnega razvrščanja po desni strani v orodju SkE	21
Slika 19: primer nastavitve za Preprosto razvrščanje oblik iskane leme v orodju SkE	21
Slika 20: rezultat Preprostega razvrščanja oblik leme stisniti v orodju SkE	21
Slika 21: primer nastavitve za Večnivojsko razvrščanje v orodju SkE, predhodno iskanje po lemi brcniti ..	22
Slika 22: rezultat Večnivojskega razvrščanja v orodju SkE	22
Slika 23: okno za nastavitve v funkciji Pogostost (Frequency) v orodju SkE	23
Slika 24: osnovna nastavitve pri Večnivojski distribuciji pogostosti v orodju SkE	24
Slika 25: rezultat osnovne nastavitve pri Večnivojski distribuciji pogostosti – enako kot funkcija Node forms, v orodju SkE (primer 1)	25
Slika 26: nastavitve za prikaz dveh nivojev pri Večnivojski distribuciji pogostosti.....	26
Slika 27: rezultat iskanja po dveh nivojih pri Večnivojski distribuciji pogostosti v orodju SkE (primer 2) .	26
Slika 28: nastavitve za prikaz treh nivojev pri Večnivojski distribuciji pogostosti za lemi ocenjevanje in vino v orodju SkE	27
Slika 29: rezultat iskanja po treh nivojih pri Večnivojski distribuciji pogostosti v orodju SkE (primer 3a) .	28
Slika 30: nastavitve za prikaz treh nivojev pri Večnivojski distribuciji pogostosti za lemo glodati v orodju SkE	29
Slika 31: rezultat iskanja po treh nivojih pri Večnivojski distribuciji pogostosti v orodju SkE (primer 3b) .	29
Slika 32: nastavitve za prikaz Distribucije pogostosti glede na besedilo v orodju SkE	30
Slika 33: rezultat Distribucije pogostosti glede na besedilo: leto in prenosnik.....	30
Slika 34: distribucija pogostosti oblikoskladenjskih oznak pri lemi glodati (Node tags).....	31
Slika 35: nastavitve v funkciji Kolokacije v orodju SkE (Collocation candidates)	31
Slika 36: primer nastavitve za izdelavo seznama možnih kolokacij v orodju SkE (funkcija Collocation)	32
Slika 37: seznam možnih kolokatorjev zveze ocenjevanje vin v razponu do tri mesta levo od leme ocenjevanje, v orodju SkE.....	32
Slika 38: okno z nastavitvami za izdelavo Besedne skice v orodju SkE beta	34



Slika 39: predstavitev Besedne skice (Word Sketch) v orodju SkE	35
Slika 40: gručenje kolokacij leme rana v orodju SkE	35
Slika 41: iskalno okno za izdelavo Primerjalne skice v orodju SkE	36
Slika 42: lestvica izračunov pri lemah, udeleženi v Primerjalni skici v orodju SkE	37
Slika 43: Primerjalna skica lem situacija in položaj: nekatere od primerjanih slovničnih relacij	37
Slika 44: Primerjalna skica lem situacija in položaj: nekatere od slovničnih relacij, v katerih nastopa lema položaj, lema situacija pa ne	38
Slika 45: posebne programske nastavitve v SkE beta: Besedna skica (Word Sketch)	39
Slika 46: posebne programske nastavitve v SkE beta: Možnosti prikaza (View options)	40
Slika 47: vsebinske nastavitve v Besedni skici kot izhodišče pri izdelavi leksikalne baze za slovenščino	40
Slika 48: označitev kolokatorjev (TBL) v Besedni skici za uporabo funkcije Izbor dobrih zgledov (GDEX) v orodju SkE beta	41
Slika 49: avtomatsko izdelan izbor dobrih zgledov (GDEX) v orodju SkE	42
Slika 50: osnovno okno modula Corpus builder	44
Slika 51: funkcije v modulu Corpus builder	45
Slika 52: primer slovnične relacije v orodju SkE	47
Slika 53: interpretacija definicije slovnične relacije v orodju SkE	47
Slika 54: vzorčna slovnična relacija v Besedni skici	48
Slika 55: rezultat slovnične relacije v_rodil-s	49
Slika 56: rezultat slovnične relacije s-koga-česa	49
Slika 57: podatki o slovničnih relacijah v SBS	50
Slika 58: preveritev, ali je .net Framework nameščen	74
Slika 59: programsko okno Pretvornika za konkordance in kolokacije	75
Slika 60: okno za shranjevanje konkordanc v orodju SkE beta	76
Slika 61: rezultat prvega postopka pretvorbe s pretvornikom	76
Slika 62: okno za shranjevanje Besedne skice v orodju SkE beta	77
Slika 63: rezultat drugega postopka pretvorbe s pretvornikom (izsek tabele)	77
Slika 64: okno za shranjevanje konkordanc v konkordančniku ASP32	78
Slika 65: rezultat tretjega postopka pretvorbe s pretvornikom	79
Slika 66: Program DPS: celostno delovanje programa: komunikacija med posameznimi deli programa in uporabniki	84
Slika 67: mesto, kamor uporabnik namesti konfiguracijo za DPS vmesnik	86
Slika 68: DPS vmesnik: prijava na DPS strežnik	87
Slika 69: DPS vmesnik: dostop do paketov na DPS strežniku (gumb Home)	87
Slika 70: ikone pred iztočnico: pravica do urejanja, skladnost z DTD, sinhroniziranost z DPS strežnikom (1. okno)	88
Slika 71: DPS vmesnik: deli programskega okna	89
Slika 72: DPS vmesnik, 1. okno, del z gumbi za upravljanje z gesli v paketu	90
Slika 73: statusi gesla (Entry Status)	92
Slika 74: statusi gesla v povezavi s komunikacijo med DPS vmesnikom in DPS strežnikom	92
Slika 75: glavni meniji in podmeniji v DPS vmesniku	93
Slika 76: opis podmenijev v DPS vmesniku	95
Slika 77: DPS vmesnik: gumbi v drevesni strukturi (3. okno)	95
Slika 78: DPS vmesnik: gumbi v priročnem meniju (desni klik)	96
Slika 79: režim pomikanja po geslu v DPS vmesniku	97
Slika 80: prikaz funkcije Osveži element (Refresh element)	98
Slika 81: razloženi in strnjeni izgled strukture elementov (Swap inline state)	99



<i>Slika 82: menjanje v strnjeni izgled ob vnosu elementa, ki lahko vsebuje besedilo in podrejene elemente (Swap inline state)</i>	<i>99</i>
<i>Slika 83: okno Izbor šablone z opisom funkcij</i>	<i>100</i>
<i>Slika 84: okno Išči/zamenjaj v DPS vmesniku</i>	<i>101</i>
<i>Slika 85: način delovanja validacije v DPS vmesniku</i>	<i>103</i>
<i>Slika 86: urejanje atributov in vnos opomb v 4. oknu DPS vmesnika</i>	<i>103</i>
<i>Slika 87: osnovno okno modula za primerjanje različic gesla (Version manager) v DPS vmesniku</i>	<i>105</i>
<i>Slika 88: primerjava dveh različic gesla igrati: funkcija Diffence v modulu za primerjanje različic gesla (Version manager)</i>	<i>105</i>
<i>Slika 89: oznaka arhivske različice gesla, ko jo odpremo v DPS vmesniku</i>	<i>106</i>
<i>Slika 90: DPS strežnik: okno za vnos podatkov za dostop</i>	<i>107</i>
<i>Slika 91: DPS strežnik: razdelki za upravljanje s posameznimi podatkovnimi bazami (projekti) v programu DPS</i>	<i>107</i>
<i>Slika 92: DPS strežnik: Projekti</i>	<i>108</i>
<i>Slika 93: pregled uporabniških pravic pri profilu FLBrainiac</i>	<i>109</i>
<i>Slika 94: DPS strežnik: podatki o uporabniku in uporabniških pravicah pri posameznem projektu</i>	<i>109</i>
<i>Slika 95: DPS strežnik: konfiguracija za izdelavo statistike</i>	<i>110</i>
<i>Slika 96: DPS strežnik: Iskanje (Search)</i>	<i>110</i>
<i>Slika 97: primer opredelitve delovnega postopka, posameznih postopkov v njem ter razdelitev na naloge</i>	<i>111</i>
<i>Slika 98: potek dela in razvrščanje gradiva pri projektu v programu DPS</i>	<i>112</i>
<i>Slika 99: DPS strežnik: osnovno okno za organiziranje poteka dela in razvrščanje gradiva (Workflow)</i>	<i>113</i>
<i>Slika 100: prikaz poteka dela v paketu: statusna pasica</i>	<i>114</i>
<i>Slika 101: prikaz poteka dela za vse pakete v posameznem koraku naloge (View task)</i>	<i>114</i>
<i>Slika 102: DPS strežnik: tehnični postopki za delo s podatkovno bazo</i>	<i>115</i>
<i>Slika 103: DPS strežnik: okno za nastavitve zahtevnega iskanja (SkXml Box); primer iskanja: element <zveza>, ki je podrejen elementu <skladenjska zveza>, izpis od <zveza> dalje (znak %)</i>	<i>115</i>
<i>Slika 104: DPS strežnik: prikaz rezultatov zahtevnega iskanja (SkXml Box), izgled Pretty Printed Xml</i>	<i>116</i>
<i>Slika 105: DPS strežnik: statistika, podatki o elementih</i>	<i>116</i>
<i>Slika 106: DPS strežnik: statistika, podatki o XML poteh</i>	<i>117</i>
<i>Slika 107: glava gesla pasti in prvi pomen s skladenjskimi zvezami, predogled v programu DPS</i>	<i>126</i>
<i>Slika 108: podpomen 1.2 glagola pasti, predogled v programu DPS</i>	<i>127</i>
<i>Slika 109: frazeološke zveze pri glagolu pasti, odlomek iz predogleda v programu DPS</i>	<i>127</i>
<i>Slika 110: idiomatske zveze pri glagolu pasti, predogled v programu DPS</i>	<i>128</i>



1

Uvod

*Programska
oprema
za analizo
referenčnega korpusa
in izdelavo
leksikalne baze
za slovenščino*



1 Uvod: programska oprema za analizo referenčnega korpusa in izdelavo leksikalne baze za slovenščino

V raziskovalni fazi izdelave leksikalne baze za pridobivanje podatkov iz besedilnega korpusa uporabljamo trenutno največji referenčni korpus slovenskega jezika z imenom FidaPLUS. Ta je prosto dostopen na naslovu <http://www.fidaplus.net/>, vgrajen pa je tudi v korpusno orodje Sketch Engine (<http://www.sketchengine.co.uk/>, gl. poglavje 2). Tekom projekta bo na voljo tudi novi korpus, ki nastaja pri projektu SSJ v skladu s predvideno projektno časovnico.

V kazalniku 5 – Opis analize referenčnega korpusa – opisujemo naslednje aplikacije:

- 1) orodje *Sketch Engine*, ki je namenjeno pridobivanju jezikovnih podatkov iz besedilnega korpusa FidaPLUS,
- 2) *Pretvornik za konkordance in kolokacije* lastne izdelave,
- 3) programsko opremo za sestavljanje slovarjev *Dictionary Production System*.

Uporabljene kratice v kazalniku 5 so:

SSJ	Sporazumevanje v slovenskem jeziku
DPS	<i>Dictionary Production System</i>
SkE	<i>SketchEngine</i>
SBS	slovnica besednih skic (<i>Sketch grammar</i>)
DTD	definicija tipa dokumenta (<i>Document Type Definition</i>)

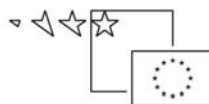
Opomba: pri opisu programske opreme tuje izdelave (gl. 2. in 4. poglavje) so ob slovenskih poimenovanjih navedena tudi angleška. Zaradi tega je orientacija v programski opremi in navodilih lažja, kajti programa nista lokalizirana za slovenski jezik. Uporabljena poimenovanja so tudi zbrana na koncu ustreznih poglavij.



2

Sketch Engine

*Programska
oprema
za
analizo
referenčnega
korpusa*



2 Programska oprema za analizo referenčnega korpusa: *Sketch Engine*

2.1 Uvod

Sketch Engine (dalje: SkE) je orodje za prikaz in analizo jezikovnih podatkov v besedilnem korpusu. Orodje je razvilo podjetje Lexical Computing Ltd. za analizo besedilnega korpusa, uporabljenega pri sestavljanju slovarja *Macmillan English Dictionary* (Rundell 2002, Rundell 2007). V začetku je bilo orodje izdelano le za angleški jezik (Kilgarriff in Rundell 2002), nato pa so ga zaradi velikega povpraševanja stroke začeli nadgrajevati tudi za korpusne drugih jezikov (Kilgarriff et al. 2004), med drugim tudi za slovenščino (Krek in Kilgarriff 2006). Za korpus, ki ga želimo vključiti v to orodje, je potrebno sodelovanje tako razvijalcev orodja na strani proizvajalca kot ekip na strani uporabnikov, torej tistih, ki želijo svoj korpus analizirati s pomočjo SkE. Korpus je namreč treba tako tehnično dodatno pripraviti kot tudi izdelati specifikacijo slovničnih relacij, na kateri temeljijo bistvene funkcije orodja SkE; priprava tako združuje delo korpusnih in računalniških jezikoslovcev ter računalniških strokovnjakov.

Ciljni uporabniki orodja SkE so slovaropisne ekipe pri različnih založbah, raziskovalne ekipe pri jezikovnotehnoloških projektih, uporabljajo ga za poučevanje in raziskovanje na univerzah in v raziskovalnih centrih, med drugim so ga uporabili tudi pri samodejnem razdvoumljanju večpomenskih besed (*Word Sense Disambiguation*).

Orodje SkE obstaja v dveh različicah:

osnovni, dostopni na <http://www.sketchengine.co.uk/>, ter

beta, dostopni na <http://beta.sketchengine.co.uk/> ali na strani osnovne različice.

Različica beta, do katere dostopajo tudi sodelavci pri projektu SSJ, vsebuje več možnosti pri nastavitvah analiz, zlasti pa dodatne funkcije, ki omogočajo lažji prenos podatkov v programe za sestavljanje slovarjev.

Dostop do korpusov v orodju SkE sodelavcem pri projektu SSJ dodeli sistemski skrbnik.

V tem poglavju bomo okvirno predstavili možnosti analize korpusnih podatkov v orodju SkE, več o njegovem razvoju in lastnostih pa je na voljo npr. v Kilgarriff et al. 2004, za slovenščino pa v Krek in Kilgarriff 2006. Drugi podrobnejši dokumenti, ki so sicer navedeni tudi na ustreznih mestih pri opisu posameznih funkcij, so:

izvorna navodila:

<http://trac.sketchengine.co.uk/wiki/SkE/DocsIndex>,

sintaksa iskanj v poizvedbenem jeziku CQL in sintaksa slovničnih relacij:

<http://trac.sketchengine.co.uk/wiki/SkE/CorpusQuerying>,

seznam oblikoskladenjskih oznak za slovenščino:

<http://www.sketchengine.co.uk/tagsets/slovene.html>

vrste statistik v orodju SkE:

<http://trac.sketchengine.co.uk/wiki/SkE/DocsIndex>



Osnovno okno orodja SkE s seznamom korpusov je prikazano na naslednji sliki:

LEX COM Sketch Engine beta
Lexical Computing
user: Polonca Kocjančič

Help Change passwd WebBootCaT CorpusBuilder Bug reporting News Logout

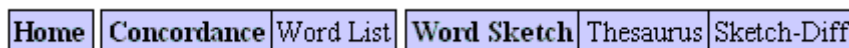
Corpora

Language	Name	Tokens [?]	
Chinese	Chinese GW, simpl	706 427 624	info
Chinese	Chinese GW, trd	706 428 333	info
English	British Academic Written English Corpus (BAWE)	8 336 262	info
English	British National Corpus	111 244 375	info
English	NCI English	257 900 780	
English	ukWaC	1 526 599 198	info
French	French web corpus	126 850 281	info
German	deWaC	1 644 785 836	info
Irish	NCI Irish	34 358 267	
Italian	itWaC	1 909 535 984	info
Japanese	JpWaC	409 384 405	info
Russian	Russian Web Corpus	187 965 822	info
Slovenian	Fida PLUS 620m	738 460 168	info
Spanish	Spanish web corpus	116 900 060	info

[more corpora >>](#)

Slika 1: osnovno okno orodja SkE beta s povezavami na besedilne korpuse, module, nastavitve in pomoč

Osnovne funkcije orodja SkE so:



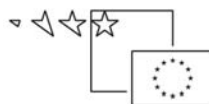
Slika 2: osnovne funkcije orodja SkE

Za vsebinsko delo z izbranim korpusom so predvidene:

- 1) Konkordance (*Concordance*),
- 2) Seznam besed (*Word List*),
- 3) Besedna skica (*Word Sketch*),
- 4) Tezaver (*Thesaurus*),
- 5) Primerjalna skica (*Sketch-Diff*).

V kazalniku 5 opisujemo tiste dele orodja SkE, ki jih tudi konkretno uporabljamo pri sestavljanju leksikalne baze za slovenščino:

- **osnovne funkcije orodja SkE:** Konkordance, Besedne skice in Primerjalne skice,
- **dodatne funkcije v različici SkE beta:** Kliksikografija in Izbor dobrih zgledov (TBL in GDEX),

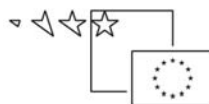


- **modul za razvoj in dopolnjevanje slovničnih relacij v vzorčnem korpusu** (*Corpus Builder*),
- **razvojno delo na področju slovničnih relacij** (*gramrel*).

Funkcij in modulov, ki jih za delo pri leksikalni bazi ne uporabljamo neposredno, v tem dokumentu ne opisujemo. To so:

- funkcija Seznam besed (*Word List in Find X*),
- funkcija Tezaver (*Thesaurus*),
- modul *WebBootCaT*, ki je namenjen izdelavi lastnega korpusa na podlagi ključnih besed in pridobivanju strokovnega izrazja s spleta.

Za njihovo uporabo napotujemo na podrobna izvorna navodila (<http://trac.sketchengine.co.uk/wiki/SkE/DocsIndex>).



2.2 Osnovne funkcije orodja SkE: Konkordance, Besedne skice, Primerjalne skice

V tem poglavju so opisane tri od osnovnih funkcij orodja SkE:

- avtomatično pridobivanje kolokacij (*Concordance*),
- izdelava Besednih skic (*Word sketch*),
- izdelava Primerjalnih skic dveh lem (*Sketch-Diff*).

2.2.1 Konkordance (*Concordance*)

2.2.1.1 Izdelava seznama konkordanc

Iskalno okno za izdelavo seznama konkordanc:

Slika 3: funkcija za iskanje konkordanc v orodju SkE z možnostmi iskanja po podatkih o ključnih besedah, sobesedilu in besedilu (Keywords, Context, Text Types)

Osnovno iskanje vpišemo v okence *Query*; če je korpus označen in lematiziran, kar FidaPLUS je, iskalnik upošteva lemo. Rezultat osnovnega iskanja prikazuje naslednja slika:



Home | Concordance | Word List | Word Sketch | Thesaurus | Sketch-Diff
View options | Sample | Filter | Sort | Frequency | Collocation | Save

Corpus: Fida PLUS 620m
Hits: 224472
[conc description](#)

First | Previous | Page 25 of 11224 | Go | Next | Last

časopis	skupinah. " </p><p> Nemaški predstavnik: Werder iz Bremna bo	igral	proti švicarskemu Baslu, francoski Monaco, podprvak lige prvakov	
časopis	špansku Betisu. Več sreče je imel škotski prvak Rangers, ki bo	igral	proti Anorthosis Famagustu iz Cipa ali turskemu Trabzonsporu	
časopis	kar z 0:5. Steaua iz Bukarešte, evropski prvak leta 1986, bo	igrala	proti rednemu udeležencu lige prvakov v zadnjih letih Rosenborgu	
časopis	pričakovati, da nas bo tekmeec povsem razbil. Menim, da smo	igrali	proti najboljšemu klubu na svetu in dokazali, da lahko igramo	
časopis	igrali proti najboljšemu klubu na svetu in dokazali, da lahko	igramo	tudi proti tako dobrim ekipam. " </p><p> * * * adu * * * </p>	
časopis	nas bo tekmeec povsem razbil, " je dejal. " Po mojem smo	igrali	proti najboljšemu klubu na svetu in dokazali, da lahko igramo	
časopis	igrali proti najboljšemu klubu na svetu in dokazali, da lahko	igramo	tudi proti tako dobrim ekipam. " </p><p> Angleško moštvo	
časopis	zakrpati luknjo v osrednji obrambi. " Vedno sem si želel	igrati	za velik evropski klub, kar Werder zagotovo je. Tu si lahko	
časopis	hočem osvojiti naslov prvaka in poleti ostati v Nemčiji ter	igrati	za državno reprezentanco na SP, " je dejal Naldo. </p><p>	
revija	deset otrok, katerim je Bil je zelo predan oče, otroci so se	igrali	v njegovi delovni sobi in risali po hrbtnih straneh znanstvenih	
revija	ustnic nakažejo smeh. Vloge niso ločene. Tudi ženske lahko	igrajo	moške vloge, razen pri nekaterih likih (npr. Hanuman ali razni	
revija	Ključno vlogo pri prenosu tujih genov v bakterijsko celico	igrajo	tako imenovani plazmidi. To so krožne molekule DNK, ki se	
Dnevnik, časopis	leta 2001 prav tako 28. izbrancev lanskah prvakov. " Francoz	igra	drugče. Predvsem bolj agresivno, moji značilnosti pa sta	
Dnevnik, časopis	pogovorih mi je trener Popovich že omenil možnost, da bi lahko	igral	okrog 15 minut, čeprav je seveda vse odvisno le od mene. "	
Dnevnik, časopis	Udarec v koleno me je prikrajšal za trening, zato sem samo	igral	in na višji nadmorski višini mi je večkrat zmanjkalo zraka	
Dnevnik, časopis	Beno Udrih (drugi z leve) je zadovoljen, ker bo v novi sezoni	igral	v ligi NBA, kjer se bo pridružil Primožu Brezcu (levo) in	
Dnevnik, časopis	pogovorih mu je trener Popovich že omenil možnost, da bi lahko	igral	okrog 15 minut, čeprav je seveda vse odvisno le od njega. </p>	
Dnevnik, časopis	Prvi je reprezentant Rusije Dimitrij Stratan, ki je dve sezoni	igral	za obalnega prvotigaša Koper, takrat še kot Ukrajinec, kasneje	
Dnevnik, časopis	želeli zaradi tekme v evropskem pokalu UEFA v torek v Beogradu	igrati	že večeraj, vendar njihova prošnja ni bila uslišana, ker	
Dnevnik, časopis	času slabo skakal, in Matevžičević, ki zaradi poškodbe ni	igrala	pol leta, zavrnil strokovno mnenje in ju utvrdil med olimpijske	

First | Previous | Page 25 of 11224 | Go | Next | Last

Slika 4: prikaz rezultatov iskanja konkordanc v orodju SkE, lema igrati; na desni so s počrtnjenimi znaki označene konkordance, po tem ko jih izberemo za prenos v odložišče (SkE beta)

Osrednji del zaslona prikazuje zadetke z nastavljenimi količinami sobesedila, v levem stolpcu je podan vir. Desno zgoraj so navedeni podatki o izbranem korpusu, številu zadetkov in iskanju. V skrajnem desnem stolpcu so ob konkordancah ikone, vgrajene kot dodatna možnost v SkE beta, namenjene prenosu celostavnih konkordanc v odložišče.

2.2.1.2 Opis možnosti v iskalnem oknu za izdelavo seznama konkordanc

2.2.1.2.1 Iskanje s podatki o ključnih besedah (Keywords)

Iskanje v korpusu je možno natančneje opredeliti:

- po lemi in z možnostjo določitve besedne vrste, npr. *dolg* + *samostalnik*, *dolg* + *pridevnik*,
- po večbesedni zvezi, pri čemer lematizacija ni upoštevana, npr. *in sicer*,
- po konkretni obliki in z možnostjo določitve besedne vrste, npr. *dolga* + *samostalnik*,
- v poizvedbenem jeziku CQL (*Corpus query language*) – primer gl. spodaj.

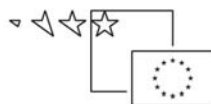
Sintaksa iskanj v CQL je opisana na:

<http://trac.sketchengine.co.uk/wiki/SkE/CorpusQuerying>.

Dodatne informacije o CQL so na voljo na:

<http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/> in

<http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/CQPSyntax.html>.



Spustni meni pri *Default attribute* določuje, na kakšen način bodo razumljena iskanja v CQL:

Atribut	Pomen atributa	Primer	Pomen primera
<i>word</i>	beseda oziroma njena oblika	word="pod"	išče besedo <i>pod</i>
<i>tag</i>	oblikoskladenjska oznaka (MSD)	tag="S.*"	išče vse samostalnike
<i>lempos</i>	lema v povezavi z besedno vrsto	lempos="dolg-s"	išče lemo <i>dolg</i> kot samostalnik
<i>lemma</i>	lema	lemma="pasti"	išče lemo <i>pasti</i>
<i>lc</i>	mala začetnica	lc="loka"	išče besedo <i>loka</i> z malo začetnico
<i>lemma_lc</i>	lema z malo začetnico	lemma_lc="lok"	išče lemo <i>lok</i> z malo začetnico

Slika 5: opis atributov za iskanja v orodju SkE

Seznam oblikoskladenjskih oznak za slovenščino je dostopen s klikom na povezavo *Tagset summary* (<http://www.sketchengine.co.uk/tagsets/slovene.html>).

Primer iskanja v CQL: `lemma="pasti"[word="pod"]]{0,2}[lemma="stopinja"]`

The screenshot shows the SkE search interface with the following elements:

- Navigation tabs: Home, Concordance, Word List, Word Sketch, Thesaurus, Sketch-Diff
- Corpus: Fida PLUS 620m
- Query input field: []
- Keyword(s) section:
 - Lemma: [] PoS: unspecified
 - Phrase: []
 - Word Form: [] PoS: unspecified Match case:
 - CQL: `lemma="pasti"[word="pod"]]{0,2}[lemma="stopinja"]`
 - Default attribute: word Tagset summary
- Context and Text Types sections (expanded)
- Buttons: Make Concordance, Clear All

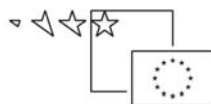
Slika 6: primer iskanja v poizvedbenem jeziku CQL v orodju SkE

Rezultat iskanja v CQL:

The screenshot shows the SkE search results page with the following elements:

- Navigation tabs: Home, Concordance, Word List, Word Sketch, Thesaurus, Sketch-Diff
- View options: View options, Sample, Filter, Sort, Frequency, Collocation, Save
- Page 1 of 7
- Corpus: Fida PLUS 620m Hits: 126
- Concordance results showing text snippets with highlighted search terms and frequency counts (e.g., "pade pod -20 stopinj", "pade pod 0 stopinj", "pade pod 5 stopinj", "pade pod +8 stopinj", "pade pod 50 stopinj", "pade pod 18 stopinj", "padla pod -2 stopinji", "padejo pod sedem stopinj", "padla pod 35 stopinj", "padla pod 35 stopinj", "pade pod 10 stopinj", "pade pod 18 stopinj", "pade pod -20 stopinj", "padla pod 0 stopinj", "pade pod minus 15 stopinj").

Slika 7: rezultat iskanja v poizvedbenem jeziku CQL v orodju SkE



2.2.1.2.2 Iskanje s podatki o sobesedilu (*Context*)

Iščemo lahko po lemah, ki so od osnovne iskane besede oddaljene do 15 mest v levo ali desno; najpogosteje se sobesedilo sicer analizira do pet besed levo ali desno. Kot sobesedilo lahko določimo bodisi konkretno besedo ali pa eno ali več besednih vrst, z izbiro *None* v spustnem meniju *Query Type* pa lahko določene podatke tudi izključimo iz iskanja.

Primer: *zmajati* (glagol) + *glava* (samostalnik) + v oddaljenosti treh mest

Možnosti za enak rezultat sta dve:

- vtipkamo *zmajati* v okno za iskanje po lemi (*Lemma* v nastavitvah *Keyword(s)*) in določimo besedno vrsto. V okno *Right context* vtipkamo lemo *glava* in določimo iskalno razdaljo (3 tokens).

- vtipkamo *glava* v okno za iskanje po lemi (*Lemma* v nastavitvah *Keyword(s)*) in določimo besedno vrsto. V okno *Left context* vtipkamo *zmajati* in določimo iskalno razdaljo (3 tokens).

The screenshot shows the SKE search interface with the following settings:

- Corpus: Fida PLUS 620m
- Query: (empty)
- Keyword(s):
 - Lemma: PoS:
 - Phrase: (empty)
 - Word Form: (empty) PoS: Match case:
 - CQL: (empty)
 - Default attribute: [Target summary](#)
- Context:
 - Query Type: of these items.
 - Window Size:
 - Left context: tokens.
 - Right context: tokens.
 - Lemma:
 - PoS:
 - Left context: (dropdown menu open showing: samostalnik, glagol, pridevnik, prislov)
 - Right context: (dropdown menu open showing: samostalnik, glagol, pridevnik, prislov)
- Text Types: (dropdown menu open)

Slika 8: iskanje s podatki o sobesedilu v orodju SKE

Rezultat iskanja s podatki o sobesedilu:

The screenshot shows the search results page with the following information:

- Corpus: Fida PLUS 620m
- Hits: 1908
- conc. description
- Page: 5 of 96

The concordance results are as follows:

grozen , zoprn , vsiliv. — Kar goriš , sine , prav potihem pa **zmaje** z **glavo** , ful zaskrbleno faco ma . Eno roko potegnem izpod grozen , zoprn , vsiliv. — Kar goriš , sine , prav potihem pa **zmaje** z **glavo** , ful zaskrbleno faco ma . Eno roko potegnem izpod se je previdno dotaknil njenih desnih senc . " Tukaj ? " **Zmajala** je z **glavo** . " Tukaj ? Tukaj ? " " Ja . " Malo levo gradbenega podjetja , privlačen soprog kričave Susie . " **Zmajala** je z **glavo** in odvila papir s čokoladne ploščice , kot bi , ker ne vedo , ali so primerno oblečeni ! To je neumnost ! **zmaje** z **glavo** . Če pomisli , je tako najbrž tudi zato , ker je Prebral ga je v drži njenih ramen , v črti stisnjenih ustnic . **Zmajala** je z **glavo** . " Se dobro počutiš ? " je vprašal tiho Elizabeth . Dane se ji je zarežal . " Prebolela boš . " **Zmajala** je z **glavo** . " Ne verjamem . Ne potrebujem težav . In razočarana , ker se je romantična zgodbica sesula v prah . **Zmajal** je z **glavo** . » Oprosti , ker sva ti lagala , toda z mano oče . » Kajne , da boš ? « se je proseče obrnila k meni . **Zmajala** sem z **glavo** . » Zal mi je , Jasna , ampak tisti kraj sovražim na vprašanja , katerih lepota je , da nanje ni odgovorov ? **Zmajal** je z **glavo** . Ne . Za hip se je razlegla tišina . Ko je komornik oče dal križati . Daj no , Anthony . To je krasna zamisel . **Zmajal** je z **glavo** . Lepa , a ne praktična . Kje boš našel 458 natančen , bi se lahko tako reklo , « sem se diplomatsko izmikal . **Zmajal** je z **glavo** . » In jaz sem že hotel poklicati zdravnika skušajte se spomniti . Kje bi lahko še skrtil te reči ? « **Zmajala** je z **glavo** . » Ne vem . « » Gotovo je predvidel , da se priljubljenost , iz vsega skupaj je nastala čudna godlja . **Zmajala** je z **glavo** . Njeni možgani so mleli prepočasi . Vse te možganska kap , sam sliš se koker tipična Mattyjeva storija . **Zmajem** z **glavo** . — To krasni povzame Mattyja . Nabav muco koker možganska kap , sam sliš se koker tipična Mattyjeva storija . **Zmajem** z **glavo** . — To ti je bil taprav Matty . Nabav muco koker tavžent ... pa še zmer tud bliz ne boš . Obzalujoče sem **zmajal** z **glavo** . Pokazal sem mu sliko sebe s Frances . Pozirala

Slika 9: rezultat iskanja s podatki o sobesedilu v orodju SKE



2.2.1.2.3 Iskanje s podatki o besedilu (Text Types)

Med vzpostavljanjem tehnoloških pogojev dela pri projektu SSJ smo v orodju SkE natančneje opredelili lestvico možnosti, ki jih korpus FidaPLUS omogoča s podatki v glavi dokumentov, kar uporabljamo v raziskovalni fazi izdelave leksikalne baze, dokler ne bo na voljo novi korpus, ki nastaja pri projektu SSJ. Možnosti za natančnejšo določitev iskanj s podatki o besedilu kot tudi za ustvarjanje trajnejših podkorpusov so prikazane na naslednji sliki:

New subcorpus name:

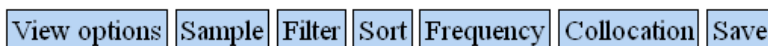
[Document counts](#) [Tokens](#)

doc. leto	#	doc. prenosnik	#	doc. zvrst	#	doc. posamezno	#
<input type="checkbox"/>	10,326,254	<input type="checkbox"/> govorni	2,890,082	<input type="checkbox"/> leposlovje	26,838,888	<input type="checkbox"/>	312,301,593
<input type="checkbox"/> 0000	1,188	<input type="checkbox"/> internet	9,960,133	<input type="checkbox"/> neoznačeno	841,497	<input type="checkbox"/> DZSeje	2,441,338
<input type="checkbox"/> 0200	6,253,016	<input type="checkbox"/> knjiga	66,556,898	<input type="checkbox"/> stvarna besedila	710,779,783	<input type="checkbox"/> DZZakon	2,985,593
<input type="checkbox"/> 1979	992	<input type="checkbox"/> neoznačeno	8,428,947	<input type="button" value="Select All"/>		<input type="checkbox"/> Delo	136,527,779
<input type="checkbox"/> 1980	239,937	<input type="checkbox"/> revija	172,401,482			<input type="checkbox"/> Dnevnik	156,008,309
<input type="checkbox"/> 1989	40,861	<input type="checkbox"/> časopis	478,222,626			<input type="checkbox"/> DoleList	3,390,513
<input type="checkbox"/> 1990	33,496	<input type="button" value="Select All"/>				<input type="checkbox"/> FP splet	9,931,228
<input type="checkbox"/> 1991	1,768,146					<input type="checkbox"/> Finance	1,306,862
<input type="checkbox"/> 1992	2,669,333					<input type="checkbox"/> GlasGos	2,872,263
<input type="checkbox"/> 1993	3,810,572					<input type="checkbox"/> GolfSlov	1,334,196
<input type="checkbox"/> 1994	8,937,294					<input type="checkbox"/> Hopla	7,854,880
<input type="checkbox"/> 1995	8,830,102					<input type="checkbox"/> Joker	2,744,819
<input type="checkbox"/> 1996	20,113,640					<input type="checkbox"/> Kapital	726,054
<input type="checkbox"/> 1997	37,741,946					<input type="checkbox"/> MBčan	1,772,628
<input type="checkbox"/> 1998	51,645,865					<input type="checkbox"/> Mag	6,874,797
<input type="checkbox"/> 1999	65,182,225					<input type="checkbox"/> MarkMag	1,130,430
<input type="checkbox"/> 2000	68,312,802					<input type="checkbox"/> Motorevija	347,329
<input type="checkbox"/> 2001	87,978,774					<input type="checkbox"/> Navtika	1,812,762
<input type="checkbox"/> 2002	85,961,072					<input type="checkbox"/> NedDnev	18,467,306
<input type="checkbox"/> 2003	97,811,398					<input type="checkbox"/> NovTed	2,692,734
<input type="checkbox"/> 2004	79,464,857					<input type="checkbox"/> Pilot	1,610,309
<input type="checkbox"/> 2005	46,857,226					<input type="checkbox"/> Podjetnik	2,582,791
<input type="checkbox"/> 2006	54,479,172					<input type="checkbox"/> Premiera	1,781,589
<input type="button" value="Select All"/>						<input type="checkbox"/> PrimNov	4,572,073
						<input type="checkbox"/> Vestnik	6,404,058
						<input type="checkbox"/> Večer	39,716,607
						<input type="checkbox"/> ZlatiKap	499,128
						<input type="checkbox"/> ŠtajTedn	5,274,735
						<input type="checkbox"/> ŽIT	968,084
						<input type="checkbox"/> Žurnal	1,527,381
						<input type="button" value="Select All"/>	

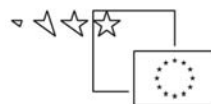
Slika 10: možnost iskanja s podatki o besedilu iz glave korpusnih dokumentov v orodju SkE

2.2.1.3 Prikaz in obdelava konkordanc

Ko imamo pred seboj zaslon s konkordancami, jih lahko tudi na različne načine obdelamo, kar je ključno za uspešno semantično in sintaktično analizo gesel. Možnosti nadaljnje obdelave konkordanc so prikazane v 2. vrsti gumbov v orodju SkE:



Slika 11: možnosti urejanja konkordanc v orodju SkE



2.2.1.3.1 Možnosti prikaza (View options)

Funkcija omogoča naslednje nastavitve prikaza konkordanc:

- v obliki stolpca z zadetkom na sredini in v celostavčni obliki (KWIC/Sentence),
- dodajanje prikaza atributov (Attributes): z oblikoskladenjskimi oznakami, lemmami ali drugimi podatki, bodisi samo za zadetek ali pa za vse besede v konkordanci,
- dodajanje prikaza začetnih in končnih kod pri strukturnih elementih (Structures), kot so stavki, odstavki in dokumenti,
- določitev, kateri podatki o viru konkordance bodo prikazani v skrajno levem stolpcu (References),
- določitev števila konkordanc na strani (Page Size box); gornja meja je 500, kar pa vpliva na hitrost odziva pri prikazu prvega zaslona zadetkov,
- določitev širine konkordanc.

V SkE beta so dodane še možnosti za urejanje Izbora dobrih zgledov (Sort good dictionary examples, Number of lines to be sorted, Icon for one-click sentence copying, Allow multiple lines selection in XML template for one-click copying); specifične nastavitve za uporabnike pri projektu SSJ so opisane v poglavju o dodatnih funkcijah v SkE beta.

Okno z možnostmi prikaza:

View options

Attributes	Structures	References
<input checked="" type="checkbox"/> word	<doc>	Token number
<input checked="" type="checkbox"/> tag	<p>	doc.posamezno
<input checked="" type="checkbox"/> lempos	<s>	doc.catref1
<input type="checkbox"/> lemma	<g>	doc.catref2
<input type="checkbox"/> lc		doc.catref3
<input type="checkbox"/> lemma_lc		doc.datum
		doc.letno
		doc.prenosnik
		doc.zvrst

Display attributes
 For each token
 KWIC tokens only

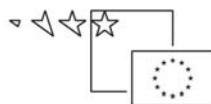
Page size (number of lines):
KWIC Context size (number of characters):

Sort good dictionary examples. Number of lines to be sorted:

Icon for one-click sentence copying
 Allow multiple lines selection

XML template for one-click copying:

Slika 12: okno z možnostmi prikaza (View options) v orodju SkE



Gornje nastavitve dajo naslednji prikaz:

Page 1 of 78 Go Next Last

NedDnev,2002	iz kraja , kjer lahko čez ograjo gledate , kako " rja	gloda /Gppste-n-----n/glodati-g	arzenal " , kot ste dejali .	</s><s> Za začetek bi dejal	
2003	<s> Nekje globoko v njem se je zganila skrb in ga začela	glodati /Gpn-----n/glodati-g	.	</s><s> » Martinsson , « je rekel počasi , » se mogoče	
Dnevnik,1996	kulinaričnih dogodkov .	gloda /Gppste-n-----n/glodati-g	proračunske vreče in žepov davkoplačevalcev , bodo		
Delo,1999	spregovoril gospod nadškof .	glodali /Gpdr-dst-----n/glodati-g	časniki (in ne samo oni) v časopisih , na radiu ,		
2005	prav taki ste tudi sami .	gloda /Gppste-n-----n/glodati-g	dvom , zato ste tudi poslali vprašanje .	</s><s> Hčerka	
1992	pomembna .	glodati /Gpn-----n/glodati-g	črv zavisti in strahu , da o njej mogoče nikoli ne bodo		
2005	Janeza , inajo kost za glodanje .	glodajo /Gppstm-n-----n/glodati-g	, se ne posvečajo s tolikšnim žarom svojemu » poslanstvu		
2005	sem , ne pa tajni agent , madona !	glodati /Gpn-----n/glodati-g	, zakaj tu Strathmore poslal strokovnjaka .	</s><s> Vstal	
2000	nisem bila čisto prepričana .	glodalo /Gpdr-est-----n/glodati-g	to , ali bo konec tega mojega filmskega življenja srečen		
Hopla,2006	podrobnosti , naj bodo še tako očitne .	gloda /Gppste-n-----n/glodati-g	slaba misel , ko lahko komu nekaj očitate .	</s><s> Toda	

Page 1 of 78 Go Next Last

Slika 13: različne možnosti prikaza konkordanc v orodju SkE

2.2.1.3.2 Vzorec (Sample)

Funkcijo uporabimo, če želimo dobiti številčno zamejen vzorčni seznam konkordanc. Pri analizi leksikalne baze za slovenščino uporabljamo od 150 do 350 zadetkov za posamezno lemo, ki jih posnamemo kot XML in jih potem s *Pretvornikom za konkordance in kolokacije* preoblikujemo v format *.xlsx* za nadaljnjo obdelavo (gl. 3. poglavje kazalnika 5).

Random sample

Make a random sample from the concordance.
Number of lines in the sample:

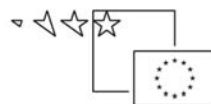
Slika 14: okno za izdelavo vzorca v SkE

2.2.1.3.3 Filter (Filter)

Funkcija omogoča oženje seznama konkordanc z dodatnimi omejitvami, ki so lahko **vključujoče** ali **izključujoče** (*positive* ali *negative*); pri vključujočem filtru lahko tudi nastavimo, kateri zadetki bodo v rezultatu izpostavljeni (*Selected token*); sistemska nastavitve je na prve. Nadalje lahko določimo še, koliko naj bodo filtrirani pogoji **oddaljeni od zadetka** (*Search span*).

Primer filtriranja: predhodno iskanje konkordanc po lemi *glodati* filtriramo z naslednjimi nastavitvami:

Filter:	<i>positive</i>
Razpon:	-5 do 5
Lema:	<i>črv</i>
Besedna vrsta:	<i>samostalnik</i>
Zveza:	<i>dvoma</i>



Concordance Filter

Filter: positive negative
 Selected token: first last
 Search Span: from to
Lemma: PoS:
Phrase:
Word Form: PoS: Match case:
CQL:
 Default attribute: [Tagset summary](#)

Slika 15: možnosti pri nastavitvi filtrov konkordanc v orodju SkE

Gornja nastavitvev prikaže samo konkordance, ki vsebujejo lemo *črv* v razdalji od pet besed levo do pet besed desno glede na predhodno iskano lemo *glodati*:

Dnevnik mestom v zasledovalnem teku potrdila svetovno kakovost , je vse **glodal** *črv dvoma* , kaj bi bilo , če bi ... Potem ko je povsem pogorela

Dnevnik rojstni dan . " Hvala bogu , da mi je steklo . Po treningu me je **glodal** *črv dvoma* in sem bil malce zaskrbljen . Rezerve so v sproščenosti

Dnevnik Breme je bilo zaradi velikih pričakovanj skoraj nevdržno . **Glodal** me je *črv dvoma* , ker v prosti tehnik nistem tako suverena bi človek mirno dejal - fantastično . *Črv dvoma* pa prične **glodati** , ko vzamemo malo bolj pod drobnogled naše odnose . Zakaj

NedDnev še naprej . Čeprav , kolikor se spoznam na to , *črv dvoma gloda* in nije globoke kanale še naprej . Včasih do popolne odtujitve

Dnevnik prizadeva verjeti v neodvisnost lastne umetnosti , ga neustavljivo **gloda** zob *dvoma* . Po uvodnem razmišljanju o naravi umetniškega

Slika 16: rezultat filtriranja v orodju SkE, vključujoči filter

Če filter nastavimo na *negative*, bodo iz rezultata izključeni zadetki, ki vsebujejo opredeljene pogoje:

NedDnev da ste iz kraja , kjer lahko čez ograjo gledate , kako " rja **gloda** arzenal " , kot ste dejali . Za začetek bi dejal , da preveč

kozi . Nekje globoko v njem se je zganila skrb in ga začela **glodati** . » Martinsson , « je rekel počasi , » se mogoče spominjate

Dnevnik informativnih in kuliniaričnih dogodkov . Glede na to , da festival ne **gloda** proračunske vreče in žepov davkoplačevalcev , bodo vstopnice

Delo vernikov spregovoril gospod nadškof . Vrgel je kost , ki jo bodo **glodali** časniki (in ne samo oni) v časopisih , na radiu , TV ...

svobodno , kajti prav taki ste tudi sami . Nekje globoko v vas pa **gloda** dvom , zato ste tudi poslali vprašanje . Hčerka je komunikativna

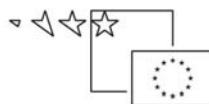
marsikaterega Janeza , imajo kost za glodanje . To je dobro , dokler **glodajo** , se ne posvečajo s tolikšnim žarom svojemu » poslanstvu

Slika 17: rezultat filtriranja v orodju SkE, izključujoči filter

2.2.1.3.4 Razvrščanje (Sort)

Funkcija omogoča različne načine abecednega razvrščanja konkordanc.

Osnovno razvrščanje je na voljo v spustnem meniju funkcije, omogoča pa standardno razvrstitev levo ali desno od konkordance ali po samem zadetku (*Left, Right, Node*).



Delo	Olimpiji , ki jo razjedajo problemi , enkrat samkrat pošteno brcnili na gol , pri čemer niso pokazali nikakršne ambicije , da	
Delo	se je po porazu z Olimpijo , ne da bi enkrat samkrat pošteno brcnila na gol , znašla prav na zadnjem mestu . Kaj se pravzaprav	
Delo	očitni premoči Prašnikarjevega novince so le dvakrat pošteno brcnili na gol - srečno izvlekli brez poraza . Ob popolni nemoči	
Dnevnik	skupaj sta Sanjsko moštvo . In Rouke je le žoga , ki jo bosta brcnila na Mars . Hongkonški režiserji so znali vedno ob steno vreči	
	pokalu tako daleč . Koliko balonov bo moral Alec Chamberlain brcniti na naslednji tekmi , nihče ne ve , verjetno pa jih bo še	
	mišljenje , takoj ko ste povohali oblast , pa čeprav so Vas brcnili na njen rob ? Kolikšna je bila Vaša cena , g. Dimitrij Rupel	
	mišljenje , takoj ko ste povohali oblast , pa čeprav so Vas brcnili na njen rob ? Kolikšna je bila Vaša cena , g. Dimitrij Rupel	
Joker	počistijo polje . Pasti pa kreature , ki prihajajo na obisk , brcenejo na pokopališče , s katerim je takisto moč ravnati ter služabnike	
Delo	sporočile , da je granato razneslo , ko jo je eden izmed dečkov brcnil na poti v šolo . Topovski izstrelek je verjetno ostal neaktiviran	
	popolnoma podzavestno . Sami naučeni refleksi . Enkrat samkrat ga brcnem na pravo mesto , je premišljevala . Toda to se je ta lup zdelo	
PrimNov	nacionalki se še nekaj trudijo , čeprav tudi Đuro velikokrat brcene na prvo žogo , nad Jonasom sem razočaran . Edino Majda Širca	
Delo	perspektivo , toda dejstvo , da Feri za Bežigradom niso pošteno brcnili na slovenski gol , je vendarle močnejše . Toda za nas je	
	Naš ata pa , namesto da bi svojo zgodbo o uspehu po kamelje brcnil na smetišče zgodovine , upogne svojo grbo in se ga v bližnji	

Slika 18: rezultat osnovnega razvrščanja po desni strani v orodju SkE

Dodatne možnosti pri Preprostem razvrščanju (*Simple Sort*) obsegajo določitev naslednjih pogojev:

- **atribut**, po katerem naj se konkordance razvrstijo (po besedi, lemi, itd.),
- **način** razvrščanja: po levi, desni ali po zadetku (*Left context, Node, Right context*),
- **razpon** razvrščanja (*Number of tokens to sort*),
- /ne/upoštevanje **velike/male začetnice** (*Ignore case*),
- **odzadnje** razvrščanje (*Backward*): razvrsti oblike, leme oziroma druge attribute, ki jih določimo, po abecedi vzvratno.

Primer nastavitve v Preprostem razvrščanju: atribut *word*, način razvrščanja: *Node*, kljukica pri *Ignore case* za neupoštevanje velike/male začetnice ter vključen *Backward*. Rezultat so razvrščene oblike iskane leme po seznamu konkordanc od zadaj naprej. Nastavitve in rezultat predstavljata naslednji dve sliki:

Simple Sort

Attribute:

Sort key: Left context Node Right context

Number of tokens to sort:

Ignore case Backward

Slika 19: primer nastavitve za Preprosto razvrščanje oblik iskane leme v orodju SkE

bomo , « mu je zagotovil major . **stisnila** sta si roki in major Liepa je
je zadržala , potem pa močno **stisnila** , še bolj je razširila nogi
tramvajaska kolesa so mu tako močno **stisnila** obe nogi , da so mu jih morali
in Adams ob začetku srečanja **stisnila** roki in sedela drug nasproti
pa je bilo nam , če smo lahko **stisnili** dlan tovarišu Titu . Sklonil
Ignjatoviča , nato pa je moral **stisniti** roke najboljšemu slovenskemu
na obrazu blažen izraz in je **stisnil** roko tudi sodnemu tolmaču za
izjave , vprašal : " Ali si morda **stisnil** roko zmagovalcu ? " Ker pa je
prevajalca , mu dal roko , nato **stisnil** roko tožilcu in še sodnici
pospešeno tiskal , Miha pa je **stisnil** roko odgovornemu uredniku Vasu
Lani me je tu na drugi tekmi ' **stisnilo** ' levi tetivi , letos pa v desni
vreči dol , pa me je pri srcu **stisnilo** , kaj bo potem z mojimi otroci

Slika 20: rezultat Preprostega razvrščanja oblik leme stisniti v orodju SkE



Pri nastavitvah in interpretiranju razvrščanj je treba upoštevati, da so ločila v korpusu FidaPLUS kodirana kot pojavnice (*token*), kar vpliva na razvrstitev: pri razvrščanju po desni tako dobimo najprej tiste konkordance, pri katerih je za zadetkom ločilo.

Dodatne možnosti pri Večnivojskem razvrščanju (*Multilevel Sort*) obsegajo določitev vrstnega reda, v katerem naj nastavljen razvrščanja potekajo.

Primer Večnivojskega razvrščanja: predhodno iskanje konkordanc po lemi *brcniti* razvrstimo z naslednjimi nastavitvami: prvi nivo po prvi besedi desno (1R), drugi nivo po drugi besedi desno (2R), tretji nivo po prvi besedi levo (1L):

Multilevel sort

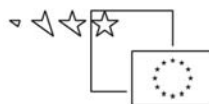
Slika 21: primer nastavitve za Večnivojsko razvrščanje v orodju SkE, predhodno iskanje po lemi *brcniti*

Rezultat Večnivojskega razvrščanja:

hišo , v kateri sama živi 75 - letna domačinka . Eden jo je **brcnil** v glavo , nato pa so od nje zahtevali denar . Ko jim je dala **brcnil** v glavo , tedaj pa sta jo podrla na tla . Eden od storilcev jo je **brcnil** v glavo , ko pa je poskušala vstati , jo je z roko potisnil naj nehajo . To je enega izmed njih tako razjezilo , da ga je **brcnil** v glavo . Ko je obrcani padel , so se pretepači razbežali , je Kukaja enkrat udaril z roko , enkrat ali dvakrat ga je **brcnil** v glavo . » Udaril sem ga , ker sta me ta dva ' balavea ' Zdravka B. za obleko , tako da je padel na tla , nato pa ga je **brcnil** v glavo . Zdravko je zaprosil za posredovanje policistov . je napadeni hotel zbežati skozi vrata , napadalec pa ga je **brcnil** v glavo in ga poškodoval . Sledi kazenska ovadba za ogrožanje je napadeni hotel zbežati skozi vrata , napadalec pa ga je **brcnil** v glavo in ga poškodoval . Sledi kazenska ovadba za ogrožanje najmanj petnajstkrat udaril v zgornji del telesa , Seferaj ga je **brcnil** v glavo in ga tepel , pa tudi Repanšek ga je brcal . Po opisu pa je potem šel nadenj z rokami in nogami . Večkrat ga je **brcnil** v glavo in udaril po obrazu , starejši Kamničan pa se je pa je potem šel nadenj z rokami in nogami . Večkrat ga je **brcnil** v glavo in udaril po obrazu , starejši Kamničan pa se je Ljubljancanka neznanecu ni hotela dati telefona , on pa jo je **brcnil** v glavo in ji vzel nahrbtnik . Prebrskal ga je in si postregel pajdaš , ki je prihital iz nasprotnne smeri . Oškodovanca je **brcnil** v glavo in ga zbil na tla . Z roke mu je snel uro , vredno poškodoval , da je med prevozom v bolnišnico umrl . J. S. Bik ga je **brcnil** v glavo Nova Gorica - V sredo zjutraj so na kmetiji v Ravnici zavrnjen . Zlatko Starčević Štirindvajset ur Konj ga je **brcnil** v glavo Šentjur - V bližini posestva Žonta v Šentjurju podtakajeni , nadaljujejo z zbiranjem obvestil . (ob) Konj ga je **brcnil** v glavo ŠENTJUR , 25 . V petek dopoldan so blizu posestva Rus LJUBLJANA Mamo sem enkrat boksnil v obraz , enkrat sem jo **brcnil** v glavo , nato pa pohodil . Na obraz sem ji zelo močno stopil daviti , jo prijel za lase in boksnil v obraz . Enkrat sem jo **brcnil** v glavo , nato pa pohodil . Na obraz sem ji zelo močno stopil nekajkrat pogledal v oči ; ko je štel tri leta , ga je kobila **brcnila** v glavo in je zdravnik rekel , da je z njim konec . Tudi drugič najboljših močeh , se je nihče ni bal . Nekoč jo je konj **brcnil** v glavo , zato je ni mogla več dvigniti zadosti visoko , da konju in je padel , ko je hotel razjahati . Tedaj ga je konj **brcnil** v glavo . Zelo sem se ustrašil zanj , Eethgow pa se je le

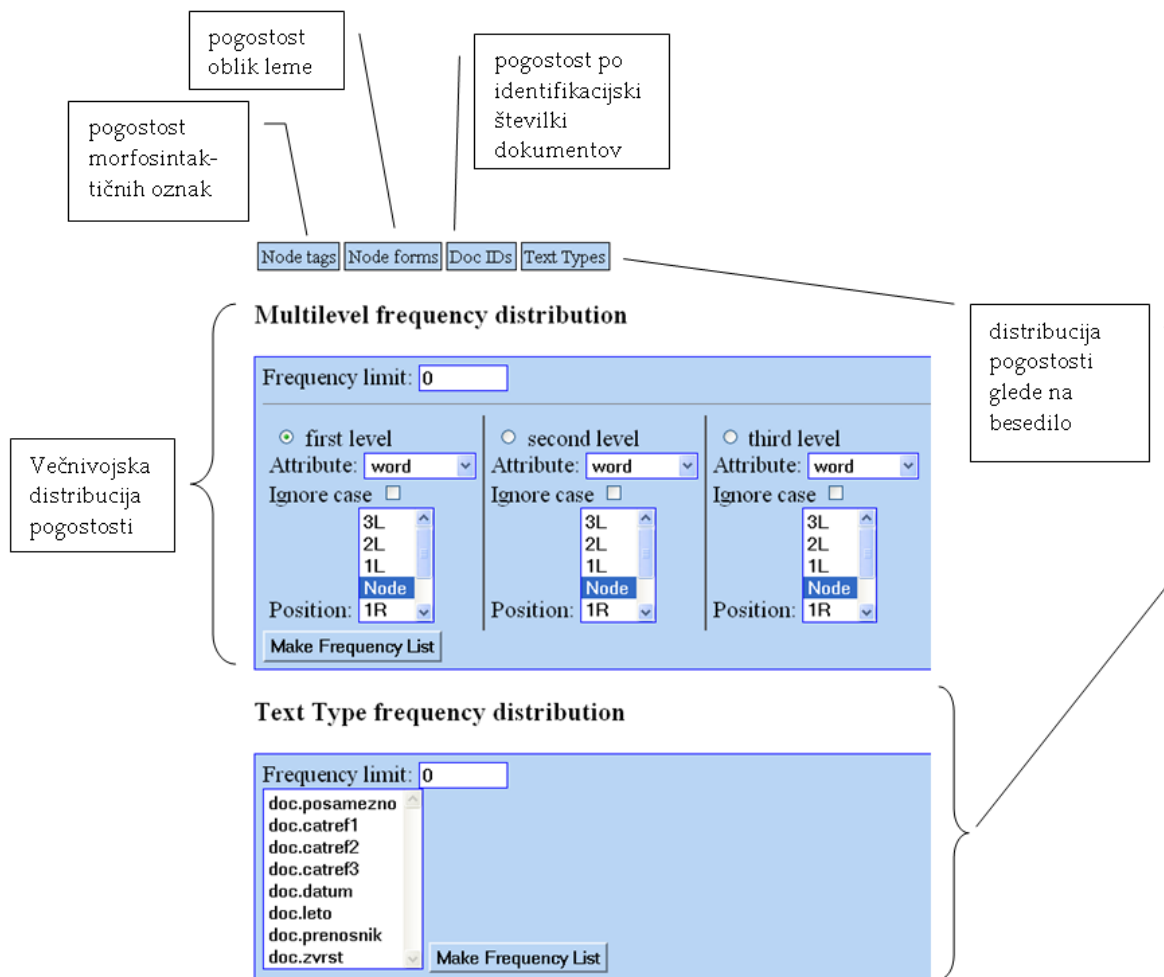
First | Previous Page 78 of 143 Go Next | Last

Slika 22: rezultat Večnivojskega razvrščanja v orodju SkE



2.2.1.3.5 Pogostost (Frequency)

Funkcija omogoča prikaz različnih podatkov o pogostosti pojavljanja iskane leme v korpusu. Osnovno okno za izdelavo seznamov pogostosti prikazuje naslednja slika:



Slika 23: okno za nastavitve v funkciji Pogostost (Frequency) v orodju SkE

Večnivojska distribucija pogostosti (*Multilevel frequency distribution*) je funkcija, najbolj vezana na same leksikalne podatke v korpusu, medtem ko so ostale vezane na podatke, ki o določeni lemi v korpusu obstajajo. Možno je izdelati naslednje distribucije pogostosti:

- po oblikoskladenjskih oznakah (*Node tags*),
- po oblikah (*Node forms*),
- po identifikacijskih številkah dokumentov (*Doc IDs*),
- po podatkih o besedilu (*Text types*).

Pri Večnivojski distribuciji pogostosti in Distribuciji pogostosti glede na besedilo je možno nastaviti več parametrov, kar je predstavljeno v nadaljevanju, dodani pa so tudi primeri ostalih distribucij pogostosti.



Večnivojska distribucija pogostosti (*Multilevel frequency distribution*)

Pri osnovni nastavitvi (vključen prvi nivo, atribut *word*, položaj *Node*) ta funkcija prikaže pogostost posameznih oblik pri lemi, ki smo jo predhodno poiskali v Konkordancah, npr. *glodam*, *glodaš*, *gloda* itd. pri lemi *glodati* (rezultat je enak, kot če izberemo gumb *Node forms*). Drugi in tretji nivo omogočata kompleksnejše kombinacije, npr.: če vključimo drugi nivo in izberemo položaj ena desno (1R), bomo dobili besede, ki se pojavljajo na tem mestu glede na iskano lemo ter kako pogoste so (gl. primer št. 1).

Primer 1: nastavev enega nivoja pri Večnivojski distribuciji pogostosti:

Izhodišče je predhodno iskanje leme *glodati*, osnovne nastavitve, kot je opisano zgoraj, pa so:

Vključen: *prvi nivo* (drugi in tretji nista aktivna)
Atribut: *word*
Prvi nivo: *Node*

Multilevel frequency distribution

Frequency limit: 0

first level second level third level

Attribute: word Attribute: word Attribute: word

Ignore case Ignore case Ignore case

3L 3L 3L
2L 2L 2L
1L 1L 1L
Node Node Node

Position: 1R Position: 1R Position: 1R

Make Frequency List

Slika 24: osnovna nastavev pri Večnivojski distribuciji pogostosti v orodju SkE

Rezultat Večnivojske distribucije pogostosti z vključenim prvim nivojem, atributom *word* ter položajem *Node* je:



Frequency list

Frequency limit:

	<u>word</u>	<u>Freq</u>	
p/n	gloda	246	
p/n	glodal	106	
p/n	glodajo	104	
p/n	glodati	84	
p/n	glodala	56	
p/n	glodali	44	
p/n	glodalo	31	
p/n	glodale	16	
p/n	glodamo	14	
p/n	glodam	9	
p/n	glodaš	8	
p/n	glodata	7	
p/n	Glodal	7	
p/n	glodate	5	

Slika 25: rezultat osnovne nastavitve pri Večnivojski distribuciji pogostosti – enako kot funkcija Node forms, v orodju SkE (primer 1)

Povezavi pred stolpcem (**p/n**) nas peljeta neposredno do relevantnih konkordanc, in sicer:

- znak "**p**" do vseh konkordanc, ki konkretni rezultat v seznamu **vsebujejo**,
- znak "**n**" pa do vseh konkordanc, ki konkretnega rezultata v seznamu **ne vsebujejo**.

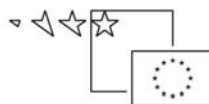
Možne so različne razvrstitve, in sicer če pritisnemo na naslove stolpcev. Osnovna nastavitev je padajoče po pogostosti.

Rezultate distribucij pogostosti lahko tudi shranimo v tekstovni ali XML obliki.

Primer 2: nastavitev dveh nivojev pri Večnivojski distribuciji pogostosti:

Izhodišče je predhodno iskanje po lemi *vrzel*, nastavitve pa so naslednje:

Vključen: *drugi nivo* (tretji nivo ni aktiven)
Atribut: na obeh mestih *lemma*
Prvi nivo: *Node*
Drugi nivo: *1L*



Multilevel frequency distribution

Frequency limit:

first level
 Attribute: Ignore case
 second level
 Attribute: Ignore case
 third level
 Attribute: Ignore case

Position: Node
 Position: Node
 Position: Node

Slika 26: nastavitve za prikaz dveh nivojev pri Večnivojski distribuciji pogostosti

Rezultat Večnivojske distribucije pogostosti z dvema vključenima nivojema – opazujemo povezavo leme *vrzel* z realizacijami na mestu ena levo od leme s skupno frekvenco posamezne kombinacije:

Frequency list

Frequency limit:

	<u>lemma</u>	<u>lemma</u>	<u>Freq</u>
p/n vrzel	privatizacijski		866
p/n vrzel	zapolniti		718
p/n vrzel			373
p/n vrzel	ta		302
p/n vrzel	velik		217
p/n vrzel	biti		181
p/n vrzel	zapolnjevat		140
p/n vrzel	zapolnitev		96
p/n vrzel	nastal		87
p/n vrzel	nastati		76
p/n vrzel	in		64
p/n vrzel	generacijski		55
p/n vrzel	veliko		53
p/n vrzel	proračunski		53

Slika 27: rezultat iskanja po dveh nivojih pri Večnivojski distribuciji pogostosti v orodju SkE (primer 2)

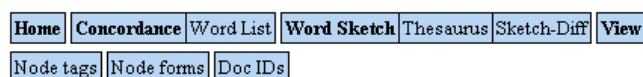
Primer 3a: nastavitve treh nivojev pri Večnivojski distribuciji pogostosti:

Izhodišče je predhodno iskanje po lemi *ocenjevanje* s kolokatorjem *vino* kot sobesedilom na položaju ena desno od iskane leme (funkcija Konkordance v povezavi z iskanjem po podatkih v sobesedilu). V Večnivojski distribuciji pogostosti nastavimo naslednje parametre:



Vključen: *tretji nivo*
Atribut: *povsod lemma*
Prvi nivo: *Node*
Drugi nivo: *1R*
Tretji nivo: *1L*

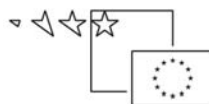
Prvi nivo predstavlja iskano lemo (*ocenjevanje*), drugi nivo kolokator na mestu ena desno glede na iskano lemo (*vino*), tretji nivo pa leme, ki se pojavljajo na mestu ena levo od iskane leme. Nastavitve v orodju SkE za konkretni primer so:



Multilevel frequency distribution

Slika 28: nastavitve za prikaz treh nivojev pri Večnivojski distribuciji pogostosti za lemi ocenjevanje in vino v orodju SkE

Kot rezultat dobimo distribucijo pogostosti na mestu ena levo glede na osnovno lemo – opazujemo povezavo zveze *ocenjevanje vin* z realizacijami na mestu ena levo s skupno frekvenco posamezne povezave:



Home	Concordance	Word List	Word Sketch	Thesaurus	Sketch-Diff	View
Frequency	Node tags	Node forms	Doc IDs	Save		

Frequency list

Frequency limit:

	lemma	lemma	lemma	Freq	
p/n	ocenjevanje	vin	mednarodn	286	
p/n	ocenjevanje	vin		123	
p/n	ocenjevanje	vin	na	122	
p/n	ocenjevanje	vin	društven	109	
p/n	ocenjevanje	vin	letošnji	63	
p/n	ocenjevanje	vin	za	61	
p/n	ocenjevanje	vin	mednarodno	48	
p/n	ocenjevanje	vin	pripraviti	46	
p/n	ocenjevanje	vin	in	45	
p/n	ocenjevanje	vin	pri	33	
p/n	ocenjevanje	vin	tudi	29	
p/n	ocenjevanje	vin	z	28	
p/n	ocenjevanje	vin	tradicionalen	24	
p/n	ocenjevanje	vin	rezultat	24	
p/n	ocenjevanje	vin	potekati	22	
p/n	ocenjevanje	vin	o	21	
p/n	ocenjevanje	vin	biti	20	
p/n	ocenjevanje	vin	prvi	19	
p/n	ocenjevanje	vin	tradicionalno	16	
p/n	ocenjevanje	vin	strokovn	15	
p/n	ocenjevanje	vin	državn	15	
p/n	ocenjevanje	vin	organizirati	13	

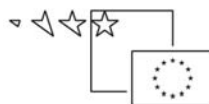
Slika 29: rezultat iskanja po treh nivojih pri Večnivojski distribuciji pogostosti v orodju SkE (primer 3a)

Primer 3b: nastavitev treh nivojev pri Večnivojski distribuciji pogostosti:

Izhodišče je predhodno iskanje konkordanc po lemi *glodati*. V Večnivojski distribuciji pogostosti nastavimo naslednje parametre:

Vključen: *tretji nivo*
Atribut: *povesod lemma*
Prvi nivo: *1L*
Drugi nivo: *1R*
Tretji nivo: *2R*

Prvi nivo predstavlja mesto ena levo (1L) od iskane leme (*glodati*), drugi nivo predstavlja kolokator na mestu ena desno (1R) glede na iskano lemo, tretji nivo pa leme, ki se pojavljajo na mestu dva desno (2R) od iskane leme. Nastavitve v orodju SkE za konkretni primer so:



Multilevel frequency distribution

Frequency limit:

first level
 Attribute: lemma
 Ignore case
 Position: 1R

second level
 Attribute: lemma
 Ignore case
 Position: 1R

third level
 Attribute: lemma
 Ignore case
 Position: 2R

Make Frequency List

Slika 30: nastavitve za prikaz treh nivojev pri Večnivojski distribuciji pogostosti za lemo glodati v orodju SkE

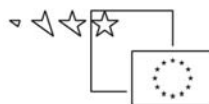
Kot rezultat dobimo distribucijo pogostosti okoli leme *glodati*, in sicer na mestu ena levo (1. stolpec), ena desno (2. stolpec) in dve desno (3. stolpec) s skupno frekvenco posamezne povezave:

Frequency list

Frequency limit:

<u>lemma</u>	<u>lemma</u>	<u>lemma</u>	<u>Freq</u>
p/n začeti	črv	dvom	8
p/n	pohištvo	in	5
p/n on	črv	dvom	4
p/n	ti	biti	4
p/n	jaz	biti	4
p/n ti	ljubosumje		3
p/n ti	kakršenkoli	težava	3
p/n skupaj	šolski	klop	3
p/n rezko	pirave	jasli	3
p/n pa	črv	dvom	3
p/n in	piroški		3
p/n in	mišičen	tkivo	3
p/n in	in	vrtati	3
p/n biti	skrb		3
p/n biti	in	glodati	3
p/n biti		dokler	3
p/n	on	biti	3
p/n			3
p/n še	črv	dvom	2
p/n čas	strah		2
p/n zob	in	drobiti	2
p/n zmeraj			2
p/n začeti		se	2
p/n začetek	vrv		2
p/n zavist	zavistnež	bolj	2
p/n vseeno	kanček	upanje	2
p/n večno	črv	dvom	2

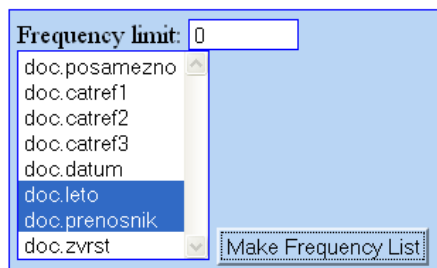
Slika 31: rezultat iskanja po treh nivojih pri Večnivojski distribuciji pogostosti v orodju SkE (primer 3b)



Distribucija pogostosti glede na besedilo (*Text type frequency distribution*)

Ta funkcija prikaže, kako je iskana lema razporejena v korpusnih besedilih glede na kriterije, kot so leto, zvrst, prenosnik, datum, oznaka tipologije ali identifikacijska številka dokumenta. Koristna je, kadar želimo raziskovati razmerja med različnimi vrstami besedil, različnimi leti in podobno. Izberemo enega ali več kriterijev, npr.:

Text Type frequency distribution



Slika 32: nastavitve za prikaz Distribucije pogostosti glede na besedilo v orodju SKI

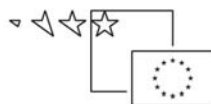
Rezultat Distribucije pogostosti glede na besedilo – predhodno iskana lema je *glodati*, izbrana kriterija sta *leto* in *prenosnik*:

Frequency list

Frequency limit: 0			
<u>doc.leto</u>	<u>Freq</u>	<u>Rel [%]</u>	
p/n 2003	122	344.5	█
p/n 2004	92	283.3	█
p/n 2005	80	405.1	█
p/n 2001	73	168.4	█
p/n 2002	72	222.9	█
p/n 1999	60	77.9	█
p/n 1998	60	33.7	█
p/n 2000	57	78.3	█
p/n 1997	42	38.9	█
p/n 2006	38	236.0	█
p/n 1996	26	56.1	█
p/n	18	89.8	█
p/n 1994	11	36.0	█
p/n 1993	8	78.1	█
p/n 1995	7	16.4	█
p/n 1992	3	99.6	█
p/n 0200	2	50.6	█
p/n 1990	1	1904.2	█

<u>doc.prenosnik</u>	<u>Freq</u>	<u>Rel [%]</u>	
p/n časopis	367	61.1	█
p/n revija	206	172.4	█
p/n knjiga	188	944.9	█
p/n neoznačeno	10	33.3	█
p/n govorni	1	60.2	█

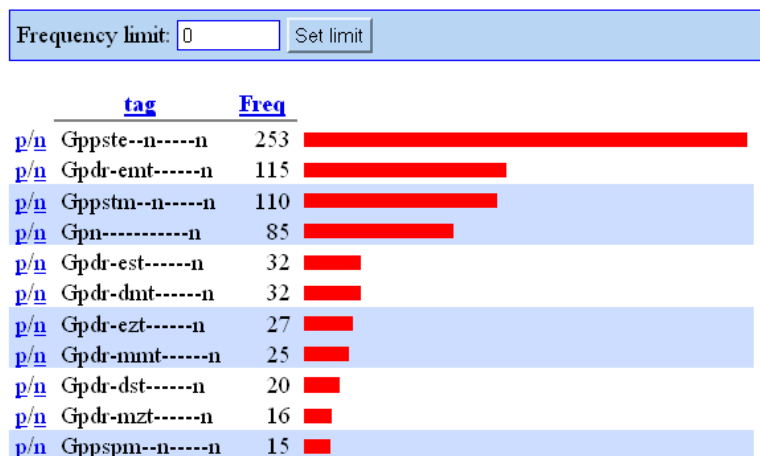
Slika 33: rezultat Distribucije pogostosti glede na besedilo: leto in prenosnik



Drugi primeri

Distribucija pogostosti s podatki o besedilu, in sicer po oblikoskladenjskih oznakah (*Node tags*):

Frequency list



Slika 34: distribucija pogostosti oblikoskladenjskih oznak pri lemi glodati (*Node tags*)

2.2.1.3.6 Kolokacija (*Collocation*)

Funkcija omogoča izdelavo seznama možnih kolokacij, torej kombinacij iskane leme ali besedne zveze z drugimi, ki se pogosto pojavljajo ob iskani lemi. Dobrodošla je, kadar ne želimo delati celotne Besedne skice za iskano lemo. V funkciji Kolokacije določimo atribut, razpon, najnižjo pogostost v korpusu, najnižjo pogostost v opredeljenem razponu, statistične izračune in vrsto razvrstitve:

Collocation candidates

Attribute: lemma In the range from: -3 to: 3
Minimum frequency in corpus: 5
Minimum frequency in given range: 3
Show functions: salience Sort by: salience
Buttons: Make Candidate List, Save Options

Slika 35: nastavitve v funkciji Kolokacije v orodju SKE (*Collocation candidates*)



Opis statističnih izračunov je dostopen v dokumentu *Statistics used in the Sketch Engine*, ki je v prilogi 1 kazalnika 5 – SSJ_K5_Priloga1_StatisticsUsedInTheSkE.pdf – in je dostopen tudi na povezavi <http://trac.sketchengine.co.uk/wiki/SkE/DocsIndex>.

Primer za izdelavo seznama možnih kolokacij

Izhodišče je predhodno iskanje po lemi *ocenjevanje* in kolokatorju *vino* do tri mesta desno od iskane leme.

Collocation candidates

Slika 36: primer nastavitve za izdelavo seznama možnih kolokacij v orodju SkE (funkcija Collocation)

Rezultat izdelave seznama možnih kolokacij:

Collocation candidates

	<u>Freq</u>	<u>T-score</u>	<u>MI3</u>	<u>log likelihood</u>	<u>salience</u>
<u>p/n</u> mednaroden	299	17.259	25.488	3202.202	0.000
<u>p/n</u> društven	112	10.581	26.260	1748.271	0.000
<u>p/n</u> na	374	18.104	21.062	1430.194	0.000
<u>p/n</u> letošnji	112	10.535	21.403	993.279	0.000
<u>p/n</u> pripraviti	122	10.969	21.043	980.603	0.000
<u>p/n</u> mednarodno	51	7.135	21.398	610.401	0.000
<u>p/n</u> vino	59	7.646	19.538	520.206	0.000
<u>p/n</u> potekati	49	6.949	18.323	385.875	0.000
<u>p/n</u> rezultat	45	6.642	17.649	327.768	0.000
<u>p/n</u> pravilnik	29	5.370	18.222	284.592	0.000
<u>p/n</u> tudi	95	8.678	16.328	254.770	0.000
<u>p/n</u> tradicionalen	29	5.360	17.444	253.420	0.000
<u>p/n</u> tradicionalno	23	4.789	18.414	253.070	0.000
<u>p/n</u> organizirati	29	5.344	16.758	226.015	0.000
<u>p/n</u> medalja	20	4.462	17.450	204.498	0.000
<u>p/n</u> stolfa	8	2.828	23.926	189.284	0.000
<u>p/n</u> vsakoleten	16	3.995	17.643	182.099	0.000

Slika 37: seznam možnih kolokatorjev zveze ocenjevanje vin v razponu do tri mesta levo od leme ocenjevanje, v orodju SkE



2.2.1.3.7 Shrani (*Save*)

Funkcija omogoča shranjevanje konkordanc v tekstovni ali XML obliki. Način shranjevanja konkordanc za nadaljnjo uporabo pri izdelavi leksikalne baze za slovenščino je predstavljen v poglavju 3.3.1, kjer opisujemo prvega od postopkov, ki jih omogoča *Pretvornik za konkordance in kolokacije*.

2.2.2 Besedna skica (*Word Sketch*)

2.2.2.1 Opis

Besedne skice (*Word sketches*) so avtomatski na korpusu temelječi sežetki slovničnega in kolokacijskega vedenja neke besede (Krek in Kilgarriff 2006: 62). Besedna skica prikazuje t. i. leksikalni profil iskane leme s podatki o dvočlenskih in tročlenskih enotah, ki izkazujejo njeno tipično sobesedilno okolje.

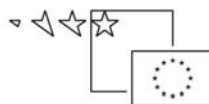
Temelj za Besedno skico so **slovnične relacije** (*grammatical relations*), ki določujejo, katere podatke, ki so v korpusu, bo mehanizem upošteval pri pripravi Besedne skice. Priprava slovničnih relacij je jezikovnotehnoški postopek, vezan na:

- poznavanje tipičnih sobesedilnih pojavov v konkretnem jeziku,
- iskanje drugih leksikalno zanimivih sobesedilnih pojavov,
- način kodiranja, ki je uporabljen v posameznem korpusu.

V poglavju 2.5 **Slovnične relacije v slovnici Besednih skic (SBS) v.15** so prikazane osnove slovnice Besednih skic in podrobno opisane same slovnične relacije (v. 15), revidirane v raziskovalni fazi projekta SSJ za referenčni korpus slovenskega jezika FidaPLUS. Ko bo na voljo novi referenčni korpus, ki nastaja pri projektu SSJ, bodo slovnične relacije prilagojene kodiranju v njem.

2.2.2.2 Izdelava Besedne skice

Okno za izdelavo Besedne skice:



Word Sketch Entry Form

Corpus:	Fida PLUS 620m
Lemma:	<input type="text" value="rana"/>
Part of speech:	<input type="text" value="samostalnik"/>
Advanced options <input type="checkbox"/>	
Subcorpus:	create new
Sort grammatical relations:	<input checked="" type="checkbox"/>
Minimum frequency:	<input type="text" value="auto"/>
Minimum salience:	<input type="text" value="0.0"/>
Maximum number of items in a grammatical relation:	<input type="text" value="10"/>
Sort collocations according to:	<input checked="" type="radio"/> Salience <input type="radio"/> Raw frequency
Textbox Lexicography template:	<input type="text" value="fidaplus_slovene"/> Examples per collocate: <input type="text" value="8"/>
Cluster collocations	<input type="checkbox"/>
Minimum similarity between cluster items:	<input type="text" value="0.15"/>
<input type="button" value="Show Word Sketch"/> <input type="button" value="Save Options"/>	

Slika 38: okno z nastavitvami za izdelavo Besedne skice v orodju SkE beta

Iskano lemo vpišemo v polje *lemma* in ji določimo besedno vrsto (*Part of speech*).

Možnosti v naprednih nastavitvah različice SkE beta so:

- določitev podkorporusa (*Subcorpus*),
- razvrščanje realiziranih slovničnih relacij (*Sort grammatical relations*),
- najnižja pogostost (zadetka v slovnični relaciji) (*Minimum frequency*),
- najnižja statistična izpostavljenost (*Minimum salience*),
- najvišje število navedenih vrednosti v slovnični relaciji (*Maximum number of items in a grammatical relation*),
- želeni način razvrščanja, tj. po izpostavljenosti (*salience*) ali po absolutni frekvenci (*Sort collocations according to*),
- šablona za Kliksikografijo (*Textbox Lexicography template*),
- število zgljedov kot rezultat Kliksikografije (*Examples per collocate*),
- gručenje kolokacij (*Cluster collocations*),
- najnižja podobnost med zadetki pri gručenju (*Minimum similarity between cluster items*).

Izdelano Besedno skico prikazuje naslednja slika:



rana Fida PLUS 20m (v.13) freq = 638

zaradi-d X 34 45.7 <input type="checkbox"/> umreti 12 21.36 <input type="checkbox"/> biti 11 12.04 biti kaksen? 13 14.9 <input type="checkbox"/> hud 5 20.11	koga-ali-cesa 49 11.8 <input type="checkbox"/> lizati 5 22.45 <input type="checkbox"/> celiti 4 20.14 <input type="checkbox"/> zaceliti 4 19.0 <input type="checkbox"/> oskrbeti 3 13.56 <input type="checkbox"/> čuhtiti 3 9.55	na-d 60 5.0 <input type="checkbox"/> želodec 4 19.81 <input type="checkbox"/> glava 7 19.74 <input type="checkbox"/> noga 6 19.09 <input type="checkbox"/> telo 5 16.04 <input type="checkbox"/> prsi 3 15.62 <input type="checkbox"/> uho 3 14.58 <input type="checkbox"/> levi 3 13.08 <input type="checkbox"/> koža 3 12.86	komu-ali-cemu 10 3.9 <input type="checkbox"/> podleči 5 23.72	koga-ali-kaj 70 3.8 <input type="checkbox"/> celiti 9 30.33 <input type="checkbox"/> lizati 5 21.53 <input type="checkbox"/> zaceliti 3 20.8 <input type="checkbox"/> zadajati 3 18.66 <input type="checkbox"/> obvezati 4 16.82 <input type="checkbox"/> povzročiti 3 7.71 <input type="checkbox"/> dobiti 3 4.27 <input type="checkbox"/> imeti 5 3.32
--	--	---	---	---

številu pojavitev kolokatorja

ime slovnične relacije in dostop do nje

iz-d 11 3.5 <input type="checkbox"/> vojna 3 13.31	kaksen? 236 2.8 <input type="checkbox"/> strelen 18 41.48 <input type="checkbox"/> vboden 14 40.3 <input type="checkbox"/> zaceljen 8 34.11 <input type="checkbox"/> hud 26 31.05 <input type="checkbox"/> gnojen 7 29.36 <input type="checkbox"/> nezaceljen 4 24.78 <input type="checkbox"/> povečičan 4 23.37 <input type="checkbox"/> rakav 5 22.39 <input type="checkbox"/> Kristusov 6 22.05 <input type="checkbox"/> boleč 6 20.36	na-d X 25 2.4 <input type="checkbox"/> obliž 8 37.34 <input type="checkbox"/> položiti 3 12.1	s-koga-cesa 57 2.4 <input type="checkbox"/> celjenje 9 38.55 <input type="checkbox"/> obvezovanje 3 25.36 <input type="checkbox"/> liv 3 19.89 <input type="checkbox"/> zdravljenje 5 18.41 <input type="checkbox"/> film 6 16.14 <input type="checkbox"/> čiščenje 3 15.09	osebek od 64 2.1 <input type="checkbox"/> krvaveti 5 20.85 <input type="checkbox"/> zaceliti 3 15.25 <input type="checkbox"/> odpirati 3 10.16 <input type="checkbox"/> skrivati 3 9.1 <input type="checkbox"/> povzročiti 3 7.89	oba v rod 76 2.0 <input type="checkbox"/> hud 20 32.51 <input type="checkbox"/> strelen 7 29.86 <input type="checkbox"/> nezaceljen 4 27.41 <input type="checkbox"/> gnojen 4 24.05 <input type="checkbox"/> boleč 5 21.21 <input type="checkbox"/> vboden 3 19.82 <input type="checkbox"/> majhen 3 9.32
--	--	--	--	---	---

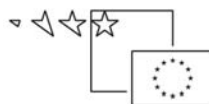
številu pojavitev in dostop do vseh konkordanc v slovnični relaciji

Slika 39: predstavitev Besedne skice (Word Sketch) v orodju SkE

Če vključimo možnost gručenja kolokacij (*Cluster collocations*), so skupaj navedeni statistično sorodni kolokatorji v posamezni slovnični relaciji; pri lemi *rana* tako pridejo skupaj *zadati rano* in *prizadejati rano*, *oskrbeti rano* in *oskrbovati rano* (slovnična relacija *koga-kaj/v_tožil* z lemo v tožilniku) in podobno:

celiti	328	599	77.36
zaceliti	239	ozdraviti	32
<input type="checkbox"/> lizati	104		56.48
zadati	53	75	36.61
prizadejati	22		
oskrbeti	50	59	35.9
oskrbovati	9		
obvezati	34	45	34.53
obvezovati	11		
zdraviti	107	198	34.27
povzročiti	64	povzročati	27
<input type="checkbox"/> zadajati	14		31.2
zašiti	21	29	30.87
šivati	8		

Slika 40: gručenje kolokacij leme *rana* v orodju SkE



2.2.3 Primerjalna skica (*Sketch-Diff*)

2.2.3.1 Opis

Primerjalna skica je statistično urejena tabela s primerjavo dveh lem, ki sta si na nek način podobni, npr. po rabi, pomenu ali obliki. Primerjalna skica za vsako slovnično relacijo prikaže, kateri kolokatorji se večkrat vežejo s prvo oziroma drugo primerjano lemo, potem pa še, katere kolokatorje privlači samo prva ali samo druga lema.

Do okna za Primerjalno skico dostopamo iz različnih delov orodja SkE:

- iz Tezavra s klikom na zadetke v seznamu (to je že neposredno povezano z izdelano Primerjalno skico iskane leme in zadetka v seznamu),
- iz drugih delov orodja s klikom na gumb *Sketch-Diff*.

Primerjalne skice so namenjene raziskovanju semantičnih in sintaktičnih podobnosti in razlik pri različnih razmerjih med skupinami besed, npr.:

- tujka proti netujki, npr. *problem* – *težava*, *cona* – *območje*,
- nedovršnost proti dovršnosti, npr. *povzročati* – *povzročiti*,
- samostalnik proti glagolniku, npr. *spodbuda* – *spodbujanje*,
- glagolnik proti glagolniku, npr. *soočanje* – *soočenje*.

2.2.3.2 Izdelava Primerjalne skice

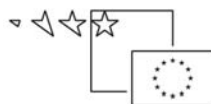
Iskalno okno pri funkciji Primerjalna skica:

Word Sketch Differences Entry Form

Corpus:	Fida PLUS 620m
First lemma:	<input type="text" value="položaj"/>
Second lemma:	<input type="text" value="situacija"/>
Part of speech:	<input type="text" value="samostalnik"/>
Sort grammatical relations:	<input checked="" type="checkbox"/>
Separate blocks:	<input type="radio"/> all in one block <input checked="" type="radio"/> common/exclusive blocks
Minimum frequency:	<input type="text" value="auto"/>
Maximum number of items in a grammatical relation of the common block:	<input type="text" value="25"/>
Maximum number of items in a grammatical relation of the exclusive block:	<input type="text" value="12"/>

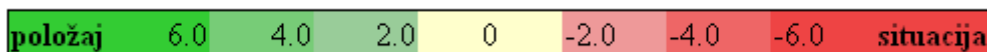
Slika 41: iskalno okno za izdelavo Primerjalne skice v orodju SkE

Rezultat Primerjalnih skic so tabele, ki skupaj s statističnim izračuni za posamezno lemo prikazujejo, kateri kolokator je pri določeni lemi prevladujoč. Poleg Tabele skupnih vzorcev (*Common patterns*) so na voljo tudi tabele, ki kažejo vezavo kolokatorjev samo s prvo oziroma drugo lemo (npr. "*položaj*" *only patterns*, "*situacija*" *only patterns*). Prevladujoča distribucija kolokatorjev po primerjanih lemah je prikazana tako s statističnimi podatki kot tudi vizualno



na podlagi barvne lestvice, definirane na vrhu Primerjalne skice. Na klik so neposredno dosegljive tudi konkordance posameznih kolokatorjev pri primerjanih lemah.

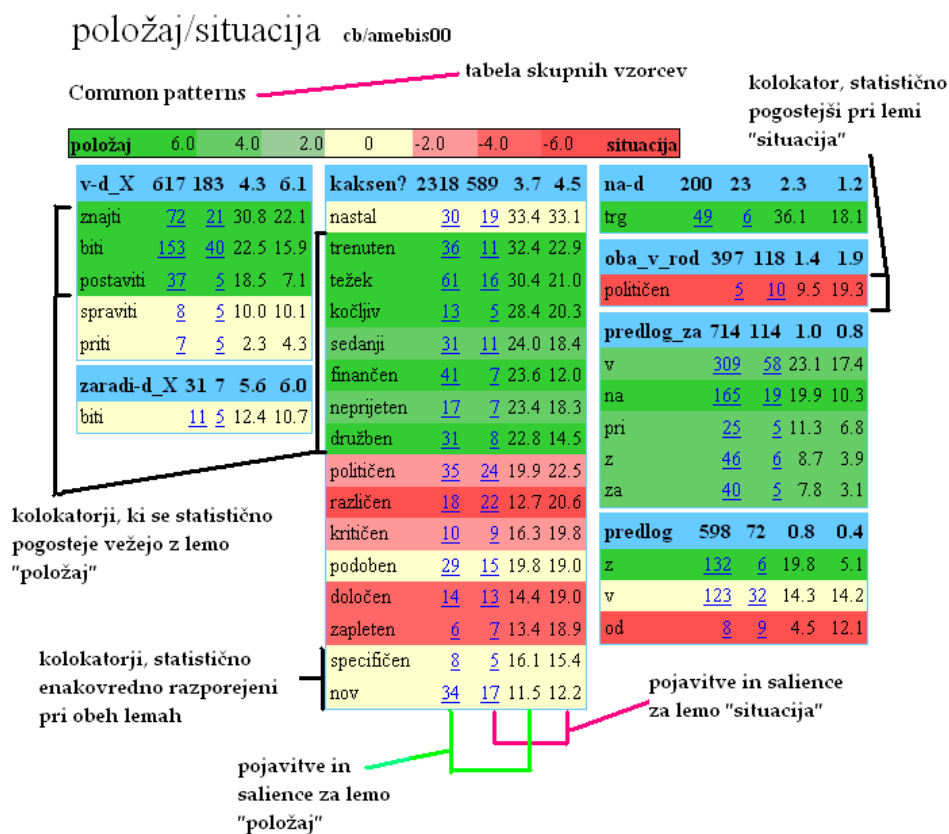
Lestvica, ki prikazuje podobnost oziroma različnost primerjanih lem, je prikazana na naslednji sliki:



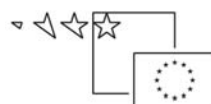
Slika 42: lestvica izračunov pri lemah, udeleženi v Primerjalni skici v orodju SkE

Kolokatorji, ki se večinsko vežejo z lemo *položaj*, so obarvani v odtenkih zelene barve, tisti, ki se večinsko vežejo z lemo *situacija*, pa v odtenkih rdeče barve. Manj ko je primesi druge leme pri določenem kolokatorji, intenzivnejši je odtenek; rezultati, ki so svetlo rumeni ali v razponu od 2.0 do -2.0, torej izkazujejo statistično največjo podobnost pri obeh lemah.

Opisano prikazujeta naslednji sliki:



Slika 43: Primerjalna skica lem situacija in položaj; nekatere od primerjanih slovničnih relacij



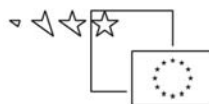
V Tabeli skupnih vzorcev so ob vsakem kolokatorju štiri številke. Prvi dve predstavljata frekvenco pojavljanja s prvo in drugo lemo, drugi dve pa izračunani izpostavljenosti (*salience*) med kolokatorjem in vsako lemo posebej. Kolokatorji so razvrščeni padajoče po seštevku obeh izračunov izpostavljenosti.

"položaj" only patterns

prec_glede-d 113 4.4		prec_o-d 3068 4.4		prec_zaradi-d 658 3.7		prec_na-d 9385 3.4	
biti	9 7.4	resolucija	57 37.4	zaskrbljenost	15 29.4	naslednik	84 35.5
		spregovoriti	74 30.7	razveljaviti	23 28.7	imenovanje	107 35.4
		posvet	41 30.2	zaskrbeti	16 24.8	zavihteti	54 32.7
		podatek	144 30.0	gol	10 18.2	uradnik	69 31.1
		govor	44 28.4	protestirati	9 17.4	izvolitev	55 31.0
		sporazum	71 28.4	sodnik	10 14.7	napad	164 30.1
		informacija	82 27.6			predhodnik	39 29.7
		pogovarjati	73 25.4			človek	374 29.6
		miza	50 24.3			okrepitev	46 29.0
		razmislek	16 23.2			rotacija	20 28.7
		raziskava	43 23.0			ženska	176 27.8
		študija	26 21.9			ostati	261 25.6

is_obj4 18002 3.2		prec_v-d 14523 3.1		a_modifier 99954 2.8		post_kot-d 609 2.6	
zlorabiti	497 60.5	vrniti	457 30.9	neenakopraven	925 79.6	potrošnik	12 21.8
utrditi	630 56.2	dvigniti	171 25.6	podrejen	971 78.3	moški	13 16.4
zasedati	365 47.9	spustiti	126 24.7	privilegirani	639 66.3	Slovenija	25 15.9
obstreljevati	141 46.9	ročka	18 24.3	sedeč	524 63.2	članica	10 15.0
zasesti	556 43.9	namestiti	84 21.5	gmoten	685 58.4	oseba	10 13.2
zavzeti	346 41.2	potiskati	30 20.6	premierski	301 57.7	večina	9 11.0
utrjevati	153 40.0	ročica	25 20.5	vodilen	2116 57.2	leto	13 7.5
okrepiti	347 39.5	trup	27 20.4	enakopraven	659 57.1	človek	8 7.5
zlorabljeni	124 39.0	vztrajanje	17 19.3	dohodkoven	344 56.3		
zadržati	265 35.5	medenica	11 19.0	pat	172 55.1		
obdržati	233 33.2	položiti	61 18.0	štarten	410 53.2		
izboriti	83 32.5	vračanje	24 17.6	neenak	231 53.1		

Slika 44: Primerjalna skica lem situacija in položaj: nekatere od slovničnih relacij, v katerih nastopa lema položaj, lema situacija pa ne



2.3 Dodatne funkcije v SkE beta: Kliksikografija (TBL), Izbor dobrih zgledov (GDEX)

2.3.1 Posebne programske nastavitve

Kot smo povedali v uvodu h kazalniku 5, različica beta orodja SkE, do katere dostopajo tudi sodelavci pri projektu SSJ, vsebuje več možnosti pri nastavitvah analiz, zlasti pa dodatne funkcije, ki omogočajo lažji prenos podatkov v programe za sestavljanje slovarjev – te so opisane v nadaljevanju poglavja, specifične nastavitve zanje po prijavi v sistem in izbiri besedilnega korpusa FidaPLUS pa so naslednje:

1) vzpostavitev funkcije Kliksikografija oziroma TBL (*Tick-box lexicography*):

- izberemo funkcijo **Besedna skica** (*Word Sketch*),
- v nastavitvah iskalnega okna (*Word Sketch Entry Form*) pri *Tickbox Lexicography template* izberemo šablono **fidaplus_slovene**,
- v istem oknu določimo tudi **število zgledov** (*Examples per collocate*), ki naj jih sistem prikaže po tem, ko v Besedni skici označimo kolokatorje,
- **shranimo** nastavitvev (*Save options*).

Tickbox Lexicography template: fidaplus_slovene Examples per collocate: 6

Cluster collocations

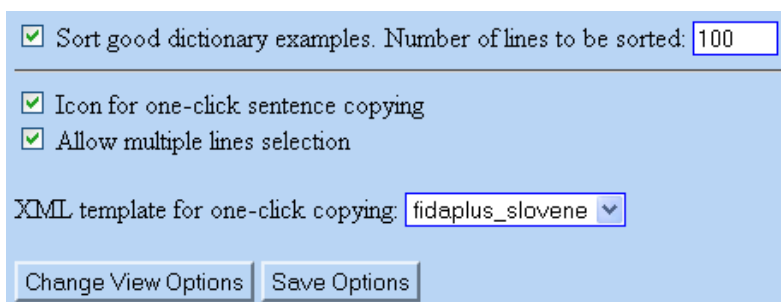
Minimum similarity between cluster items: 0.15

Show Word Sketch Save Options

Slika 45: posebne programske nastavitve v SkE beta: Besedna skica (*Word Sketch*)

2) vzpostavitev funkcije Izbor dobrih zgledov oziroma GDEX (*Good examples*) in prenosa zgledov v odložišče v formatu leksikalne baze za slovenščino:

- izvedemo katerokoli **iskanje v funkciji Konkordance** (*Concordance*),
- kliknemo na **Možnosti prikaza** (*View options*) in:
 - dodamo kljukico v potrditveno polje za razvrščanje slovarskih zgledov (*Sort (out) dictionary examples*); število lahko ostane sistemsko nastavljeno na 100,
 - dodamo kljukico pri *Icon for one-click sentence copying*,
 - dodamo kljukico pri *Allow multiple lines selection* ter
 - pri *XML template for one-click copying* izberemo **fidaplus_slovene**,
- **shranimo** nastavitvev (*Save options*).



Slika 46: posebne programske nastavitve v SkE beta: Možnosti prikaza (View options)

Opomba: če nam odložišče pri označevanju konkordanc v programu Mozilla Firefox ne deluje, težavo odpravimo z navodilom na <http://trac.sketchengine.co.uk/wiki/ProblemSolving/OneClick>.

2.3.2 Vsebinske nastavitve pri izdelavi leksikalne baze za slovenščino

Vsebinsko delo pri leksikalni bazi pogosto narekuje nastavitve parametrov, ki omogočijo natančnejše raziskave, kot izhodišče pa lahko določimo naslednje:

Nastavitve za izdelavo vzorca konkordanc (*Sample*)

V raziskovalni fazi projekta smo za leksikalno obravnavo konkordanc tipično analizirali med 250 in 300 konkordanc za vsako geslo. Izkazuje se, da se pri večini gesel, ki jih uvrščamo v leksikalno bazo za slovenščino, pomenske smernice pri približno 100-200 konkordancah izrazijo dovolj očitno. Na njihovi podlagi lahko postavimo semantično strukturo gesla, zato pri sestavljanju novih gesel izhodiščno analiziramo **150 konkordanc**, pri semantično kompleksnih geslih pa **do 350 konkordanc**.

Nastavitve v Besedni skici	Privzete vrednosti	Vrednosti pri leksikalni bazi
<i>Minimum frequency</i>	auto (=5)	10
<i>Minimum salience</i>	0.0	0.0
<i>Maximum number of items in a grammatical relation</i>	25	25

Slika 47: vsebinske nastavitve v Besedni skici kot izhodišče pri izdelavi leksikalne baze za slovenščino

2.3.3 Kliksikografija (TBL, Tick-box lexicography)

Kliksikografija omogoča, da v Besedni skici izberemo zelene kolokatorje in nato s skupnim klikom na ikono na dnu seznama pri posamezni slovnični relaciji dobimo Izbor dobrih zglede. Te potem lahko z nadaljnjo označitvijo prenesemo v odložišče in dalje v podatkovno bazo.

Orodje SkE z vzpostavljeno funkcijo Kliksikografija in označenimi kolokatorji je videti takole:



Gamma Fida PLUS 620m freq = 20319

a modifier	8068	2.1	prec zaradi-d	589	30.3
<input checked="" type="checkbox"/> vboden	353	86.89	<input type="checkbox"/> biti	320	34.42
<input checked="" type="checkbox"/> streljen	583	83.91	<input type="checkbox"/> umreti	78	34.12
<input checked="" type="checkbox"/> zaceljen	162	77.37	<input type="checkbox"/> morati	17	21.54
<input type="checkbox"/> nezaceljen	147	77.33	<input type="checkbox"/> bolnišnica	9	14.16
<input checked="" type="checkbox"/> hud	963	62.12	<input type="checkbox"/> zdraviti	9	13.21
<input type="checkbox"/> krvaveč	99	60.53	<input type="checkbox"/> ura	7	8.88
<input type="checkbox"/> rakav	168	57.02	<input type="checkbox"/> pomoč	6	7.53
<input type="checkbox"/> gnojen	94	53.28	>>		
<input checked="" type="checkbox"/> globok	315	52.55			
<input type="checkbox"/> razpocen	34	51.95	prec kljub-d	36	16.2
>>			<input type="checkbox"/> biti	25	20.59
			>>		

Slika 48: označitev kolokatorjev (TBL) v Besedni skici za uporabo funkcije Izbor dobrih zgledov (GDEX) v orodju SkE beta

2.3.4 Izbor dobrih zgledov (GDEX, Good examples)

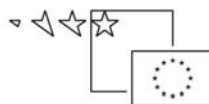
Ko v Besedni skici označimo kolokatorje in sprožimo iskanje na dnu stolpca s posamično slovnično relacijo, dobimo Izbor dobrih zgledov. Namen funkcije je skrajšati časovno obremenilno redaktorsko delo na področju izbora korpusnih zgledov, ki bi čimbolj ustrezali ilustraciji pomenov v slovarju ali leksikalni bazi. Te je možno zajeti z določljivimi mehanizmi in tako najprej prikazati tiste zglede, ki dobijo najboljšo oceno ustreznosti.

Dobri zgledi v slovarju naj bi:

- kazali tipično rabo in pogoste vzorce,
- bili informativni in naj bi osvetlili razlago,
- bili uporabnikom razumljivi, torej naj bi se izogibali težjemu besedišču ali strukturam, motečim ali nerazumljivim imenom, anaforičnemu navezovanju in podobnim besedilnim elementom, ki niso razumljivi brez več sobesedila.

Več o vsebinskem izboru ustreznih zgledov najdemo v Atkins in Rundell 2008, funkcija Izbor dobrih zgledov v SkE in kriteriji zanje pa so predstavljeni v Kilgarriff et al. 2008. Povzeto gre za naslednje kriterije izbora (Kilgarriff et al. 2008, 426-427):

- dolžina stavka: ustrenejši so stavki, ki vsebujejo od 10 do 25 besed;
- pogostost besed: ustrenejši so stavki, v katerih so uporabljene najbolj pogoste besede v jeziku (v poskusu so uporabili 17.000 najpogostejših besed v korpusu UKWaC z dvema milijardama besed);
- stavki, ki vsebujejo zaimke ali anaforične izraze (ta, tisti, itd.), imajo pogosto premalo sobesedila, da bi jih lahko dobro razumeli, zato jih v izboru ne upoštevajo;
- ustrenejši so stavki, v katerih se ciljna kolokacija pojavi v glavnem stavku;
- ustrenejši so celostavčni zgledi, ki jih je možno določiti z veliko začetnico in ločilom na koncu (pika, vprašaj, klicaj);
- ustrenejši so stavki z razširjenimi kolokacijami, to so take, ki presegajo klasične dvodelne kolokacije (*mednarodno ocenjevanje vin, pasti pod določeno, magično, psihološko, kritično mejo*);



g) opazili so tudi, da so ustrežnejši zgledi tisti, v katerih se najprej pojavi dovolj sobesedila, za katerim se pojavi kolokacija (tipično proti koncu stavka); gre za naravne stavke, ki s sobesedilom idealno pripravijo teren za ilustracijo kolokacije. Pri njih uporabniki najlaže iz sobesedila sklepajo na pomen kolokacije ali strukture.

Izbor dobrih zgledov v orodju SkE opravimo po tem, ko v Besedni skici označimo kolokatorje (gl. prejšnje poglavje, 2.3.2). Orodje predlaga želeno število zgledov, ki jih spet označimo v potrditvenih okencih ter vse skupaj prenesemo v odložišče (*Copy to clipboard*).

Tickbox Lexicography - Select Examples

Lemma: rana

Gramrel: a_modifier

Template: vanilla

vboden

- V kliničnem centru so sprejeli 160 ljudi , med njimi tudi človeka z vbodnimi **ranami** , j
- Osebe je oskrbelo še človeka z vbodnimi **ranami** ter nesrečnika z opeklina .
- Janko Makoter je v nočnih urah 26. septembra 1999 v lastni spalnici umrl zaradi vb
- Oskrbeli človeka z vbodnimi **ranami**
- Oskrbeli so vbodno **rano**

strelen

- Oskrbeli so tudi osebo s strelno **rano** .
- » Strelna **rana** , « si je rekel , » ne laserska .
- Uporabil je pištolo Makarov , kalibra 7,65 milimetra , strelna **rana** skozi glavo in mož
- Urgenco je obiskal poškodovanec s strelnimi **ranami** in drugi z vbodnimi ranami , dva
- Po njihovih podatkih ob truplu , ki je imelo dve strelni **rani** na zatilju , ni bilo nobenih c

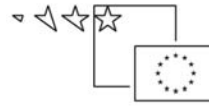
zaceljen

- Moje telo je žarelo iz teh točk kot znova oživiljeni človek , ki uživa v zaceljenih **ranah**
- Nekatere bo začelo trgati tudi po starih , zaceljenih **ranah** .

Slika 49: avtomatsko izdelan izbor dobrih zgledov (GDEX) v orodju SkE

2.3.5 Prenos zgledov v odložišče

Prenos korpusnih zgledov iz orodja SkE v odložišče je pri nastavitvah za projekt SSJ prilagojeno za uporabo pri sestavljanju leksikalne baze. To velja tako pri prenosu zgledov iz Konkordanc kot iz Besedne skice. V odložišče se prenese označen celostavčni zgled, ki je tudi ustrezno strukturno označen. Zglede lahko v odložišče prenašamo iz Konkordanc ali pa iz Izboru dobrih zgledov.



Prenos zgledov v **odložišče iz Konkordanc** poteka s klikom na ikono desno od konkordanc (📄), rezultat pa je:

Havelange bo moral spoznati, da se mora <i>pogajati</i> z nami.

To pa že nekaj pomeni, saj se s takšnimi močnimi državami ne moreš <i>pogajati</i> za delitev točk.

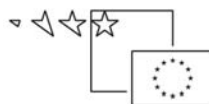
Unija želi v svoje vrste sprejeti združen otok, čeprav se o članstvu <i>pogaja</i> le z grškim delom.

Prenos zgledov v **odložišče iz Izbora dobrih zgledov (GDEX)** poteka s klikom v potrditvena okenca in na ikono za prenos v odložišče, rezultat pa je:

<zgled>Oskrbeli človeka z vbodnimi <i>ranami</i></zgled>

<zgled>Urgenco je obiskal poškodovanec s strelnimi <i>ranami</i> in drugi z vbodnimi ranami, dva z opeklinami in štirje, ki so se poškodovali pri ravnanju s petardami.</zgled>

<zgled>Nekatere bo začelo trgati tudi po starih, zaceljenih <i>ranah</i>.</zgled>



2.4 Napredne funkcije orodja SkE: *Corpus Builder*

2.4.1 Uvod

Osnovno orodje SkE dopolnjujeta dva posebna modula:

- **modul za razvoj slovnčnih relacij v vzorčnem korpusu (*Corpus Builder*)**, ki ga pri izdelavi leksikalne baze za slovenščino uporabljamo za razvoj in dopolnjevanje slovnčnih relacij, in
- **modul za izdelavo lastnega korpusa spletnih besedil (*WebBootCaT*)**. Ker ga pri izdelavi leksikalne baze za slovenščino ne uporabljamo, ga v kazalniku 5 ne opisujemo in tu le napotujemo na osnovno literaturo (Baroni et al. 2006a, Baroni et al. 2006b).

2.4.2 Predstavitev modula za razvoj slovnčnih relacij v vzorčnem korpusu (*Corpus Builder*)

Corpus Builder je modul za razvoj in dopolnjevanje slovnčnih relacij v vzorčnem korpusu. Namenjen je torej preizkušanju različnih nastavitvev in ogledu rezultatov v vzorčnem korpusu, ki ga proizvajalec glede na dogovor ustvari iz matičnega korpusa. Manjši korpus uporabljamo zato, da se tehnično dlje trajajoči postopki skrajšajo in da so vzorčni rezultati preglednejši.

Za sodelavce pri leksikalni bazi za slovenščino so v modulu *Corpus Builder* vzpostavljene naslednje možnosti:

- 20-milijonski vzorčni korpus iz FidePLUS, namenjen za preizkušanje,
- 20-milijonski kontrolni korpus iz FidePLUS,
- prostor za izdelavo korpusov do količine 120 milijonov besed.

Dostop do korpusov v modulu dodeli skrbnik pri proizvajalcu. Uporabniki kontrolnega korpusa ne spreminjajo, lahko pa vsak za lastno rabo ustvari kopijo vzorčnega korpusa in dodeli dostop do svojega korpusa ostalim članom skupine. Osnovna stran modula je:

Corpus Builder beta

[homepage](#) | [corpora](#) user: Polonca Kocjančič, free space: 120000000 tokens

Corpus list

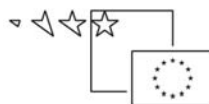
Create new corpus: [from template](#)

Corpus ID	Name	Metadata	Structures	Tokens [?]	
sgdev.fidaplus_20m.new	Fida PLUS 20m Polona	title	doc, p, s, g	0	[open in SkE] [expert config] [access] [edit] [del]
sgdev.fidaplus_20m.new.copy	Fida PLUS 20m SLOVASKI	title	doc, p, s, g	0	[open in SkE] [expert config] [access] [edit] [del]

My group's corpora

Owner ID	Owner name	Corpus	Metadata	Structures	Tokens [?]	
amebis01	Simon Krek	Fida PLUS 20m SimonK	title	doc, p, s, g	19070217	[del]
amebis01	Simon Krek	Fida PLUS 20m (v.13)	title	doc, p, s, g	19070217	[del]

Slika 50: osnovno okno modula *Corpus builder*



Za delo s slovničnimi relacijami so predvideni naslednji postopki:

```
merge >>> compile word sketches >>> recompute scores in ws >>> compile thesaurus
```

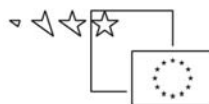
Slika 51: funkcije v modulu Corpus builder

merge: priprava korpusnega gradiva,

compile word sketches: uvoz datoteke s slovničnimi relacijami,

recompute scores in ws: izdelava Besednih skic z različnimi statističnimi izračuni,

compile thesaurus: izdelava tezavra.



2.5 Slovnične relacije v slovnici Besednih skic (SBS), v.15

2.5.1 Osnove sintakse slovničnih relacij

Sintaksa slovnice Besednih skic (dalje: SBS) uporablja sistem iskanja po vsebini atributov posamezne korpusne pojavnice s pomočjo regularnih izrazov. Atributi korpusne pojavnice so trije:

- word: besedna oblika oziroma zapis pojavnice, kakršen se nahaja v korpusu
- lemma: osnovna oblika besede pri pregibnih besednih oblikah
 - glagol: nedoločnik
 - samostalnik:
 - imenovalnik ednine
 - imenovalnik množine, če edninska oblika ne obstaja
 - zaimek (pregibni): moški spol, ednina, imenovalnik, nedoločna oblika
 - števnik:
 - moški spol, ednina, imenovalnik
 - imenovalnik moškega spola ustreznega števila, če druge oblike ne obstajajo
 - pridevnik: moški spol, ednina, imenovalnik, nedoločna oblika,¹ osnovnik
 - prislov: osnovnik
- tag: zapis oblikoskladenjske oznake po modelu Multext-East

Za lažjo predstavo navajamo nekaj primerov:

[word="levega"] – poiščemo vse pojavnice z zaporedjem črk "levega"

[lemma="most"] – poiščemo vse pojavnice z osnovno obliko samostalnika "most" (most, mosta, mostu, ...)

[tag="Somei"] – poiščemo vse pojavnice, pri katerih so bile pri oblikoskladenjskem označevanju prepoznane naslednje lastnosti: *samostalnik*, *občni*, *moški spol*, *ednina*, *imenovalnik*

[tag="S.*"] – poiščemo vse pojavnice, pri katerih je bila pri oblikoskladenjskem označevanju prepoznana lastnost *samostalnik*

[tag!="S.*"] – izločimo vse pojavnice, pri katerih je bila pri oblikoskladenjskem označevanju prepoznana lastnost *samostalnik*

[] – označuje katerokoli pojavnico s kakršnokoli lastnostjo

Podroben seznam oblikoskladenjskih oznak za slovenščino je na naslovu:

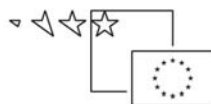
<http://www.sketchengine.co.uk/tagsets/slovene.html>, celotna dokumentacija pa na naslovu

<http://nl.ijs.si/ME/V3/msd/html/>. Natančnejši opis sintakse slovničnih relacij je na naslovu:

<http://trac.sketchengine.co.uk/wiki/SkE/CorpusQuerying>.

Poleg iskanja po vsebini atributov posamezne korpusne pojavnice s pomočjo regularnih izrazov sistem SkE uporablja še pet različnih vrst relacij. Te so lahko:

¹ V primerih, kjer obstajata nedoločna in določna oblika pridevnika, števnik, zaimka, sicer se uporabi edina oblika.



- enodelne: iskani izraz (*query*) vzpostavi razmerje z odvisnim izrazom, odnos je enosmeren; če poiščemo npr. samostalnik in s pomočjo SBS vzpostavimo razmerje s pridevnikom kot levim prilastkom, bo sistem pri samostalniških iskanjih izpisal niz pridevnikov, pri iskanju pridevnika pa samostalniški niz iz obratnega odnosa ne bo izpisan
- recipročne (DUAL): iskani izraz (*query*) vzpostavi razmerje z odvisnim izrazom, odnos je dvosmeren; če poiščemo npr. samostalnik in s pomočjo SBS vzpostavimo razmerje s pridevnikom kot levim prilastkom, bo sistem pri samostalniških iskanjih izpisal niz pridevnikov, pri iskanju pridevnika pa samostalniški niz iz obratnega odnosa
- tridelne (TRINARY): iskani izraz (*query*) vzpostavi razmerje z odvisnim izrazom, vmes pa se pojavlja še tretji element, tipično predlog; če poiščemo npr. glagol in s pomočjo SBS vzpostavimo razmerje s samostalnikom v predložni zvezi kot odvisnim izrazom, bo sistem pri glagolskih iskanjih izpisal niz samostalnikov, ki se pojavljajo za določenim predlogom
- simetrične (SYMMETRIC): iskani izraz (*query*) vzpostavi razmerje z drugim izrazom, odnos je simetričen, enakovreden; če poiščemo npr. samostalnik in s pomočjo SBS vzpostavimo razmerje z drugim samostalnikom, ki se pojavlja za prirednim veznikom, bo sistem pri samostalniških iskanjih izpisal drug niz samostalnikov
- unarne (UNARY): iskani izraz (*query*) ne vzpostavlja razmerja do drugega elementa z izpisom niza, temveč le zabeleži nadpovprečno pojavljanje tega razmerja v korpusu

2.5.2 Primer slovnične relacije

Za lažjo orientacijo pri delu z Besednimi skicami v nadaljevanju najprej predstavljamo eno od slovničnih relacij in njeno interpretacijo:

Opis delov slovnične relacije	Primer slovnične relacije #21
vrsta relacije =>	recipročna (*DUAL)
ime relacije =>	=v_rodil-s/s-koga-česa
definicija relacije =>	1:[tag="S.*"] 2:[tag="S...r.*"]

Slika 52: primer slovnične relacije v orodju SkE

Interpretacija definicije relacije:

Deli definicije slovnične relacije	Interpretacija
1:[tag="S.*"]	prvi element relacije je samostalnik
2:[tag="S...r.*"]	drugi element relacije je samostalnik v rodilniku

Slika 53: interpretacija definicije slovnične relacije v orodju SkE



Ker je relacija recipročna, bosta pri samostalniških iskanjih prikazana oba niza – prvi, kjer se ob iskanem samostalniku na desni strani pojavlja drug samostalnik v roditelju, in drugi, kjer se pri iskanem samostalniku ob roditeljski obliki tega samostalnika pojavlja na levi strani drug samostalnik. Rezultat, ki ga za lemo *ocenjevanje* prikaže ta slovnična relacija, je:

v_rodil-s	168	10.2	s-koga-česa	59	3.6
znanje	31	35.5	kriterij	5	20.33
vino	28	35.43	način	9	19.71
kakovost	9	19.77	metodologija	3	17.59
škoda	5	13.58	postopek	5	14.79
ustreznost	2	11.99	namen	2	8.93
urejenost	2	11.8	podlaga	2	8.93
skladnost	2	11.37	akcija	2	8.42
mleko	3	11.21	rezultat	2	7.99
živina	2	10.52	vrsta	2	7.08
vpliv	3	9.96	sistem	2	6.63
zadovoljstvo	2	9.42	leto	2	4.01
ponudba	3	9.21			
sistem	4	8.9			
rezultat	3	8.8			
izdelek	3	8.65			
učenec	2	8.1			
grad	2	7.97			
vzorec	2	7.8			
pes	2	7.25			
banka	2	5.88			

Slika 54: vzorčna slovnična relacija v Besedni skici

Kombinacijo pri razmerju "v_rodil-s" je torej potrebno brati:

ocenjevanje +
v_rodil-s = znanja, vin, kakovosti,...

Kombinacijo pri razmerju "s-koga-česa" je potrebno brati:

s = kriterij, način, metodologija,...
koga-česa = ocenjevanja

Primeri konkordanc pri tej slovnični relaciji so:



v_rodil-s:

1. element: lema *ocenjevanje* v kateremkoli sklonu,
2. element: samostalniški kolokator v rodilniku:

kakovost

- **Ocenjevanje** kakovosti storitve in organizacije, ki nudi storitev.
- Po namenu imamo dve vrsti **ocenjevanja** kakovosti storitve.
- Za operativni namen **ocenjevanja** kakovosti, velja osnovno pravilo, da so glavni ocenjevalci ljudje, ki so uporabniki oz. kupci storitev, ki se jo ocenjuje.

škoda

- Komisija za **ocenjevanje** škode je prejela 310 prijav.
- Podrta so posamična drevesa, škoda je razpršena, zato je delo gozdarjev in **ocenjevanje** škode še zahtevnejše.
- Škodo je državna komisija za **ocenjevanje** škode po naravnih in drugih nesrečah ocenila na 31,276 milijarde tolarjev.

Slika 55: rezultat slovnice relacije v_rodil-s

s-koga-česa:

1. element: kolokator v kateremkoli sklonu,
2. element: lema *ocenjevanje* v rodilniku:

kriterij

- Kriteriji **ocenjevanja**, v modelih odličnosti, nudijo osnovo, da lahko organizacija primerja svoje delovanje z delovanjem ostalih organizacij.
Sicer pa se bo na filatelistični razstavi 75 eksponatov iz devetih držav potegovalo po
- kriterijih **ocenjevanja** svetovnih filatelističnih razstav za osvojitev najvišjega odličja, zlatega poštnega roga.
- Prav tako definirajo tudi cilje, vsebino in metode pouka ter kriterije **ocenjevanja**.

način

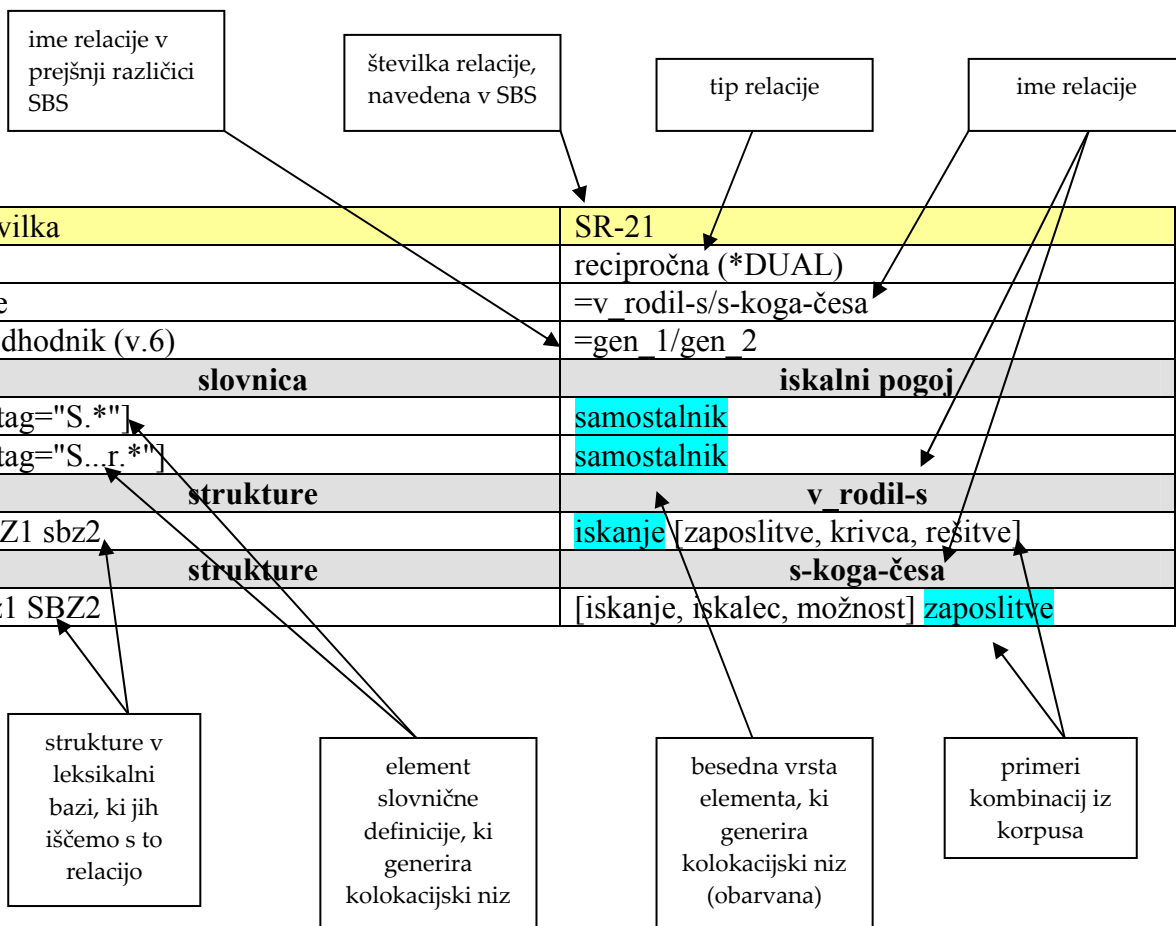
- Obsega dva sestavna dela, za vsakega je na razpolago tudi način **ocenjevanja**.
- Način dela pri predmetu in način **ocenjevanja** znanja Predavanja in vaje.
- 9 Ekonomska politika Način dela pri predmetu in način **ocenjevanja** znanja Predavanja in vaje.

Slika 56: rezultat slovnice relacije s-koga-česa



V nadaljevanju navajamo vse slovnične relacije v SBS, verzija 15, po vrsti, z naslednjimi podatki:

številka	SR-21
tip	recipročna (*DUAL)
ime	=v_rodil-s/s-koga-česa
predhodnik (v.6)	=gen_1/gen_2
slovnica	iskalni pogoj
1:[tag="S.*"]	samostalnik
2:[tag="S...r.*"]	samostalnik
strukture	v_rodil-s
SBZ1 sbz2	iskanje [zaposlitve, krivca, rešitve]
strukture	s-koga-česa
sbz1 SBZ2	[iskanje, iskalec, možnost] zaposlitve



Slika 57: podatki o slovničnih relacijah v SBS

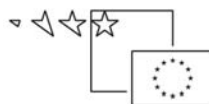


2.5.3 Slovnice relacije v SBS v.15

številka	SR-01
tip	recipročna (*DUAL)
ime	=kakšen?/kdo-kaj?
predhodnik (v.6)	=a_modifier/modifies
slovnica	iskalni pogoj
1: [tag="S.*"]	samostalnik
2: [tag="P.*"]	pridevnik
strukture	kakšen?
pbz SBZ1	[boleč, lep] spomin
strukture	kdo-kaj?
PBZ sbz1	rdeča [žoga]

številka	SR-02
tip	recipročna (*DUAL)
ime	=kako-kdaj?/je_pred
predhodnik (v.6)	=adv_modifies =adj_modified
slovnica	iskalni pogoj
1:[tag="[PG].*"]	pridevnik, glagol
2:[tag="R.*"]	prislov
strukture	kako-kdaj?
rbz GBZ	[nesrečno] pasti, [neznosno] boleti
rbz PBZ	[kristalno] čist; [neznansko] vesel
strukture	je_pred
RBZ gbz	močno [deževati], drastično [se znižati]
RBZ pbz	bledo [rdeč], hudo [zapat]

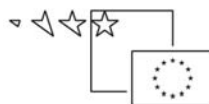
številka	SR-03
tip	enodelna
ime	=veznik
predhodnik (v.6)	--
slovnica	iskalni pogoj
1:[tag="[GPRS].*"]	glagol, pridevnik, prislov, samostalnik
strukture	veznik
PBZ Odv-da	vesel, da je kaj
SBZ1 Odv-ali	vprašanje, ali...
RBZ Odv- da	preprosto, da (bolj ne more biti)
GBZ Odv-da	vedeti, da ...



številka	SR-04
tip	enodelna
ime	=predlog
predhodnik (v.6)	=prec_prep
slovnica	iskalni pogoj
1:[]	glagol, pridevnik, prislov, samostalnik
strukture	predlog
	kot nor
	na bolje
	pred časom

številka	SR-05
tip	enodelna
ime	=predl-za
predhodnik (v.6)	=post_prep
slovnica	iskalni pogoj
1:[]	glagol, pridevnik, prislov, samostalnik
strukture	predl-za
	verjeti v
	zanimiv za
	najpozneje do
	voda iz

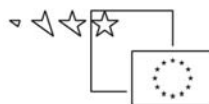
številka	SR-06
tip	tridelna (*TRINARY)
ime	=%s
predhodnik (v.6)	=post_%s
slovnica	iskalni pogoj
1:[tag="G.*"]	glagol
strukture	=%s
sbz1 GBZ sbz4 (s sbz6 / med sbz6)	stisniti s [prsti, palci]
sbz1 GBZ sbz4 (s sbz6 / med sbz6)	stisniti med [prsti]
sbz1 GBZ (o sbz5 glede sbz2 za sbz4)	pogajati se o [vdaji, izpustitvi]



številka	SR-07
tip	tridelna (*TRINARY)
ime	=%s
predhodnik (v.6)	=post_%s
slovnica	iskalni pogoj
1:[tag="S.*"]	samostalnik
strukture	=%s
SBZ1 na sbz4	spomin na [otročtvo, mladost]
SBZ1 pod sbz6	rezultat pod [pričakovanji]
SBZ1 okoli sbz2	ograja okoli [hiše]
SBZ1 ob sbz5	incident ob [meji]

številka	SR-08
tip	tridelna (*TRINARY)
ime	=%s_X
predhodnik (v.6)	=prec_%s
slovnica	iskalni pogoj
1:[tag="S.*"]	samostalnik
strukture	%s_X
sbz1 glede SBZ2	[pričakovanja] glede zaposlovanja
sbz1 nad SBZ6	[skrbništvo] nad otrokom
sbz1 čez SBZ4	[most] čez reko
sbz1 do SBZ2	[čas] do volitev

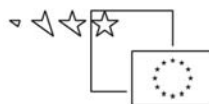
številka	SR-09
tip	tridelna (*TRINARY)
ime	=%s_X
predhodnik (v.6)	=prec_%s
slovnica	iskalni pogoj
1:[tag="S.*"]	samostalnik
strukture	%s_X
gbz k SBZ3	[privezati] k postelji
gbz na SBZ4	[pomisliti] na otroštvo
gbz preko SBZ2	[stopiti] preko potoka
gbz pri SBZ5	[ostati] pri življenju



številka	SR-10
tip	recipročna (*DUAL)
ime	=koga-česa/v_rodil
predhodnik (v.6)	=is_obj2/has_obj2
slovnica	iskalni pogoj
1:[tag="S...r.*"]	samostalnik
2:[tag="Gp.*"]	glagol
strukture	koga-česa
sbz1 GBZ sbz2	izogibati se [stikom]
sbz1 GBZ sbz2	bati se [maščevanja]
strukture	v_rodil
gbz SBZ2	[manjkati] denarja
gbz SBZ2	[braniti se] očitkov

številka	SR-11
tip	recipročna (*DUAL)
ime	=komu-čemu/v_dajal
predhodnik (v.6)	=is_obj3/has_obj3
slovnica	iskalni pogoj
1:[tag="So..d.*"]	samostalnik
2:[tag="Gp.*"]	glagol
strukture	komu-čemu
sbz1 GBZ sbz3	izogibati se [naporu], uiti [smrti]
strukture	v_dajal
gbz SBZ3	[predajati se] spominom [odpovedati se] ljubezni

številka	SR-12
tip	recipročna (*DUAL)
ime	=koga-kaj/v_tožil
predhodnik (v.6)	=is_obj4/has_obj4
slovnica	iskalni pogoj
1:[tag="So..t.*"]	samostalnik
2:[tag="Gp.*"]	glagol
strukture	koga-kaj
sbz1 GBZ sbz4	videti [priložnost]
sbz1 GBZ sbz4	sezuti [škornje]
strukture	v_tožil
gbz SBZ4	[zapravljati, izgubljati] čas
gbz SBZ4	[izgubiti] nedolžnost



številka	SR-13
tip	recipročna (*DUAL)
ime	=osebek_od/osebek_je
predhodnik (v.6)	=is_subj/has_subj
slovnica	iskalni pogoj
1:[tag="So..i.*"]	samostalnik
2:[tag="Gpp.t.*"]	glagol
strukture	osebek_od
SBZ1 gbz	spomini [oživijo]
SBZ1 gbz	zastava [visi]
strukture	osebek_je
sbz1 GBZ sbz3	[sodišče] verjame
sbz1 GBZ	[odbor] se sestane

številka	SR-14
tip	enodelna
ime	=nedoločnik
predhodnik (v.6)	=post_inf
slovnica	iskalni pogoj
1:[]	glagol, pridevnik, prislov, samostalnik
strukture	nedoločnik
GBZ Inf-gbz	uspeti [doseči]
SBZ1 Inf-gbz	(imeti) možnost [pritožiti se]
PBZ Inf-gbz	pripravljen [oditi]; sposoben [izpeljati]
RBZ Inf-gbz	dobro [izkoristiti]

številka	SR-15
tip	enodelna
ime	=GPRS-inf
predhodnik (v.6)	--
slovnica	iskalni pogoj
1:[tag="G.n.*"]	glagol
strukture	GPRS-inf
gbz Inf-GBZ	[znati] pasti, [nehati] boleti, [morati] paziti
	[treba] vedeti



številka	SR-16
tip	recipročna (*DUAL)
ime	=zanikan/z_nikalnim
predhodnik (v.6)	--
slovnica	iskalni pogoj
1:[tag="So..r"]	samostalnik
2:[tag="Gp.*"]	glagol
strukture	zanikan
Nglag SBZ2	[ne izgubljeni, ne zapravljati] časa
strukture	zanikan
	ne najti [dokazov]

številka	SR-17
tip	recipročna (*DUAL)
ime	=količina_ob-s/količinski
predhodnik (v.6)	--
slovnica	iskalni pogoj
1:[tag="R.*"]	prislov
2:[tag="S...r.*"]	samostalnik
strukture	količinski
prislk SBZ2	[nekaj, precej] časa
strukture	količina_ob-s
Kol-RBZ sbz2	precej [pozornosti, denarja]

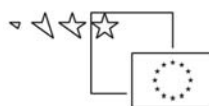
številka	SR-18
tip	recipročna (*DUAL)
ime	=s_prislovom/s_prislovom
predhodnik (v.6)	--
slovnica	iskalni pogoj
1:[tag="So..r.*"]	samostalnik
2:[tag="Gp.*"]	glagol
strukture	s_prislovom
gbz prislk SBZ2	[porabiti, preživeti] [veliko, precej] časa
strukture	s_prislovom
	kazati [veliko, precej] zanimanja



številka	SR-19
tip	recipročna (*DUAL)
ime	=kakšen-p?/kakšen-g?
predhodnik (v.6)	--
slovnica	iskalni pogoj
1:[tag="Gp.*"]	glagol
2:[tag="P...i.*"]	pridevnik
strukture	kakšen-p?
pbz GBZ	[vesel] razlagati
Vez-GBZ pbz	izgledati [utrujen], izpasti [neumen]
strukture	kakšen-g?
gbz PBZ	[vstati] neprespan, [razlagati] vesel
Vez-gbz PBZ	[ostati, postati] vesel [zdeti se] vesel

številka	SR-20
tip	recipročna (*DUAL)
ime	=kakšnega-p/kakšnega-g?
predhodnik (v.6)	--
slovnica	iskalni pogoj
1:[tag="Gp.*"]	glagol
2:[tag="P...r.*"]	pridevnik
strukture	kakšnega-p
Vez-GBZ pbz2	čutiti se [osamljenega]
strukture	kakšnega-g?
Vez-gbz PBZ2	[čutiti se] veselega

številka	SR-21
tip	recipročna (*DUAL)
ime	=v_rodil-s/s-koga-česa
predhodnik (v.6)	=gen_1/gen_2
slovnica	iskalni pogoj
1:[tag="S.*"]	samostalnik
2:[tag="S...r.*"]	samostalnik
strukture	v_rodil-s
SBZ1 sbz2	iskanje [zaposlitve, krivca, rešitve]
strukture	s-koga-česa
sbz1 SBZ2	[iskanje, iskalec, možnost] zaposlitve



številka	SR-22
tip	recipročna (*DUAL)
ime	=oba-v-rod/oba-v-rod
predhodnik (v.6)	--
slovnica	iskalni pogoj
1:[tag="P...r.*"]	pridevnik
2:[tag="S...r.*"]	samostalnik
strukture	oba-v-rod
pbz SBZ2	[šibkega] zdravja
strukture	oba-v-rod
PBZ sbz2	bledega [obraza]

številka	SR-23
tip	tridelna (*TRINARY)
ime	=%s
predhodnik (v.6)	--
slovnica	iskalni pogoj
1:[tag="P.*"]	pridevnik
strukture	%s
PBZ glede sbz2	negotov glede [meril] enoten glede [sodelovanja]
PBZ na sbz4	pogolten na [denar] prijeten na [pogled]
PBZ od sbz2	zaripel od [napora] bled od [jeze] odrezan od [sveta]

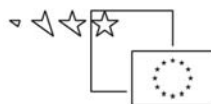
številka	SR-24
tip	recipročna (*DUAL)
ime	=v_rodil-p/p-koga-česa
predhodnik (v.6)	--
slovnica	iskalni pogoj
1:[tag="P...i.*"]	pridevnik
2:[tag="S...r.*"]	samostalnik
strukture	v_rodil-p
PBZ sbz2	[spoštovanja, omembe] vreden
strukture	p-koga-česa
pbz SBZ2	[poln, vreden] denarja



številka	SR-25
tip	recipročna (*DUAL)
ime	=v_dajal-p/p-komu-čemu
predhodnik (v.6)	--
slovnica	iskalni pogoj
1:[tag="P...i.*"]	pridevnik
2:[tag="S...d.*"]	samostalnik
strukture	v_dajal-p
PBZ sbz3	zvest [prijatelju, tradiciji]
strukture	p-komu-čemu
pbz SBZ3	[nevaren] državi

številka	SR-26
tip	recipročna (*DUAL)
ime	=v_tožil-p/p-koga-kaj
predhodnik (v.6)	--
slovnica	iskalni pogoj
1:[tag="P...i.*"]	pridevnik
2:[tag="S...t.*"]	samostalnik
strukture	v_tožil-p
PBZ sbz4	vreden [milijardo]
strukture	p-koga-kaj
pbz SBZ4	[dolžen] vsoto

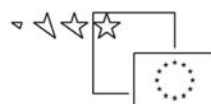
številka	SR-27
tip	recipročna (*DUAL)
ime	=biti_kakšen?/osebek+biti
predhodnik (v.6)	=biti_adj/subj_biti
slovnica	iskalni pogoj
1:[tag="S...i.*"]	samostalnik
2:[tag="P...i.*"]	pridevnik
strukture	biti_kakšen?
SBZ1 Vez-gbz pbz	spomin je [boleč, živ]
strukture	osebek+biti
Vez-gbz PBZ	[situacija, postrežba] je slaba



številka	SR-28
tip	simetrična (*SYMMETRIC)
ime	=priedje
predhodnik (v.6)	=coord
slovnica	iskalni pogoj
1:[tag="[GPRS].*"]	glagol, pridevnik, prislov, samostalnik
2:[tag="[GPRS].*"]	glagol, pridevnik, prislov, samostalnik
strukture	priedje
GBZ in gbz	pasti in se poškodovati
gbz in GBZ	spotakniti se in pasti; peči in boleti
GBZ ali gbz	kupiti ali najeti
	čas in denar
	hladen in vlažen
	sončno in vroče

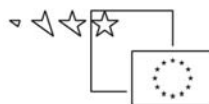
številka	SR-29
tip	recipročna (*DUAL)
ime	=s_trpnikom/trpnik
predhodnik (v.6)	=subj_of_passive/vb_passive
slovnica	iskalni pogoj
1:[tag="G.....r.*"]	glagol
2:[tag="S...i.*"]	samostalnik
strukture	s_trpnikom
	[stavba, hiša, cerkev] je zgrajena
strukture	trpnik
	cesta je [speljana, zaprta, asfaltirana]

številka	SR-30
tip	enodelna
ime	=gl-pred
predhodnik (v.6)	=prec_verb
slovnica	iskalni pogoj
1:[]	glagol, pridevnik, prislov, samostalnik
strukture	gl-pred
	sedeti [vzravhati, videti]
	lačen [nahraniti, ostati]
	slabo [igrati, počutiti]
	violina [igrati, učiti, poučevati]



številka	SR-31
tip	enodelna
ime	=gl-za
predhodnik (v.6)	=post_verb
slovnica	iskalni pogoj
1:[]	glagol, pridevnik, prislov, samostalnik
strukture	gl-za
	videti [slišati, občutiti]
	veljaven [predpisovati, obdavčiti]
	vidno [označiti, razočarati, pretresti]
	časopis [objaviti, objavljati, poročati]

številka	SR-32
tip	unarna (*UNARY)
ime	=štev-pred
predhodnik (v.6)	--
slovnica	iskalni pogoj
1:[tag="S.*"]	samostalnik
strukture	štev-pred
	mesto, dan, mesec

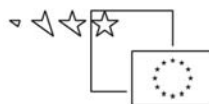


2.5.4 Slovnice relacije po besednih vrstah s primeri iz korpusa (v.15)

V nadaljevanju so zbrane slovnice relacije po besednih vrstah, s primeri iz korpusa:

		samostalnik-27
1	SR-01	boleč/lep spomin
2	SR-03	vprašanje , ali...
3	SR-04	pred časom
4	SR-05	voda iz
5	SR-07	spomin na otroštvo/mladost
6	SR-08	most čez reko
7	SR-09	pomisliti na otročstvo
8	SR-10	braniti se očitkov
9	SR-11	odpovedati se ljubezni
10	SR-12	izgubiti nedolžnost
11	SR-13	zastava visi
12	SR-14	možnost pritožiti se
13	SR-16	ne izgubljati/zapravljeni časa
14	SR-17	nekaj/precej časa
15	SR-18	porabiti/preživeti veliko/precej časa
16	SR-21	iskanje zaposlitve/krivca/rešitve
17	SR-21	iskanje/iskalec/možnost zaposlitve
18	SR-22	šibkega zdravja
19	SR-24	poln/vreden denarja
20	SR-25	nevaren državi
21	SR-26	dolžen vsoto
22	SR-27	spomin je boleč/živ
23	SR-28	čas in denar
24	SR-29	cesta je speljana/zaprta/asfaltirana
25	SR-30	violina [igrati, učiti, poučevati]
26	SR-31	časopis [objaviti, objavljati, poročati]
27	SR-32	mesto, dan, mesec

		glagol-18
1	SR-02	nesrečno pasti , neznosno boleti
2	SR-03	vedeti , da ...
3	SR-05	verjeti v
4	SR-06	pogajati se o vdaji/izpustitvi
5	SR-10	bati se maščevanja
6	SR-11	izogibati se naporu
7	SR-12	videti priložnost



8	SR-13	odbor se sestane
9	SR-14	uspeti doseči
10	SR-15	nehati boleti
11	SR-16	ne najti dokazov
12	SR-18	kazati veliko/precej zanimanja
13	SR-19	vesel razlagati, izgledati utrujen
14	SR-20	čutiti se osamljenega
15	SR-28	pasti in se poškodovati
16	SR-29	stavba/hiša/cerkev je zgrajena
17	SR-30	sedeti [vzravhati, videti]
18	SR-31	videti [slišati, občutiti]

pridevnik-17		
1	SR-01	rdeča žoga
2	SR-02	kristalno čist ; neznansko vesel
3	SR-03	vesel , da je kaj
4	SR-04	kot nor
5	SR-05	zanimiv za
6	SR-14	pripravljen oditi; sposoben izpeljati
7	SR-19	vstati neprespan , zdeti se vesel
8	SR-20	čutiti se veselega
9	SR-22	bledega obraza
10	SR-23	enoten glede sodelovanja
11	SR-24	spoštovanja/omembe vreden
12	SR-25	zvest prijatelju/tradiciji
13	SR-26	vreden milijardo
14	SR-27	situacija/postrežba je slaba
15	SR-28	hladen in vlažen
16	SR-30	lačen [nahriniti, ostati]
17	SR-31	veljaven [predpisovati, obdavčiti]

prislov-9		
1	SR-02	močno deževati, hudo zapit
2	SR-03	preprosto , da (bolj ne more biti)
3	SR-04	na bolje
4	SR-05	najpozneje do
5	SR-14	dobro izkoristiti
6	SR-17	precej pozornosti/denarja
7	SR-28	sončno in vroč
8	SR-30	slabo [igrati, počutiti]
9	SR-31	vidno [označiti, razočarati, pretresti]



2.5.5 Dokument s slovničnimi relacijami: FidaPLUS sketch grammar ver. 15

```
# #####  
  
# FidaPLUS sketch grammar ver. 15  
  
# Simon Krek  
# Sporazumevanje v slovenskem jeziku  
# http://www.slovenscina.eu/  
  
*STRUCTLIMIT s  
  
# SR-01 #####  
  
*DUAL  
=kakšen?/kdo-kaj?  
2: [tag="P.*"] [tag="[PRZKVL].*" | word=", " | word="se" | word="si"] {0,5} 1: [tag="S.*"]  
  
# SR-02 #####  
  
*DUAL  
=kako-kdaj?/je_pred  
2: [tag="R.*"] [] {0,1} 1: [tag="[PG].*"]  
  
# SR-03 #####  
  
=veznik  
1: [tag="[GPRS].*" | word=", " | tag="[DV].*" ] {0,1} 2: [tag="V.*" | tag="Z[nvz].*"]  
  
# SR-04 #####  
  
=predlog
```




2:[tag="D.*"] 1:[]

SR-05

=predl-za

1:[] 2:[tag="D.*"]

SR-06

*TRINARY

=%s

1:[tag="G.*"] [word="se"|word="si"]{0,2} 3:[tag="D.*"] [tag="[PRZKLV].*"|word=", "]{0,5} 2:[tag="S.*"]
3:[tag="D.*"] [tag="[PRZKLV].*"|word=", "]{0,5} 2:[tag="S.*"] [word="se"|word="si"]{0,1} 1:[tag="G.*"]

SR-07

*TRINARY

=%s

1:[tag="S.*"] 3:[tag="D.*"] [tag="[PRZKLV].*"|word=", "]{0,5} 2:[tag="S.*"]
1:[tag="S.*"] 3:[tag="D.*"] [tag="[RZKLV].*"|word=", "]{0,4} 2:[tag="P.*"] [tag="S.*"]

SR-08

*TRINARY

=%s_X

2:[tag="S.*"] 3:[tag="D.*"] [tag="[PRZKLV].*"|word=", "]{0,5} 1:[tag="S.*"]

SR-09

*TRINARY

=%s_X



```
2:[tag="G.*"] [word="se"|word="si"]{0,2} 3:[tag="D.*"]  
[tag="[PRZKLV].*"|word=","|word="se"|word="si"]{0,5} 1:[tag="S.*"]  
3:[tag="D.*"] [tag="[PRZKLV].*"|word=","|word="si"]{0,5} 1:[tag="S.*"] [word="se"|word="si"]{0,1} 2:[tag="G.*"]
```

SR-10

```
*DUAL  
=koga-česa/v_rodil  
2:[tag="Gp.*"] [tag!="[SDKRV].*" & tag!="Z[vn].*"]{0,4} 1:[tag="S...r.*"]
```

SR-11

```
*DUAL  
=komu-čemu/v_dajal  
2:[tag="Gp.*"] [tag!="[DVMO].*"]{0,5} 1:[tag="So..d.*"]  
[tag!="[SKPDVMO].*"] 1:[tag="So..d.*"] 2:[tag="Gp.*"]
```

SR-12

```
*DUAL  
=koga-kaj/v_tožil  
2:[tag="Gp.*"] [tag!="[GD].*"]{0,5} 1:[tag="So..t.*"]  
[tag!="D.*"] [tag!="D.*"] 1:[tag="So..t.*"] 2:[tag="Gp.*"]  
<s>1:[tag="So..t.*"] 2:[tag="Gp.*"]
```

SR-13

```
*DUAL  
=osebek_od/osebek_je  
1:[tag="So..i.*"] []{0,5} 2:[tag="Gpp.t.*"]  
1:[tag="So..i.*"] []{0,2} [lemma="biti"] [tag!="G.*"]{0,3} 2:[tag="Gpd.*"]  
2:[tag="Gpp.t.*"] []{0,5} 1:[tag="So..i.*"]  
[lemma="biti"] [tag!="G.*"]{0,2} 2:[tag="Gpd.*"] []{0,3} 1:[tag="So..i.*"]
```



SR-14

=nedoločnik
1:[] 2:[tag="G.n.*"]

SR-15

=GPRS-inf
2:[tag="G.*"][tag!="G.*"|lemma="biti"]{0,4}1:[tag="G.n.*"]
[tag="G.*"]{0,4}2:[tag="G.*" & lemma!="biti"]1:[tag="G.n.*"]
[tag="G.*" & lemma="biti"][tag!="D.*"]{0,4}2:[tag="R.*"]1:[tag="G.n.*"]
[tag="G.*" & lemma="biti"][tag!="D.*"]{0,4}2:[tag="P.*"]1:[tag="G.n.*"]
[tag="G.*"][tag!="D.*"]{0,4}2:[tag="So.*"]1:[tag="G.n.*"]

SR-16

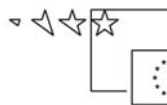
*DUAL
=zanikan/z_nikalnim
[lemma="biti" & tag=".....d.*"|lemma="imeti" & tag=".....d.*"|lemma="hoteti" &
tag=".....d.*"|word="ne"|word="ni"]{1,5} 2:[tag="Gp.*"] [tag!="[GD].*"]{0,5} 1:[tag="So..r"]

SR-17

*DUAL
=količina_ob-s/količinski
1:[tag="R.*"] [tag="[PZKL].*"]{0,3} 2:[tag="S...r.*"]

SR-18

*DUAL



```
=s_prislovom/s_prislovom
  2:[tag="Gp.*"] [tag!="[GD].*"]{0,5} [tag="R.*"] 1:[tag="So..r.*"]
  [tag!="D.*"] [tag="R.*"] 1:[tag="So..r.*"] 2:[tag="Gp.*"]

# SR-19 #####

*DUAL
=kakšen-p/?kakšen-g?
  1:[tag="Gp.*"] [tag="[RZKL].*"]{0,3}2:[tag="P....i.*"] [word="[\.,;:]"]

# SR-20 #####

*DUAL
=kakšnega-p/kakšnega-g?
  1:[tag="Gp.*"] [tag="[RZKL].*"]{0,3}2:[tag="P....r.*"] [word="[\.,;:]"]

# SR-21 #####

*DUAL
=v_rodil-s/s-koga-česa
  1:[tag="S.*"] 2:[tag="S...r.*"]

# SR-22 #####

*DUAL
=oba-v-rod/oba-v-rod
  1:[tag="P....r.*"] 2:[tag="S...r.*"]

# SR-23 #####

*TRINARY
=%s
```



```
1:[tag="P.*"] 3:[tag="D.*"] [tag="[PRZKLV].*"|word=",,"]{0,5} 2:[tag="S.*"]  
1:[tag="P.*"] 3:[tag="D.*"] [tag="[RZKLV].*"|word=",,"]{0,4} 2:[tag="P.*"] [tag="S.*"]
```

SR-24

*DUAL

=v_rodil-p/p-koga-česa

```
1:[tag="P....i.*"] [tag!="[DS].*"]{0,1} 2:[tag="S...r.*"]  
[tag!="[DS].*"]{1,4} 2:[tag="S...r.*"] []{0,1} 1:[tag="P....i.*"]
```

SR-25

*DUAL

=v_dajal-p/p-komu-čemu

```
1:[tag="P....i.*"] [tag!="D.*"]{0,1} 2:[tag="S...d.*"]  
[tag!="D.*"]{1,4} 2:[tag="S...d.*"] []{0,1} 1:[tag="P....i.*"]
```

SR-26

*DUAL

=v_tožil-p/p-koga-kaj

```
1:[tag="P....i.*"] [tag!="D.*"]{0,1} 2:[tag="S...t.*"]  
[tag!="D.*"]{1,4} 2:[tag="S...t.*"] []{0,1} 1:[tag="P....i.*"]
```

SR-27

*DUAL

=biti_kakšen?/osebek+biti

```
1:[tag="S...i.*"] [lemma="biti" & tag="G.*"] [tag!="Gp.*"]{0,5} 2:[tag="P....i.*"] [tag!="S.*" &  
tag!="Gpn.*"]{0,5}<s>  
1:[tag="S...i.*"] [lemma="biti" & tag="G.*"] [tag!="Gp.*"]{0,5} [tag="P....i.*"] [word="in" | word="ali" |  
word="ter"] []{0,2} 2:[tag="P.*"] [tag!="S.*" & tag!="Gpn.*"]{0,5}<s>
```



```
1:[tag="S...i.*"] [lemma="biti" & tag="G.*"] [tag!="Gp.*"]{0,5} 2:[tag="P....i.*"] [word="in" | word="ali" | word="ter"] []{0,2} [tag="P.*"] [tag!="S.*" & tag!="Gpn.*"]{0,5}<s>
1:[tag="S...i.*"] [lemma="biti" & tag="G.*"] [tag!="Gp.*"]{0,5} 2:[tag="P....i.*"] [word="[,;:]"]
1:[tag="S...i.*"] [lemma="biti" & tag="G.*"] [tag!="Gp.*"]{0,5} 2:[tag="P....i.*"] [word="in" | word="ali" | word="ter"] []{0,2} [tag="P....i.*"] [word="[,;:]"]
1:[tag="S...i.*"] [lemma="biti" & tag="G.*"] [tag!="Gp.*"]{0,5} [tag="P....i.*"] [word="in" | word="ali" | word="ter"] []{0,2} 2:[tag="P....i.*"] [word="[,;:]"]
2:[tag="P....i.*"] [lemma="biti" & tag="G.*"] [word!="[,;:]" & tag!="Gp.*"]{0,5} 1:[tag="So..i.*"]
[tag="P....i.*"] [word="in" | word="ali" | word="ter"] []{0,2} 2:[tag="P....i.*"] [lemma="biti" & tag="G.*"]
[word!="[,;:]" & tag!="Gp.*"]{0,5} 1:[tag="So..i.*"]
```

SR-28

=priedje

*SYMMETRIC

```
1:[tag="[GPRS].*"] [word="in"|word="ali"|word="ter"] 2:[tag="[GPRS].*"]
2:[tag="[GPRS].*"] [word="in"|word="ali"|word="ter"] 1:[tag="[GPRS].*"]
[word = "niti"] 1:[] [word = "niti"] 2:[]
[word = "bodisi"] 1:[] [word = "bodisi"] 2:[]
[word = "tako"] 1:[] [word = "kakor"] 2:[]
[word = "tako"] 1:[] [word = "kot"] 2:[]
```

SR-29

*DUAL

=s_trpnikom/trpnik

```
2:[tag="S...i.*"] []{0,5} 1:[tag="G.....r.*"]
1:[tag="G.....r.*"] []{0,5} 2:[tag="S...i.*"]
```

SR-30

=gl-pred

```
2:[tag="Gp.*"] []{0,5} 1:[]
```



SR-31

=gl-za

1:[] []{0,5} 2:[tag="Gp.*"]

SR-32

*UNARY

=štev-pred

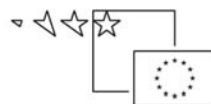
[tag="K.*"] 1:[tag="S.*"]

#####



2.6 Izrazje, povezano z orodjem SkE

<i>angleško</i>	<i>slovensko</i>
clustering	gručenje kolokatorjev
Collocation	Kolokacija
Common patterns	Tabela skupnih vzorcev
Concordance	Konkordance
Filter	Filter
Frequency	Pogostost
GDEX (Good examples)	Izbor dobrih zgledov
grammatical relations (gramrel)	slovnice relacije
Multilevel frequency distribution	Večnivojska distribucija pogostosti
Multilevel Sort	Večnivojsko razvrščanje
Sample	Vzorec
Save	Shrani
Simple Sort	Preprosto razvrščanje
Sketch grammar	slovnica Besednih skic
Sketch-diff	Primerjalna skica
Sort	Razvrščanje
TBL (Tick-box lexicography)	Klikosikografija
Text type frequency distribution	Distribucija pogostosti glede na besedilo
Thesaurus	Tezaver
View options	Možnosti prikaza
Word List	Seznam besed
Word Sketch	Besedna skica



3

Pretvornik

*za konkordance
in kolokacije*



3 Pretvornik za konkordance in kolokacije

3.1 Uvod

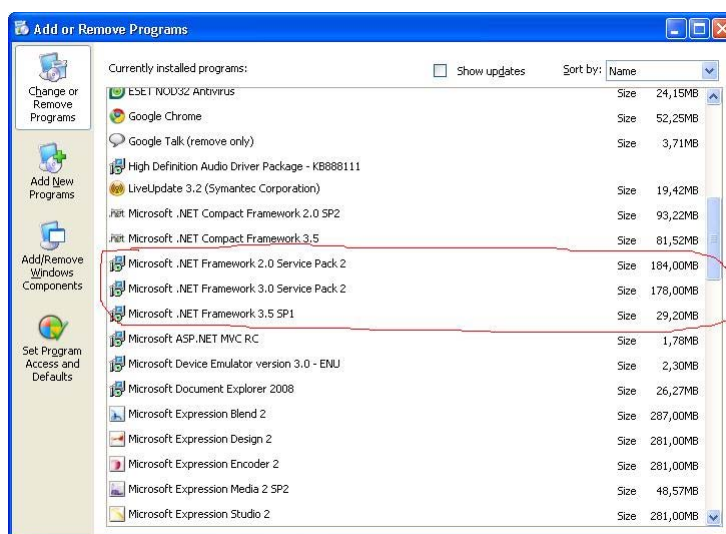
Pretvornik je namenjen preoblikovanju jezikovnih podatkov v formatu XML, ki jih dobimo iz besedilnega korpusa FidaPLUS, v obliko za uporabo v MS Wordu ali MS Excelu, v kateri je potem možno podatke obdelovati še na načine, ki jih izvorni programi ne omogočajo. Podatke iz besedilnega korpusa dobimo bodisi z orodjem *Sketch Engine beta* (dalje: SkE beta) (<http://beta.sketchengine.co.uk/>) ali s klasičnim konkordancnim ASP32 (<http://www.fidaplus.net>). Pretvornik uporabljamo sodelavci pri leksikalni bazi za vsebinsko delo s korpusnimi podatki: priprava pomenske strukture gesla, analiza kolokacij, označevanje, beleženje itd.

3.2 Namestitev

Datoteko **converter.zip** (25. 3. 2009) razširimo v mapo po svoji izbiri in z **.exe** po želji naredimo bližnjico na namizje. S tem je program pripravljen za uporabo. Preverimo pa še naslednje:

a) Če nimamo Officea 2007, namestimo filter: *Microsoft Office Compatibility Pack for Word, Excel, and PowerPoint 2007 File Formats*. Najdemo ga na naslovu:
<http://www.microsoft.com/downloads/details.aspx?FamilyId=941B3470-3AE9-4AEE-8F43-C6BB74CD1466&displaylang=en>

b) Pri starejših konfiguracijah se lahko zgodi, da ni inštaliran .net Framework; program na to opozori s sistemskim sporočilom »Pravilno inicializiranje programa ni uspelo«. V tem primeru preverimo, ali je .net Framework nameščen, in sicer v Dodaj/odstrani programe:



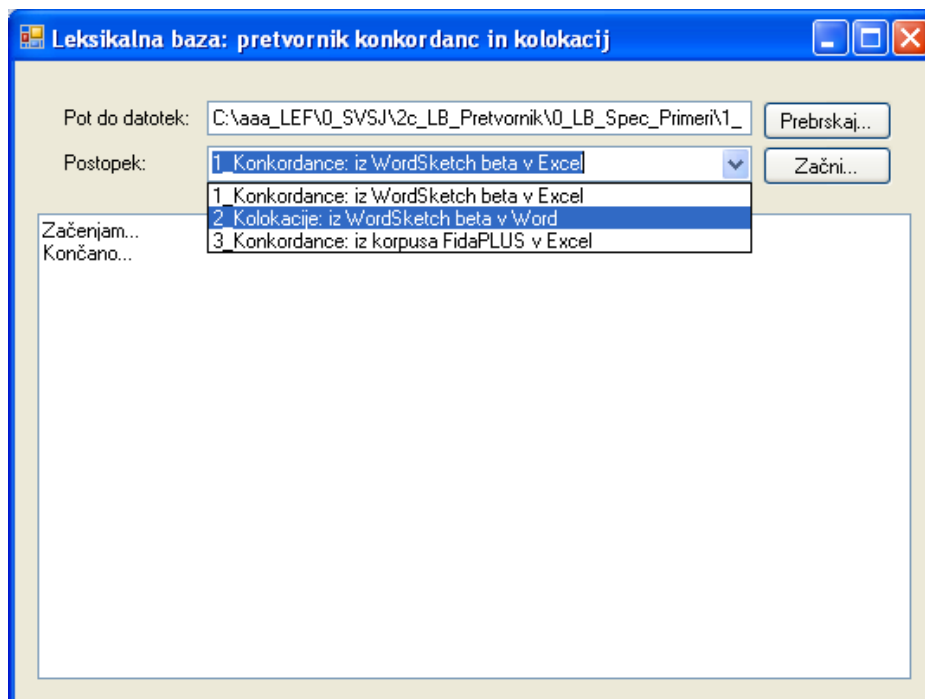
Slika 58: preveritev, ali je .net Framework nameščen



Če program ni nameščen, ga presnamemo z Microsoft Download centra (<http://www.microsoft.com/downloads/en/default.aspx>) in namestimo.

3.3 Opis in uporaba

Programsko okno pretvornika:



Slika 59: programsko okno Pretvornika za konkordance in kolokacije

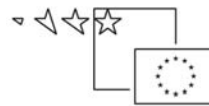
V pretvorniku najprej izberemo ciljno mapo, kjer imamo datoteke, ki jih želimo pretvoriti. Pot do datotek, gumb *Prebrskaj...*

Pretvornik omogoča **tri postopke**, opisane v nadaljevanju.

3.3.1 Prvi postopek: pretvorba konkordanc iz orodja SkE beta, funkcija Konkordance (Concordance)

1_Konkordance: iz WordSketch beta v Excel

a) **Vhodna datoteka** vsebuje konkordance iz orodja SkE beta, ki jih dobimo z iskanjem s funkcijo **Concordance**.

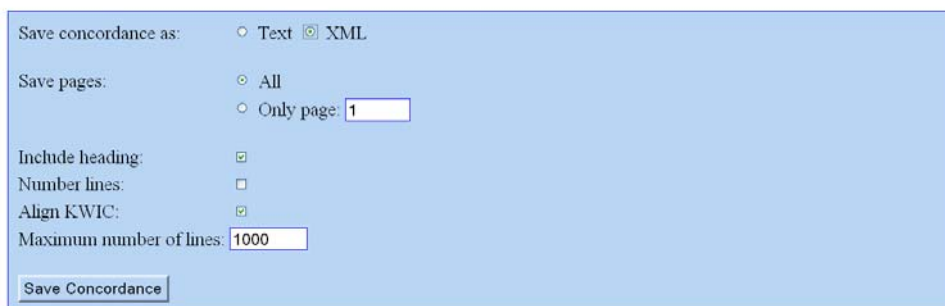


Izdelamo vzorec konkordanc (Sample) z vsebinsko dogovorjenimi nastavitvami in ga posnamemo: **Save**, izberemo **XML**, ostale nastavitve pustimo, kot so, in pritisnemo **Save Concordance**.

Orodje najprej na zaslonu prikaže datoteko, ki jo z brskalnikom shranimo še na disk: **Datoteka, Shrani stran kot**, pri **Vrsta datoteke** izberemo »Spletna stran, le XML«.



Save Concordance



Slika 60: okno za shranjevanje konkordanc v orodju SKE beta

b) V pretvorniku sprožimo prvi postopek, katerega rezultat je tabela v formatu **.xlsx**:

A	B	C	D	E	F	G	H
	----- 2000 : " Kaj naj rečem po takšnem zaključku? Dobro smo	igrati	, pa izgubili . Zavedam se , da sem zapravil dve priložnosti za zmago , in počutim se krivega . Tako				
	Koncept : Iztok Kovač MED 22.00 Slovenska kinoteka Lj . NAS	igrano	- dokumentarni prvenec mlade slovenske igralka Nataše Matjašec . DAN TOREK , 11. 12 .				
3 Dnevnik	vedno . V boje za svetovno žahovsko prvenstvo je že nekajkrat posegel , leta 1995 pa je v New Yorku	igral	celo dvoboj s svetovnim prvakom Kasparovom . V tem dvoboju je bil sicer poražen (z rezultatom 7,5:10,5				
4 Dnevnik	pred tem treniral Muro , s katero je osvojil slovenski pokal , in Primorje , s katerim je uspešno lani	igral	v Evropi . Perica je že šesta trenerska žrtev v letošnji prvi ligi . Pred tem so trenerja zamenjali				
5 Dnevnik	Ultrasi so bili na moji strani . Z njimi sem bil kolega . Na prijateljskih tekmah so kričali , da moram	igrati	. Zakaj vas pa trener ni dajal v ekipo ? Govoril je , da še nisem pripravljen , da moram				
6 Delo	tremi novinci - L. Nagyjem in M. Hadzusom (oba Phoenix Coyotes) in L. Visnavskym (LA Kings) - so	igrati	potrpežljivo in vztrajno ogrožali Sala . V 48. minuti je K. Johnsson skušal preigrati M. Šatana				
	dečkom * sv. Florijana (St. Florianer Sangerknaben) . Brucknerja je še osebno poznal . Orgle je	igral	izredno dobro in vodil tudi cerkveni zbor . Njegova posebna zasluga je , da je poučeval fante in dekleta				
8 Dnevnik	Sheldon Reynolds . - 1.55 Telefon - 3.35 Šakalovo poltje (Death Has A Bad Reputation , VB , 1989) -	igrata	Tony Lo Bianco , Garrick Hagon . Rožija Lawronco Gordon Clark . - 5.10 Airwolf AVSTRIJA 2				
	vrhove Smjak , Vršič , Drčev net , Krmico in v Tamar . Obiskali smo tudi Kekčevo deželno in se	igrati	ob jezeru Jasna . Pestrji so bili večeri s petjem , igrami , nočnim pohodom in programom , ki sta				
	kljub nekaterim oviram uspelo nekoč uresničiti vizijo o festivalu , na katerem bi v Trebnje prišli	igrat	najboljši pihalni orkestri sveta . MIM Daniel Metelko , predsednik Društva Občinski				
11 DELO	pritsnili na plin žele pri rezultatu 62:62 . » Streti bi jih morali v prah in pepel , vendar nismo	igrati	na ravni . Bili smo obupni , « je po tekmi priznal prvi zvezdnik domačega moštva Shaquille O'Neal				
	uvetlavi močna opozicija proti ustanovitvi armade . Ljudje se spominjajo politične vloge , ki jo je	igrala	" obramba rajha " , spominjajo se grozot pretekle vojne , ki so komaj pet , šest let za njimi , spominjajo				
	in metodično trgata papirnati ovoj . " Poročil se je s tukajšnjim dekletom . Nekaj let je	igral	nogomet pri Raidersih . Potem si je uničil koleno , si tako uničil kariero in žena se ga je				
	Izgubil je kvaliteto in mu ni bilo več pomoči . Slab vtis je še povečal s tem , da je dolgo časa	igral	posvem izgubljeno pozicijo . Izidi - 3. kolo : Van Wely : Griščuk remi , Bacrot : Barejev remi				
14 Delo	Kegljačič z našega konca so bili konec minulega tedna uspešni . V prvi ligi so Kočevke doma	igrate	z ekipo Mehano iz Izole 4:4 , v drugi ligi pa so Trebanjke v gosteh premagale Plivko s 6:2 , Krčanke				
	na Koroško in na Štajersko , pa naj bo to na veselice all na		tudi zabavne skladbe , med narodnozabavnimi pa poleg svojih				

Slika 61: rezultat prvega postopka pretvorbe s pretvornikom



V stolpcu A so podatki o viru besedila, v stolpcih B, C in D pa so konkordance. Nadaljnji stolpci so namenjeni vsebinskemu delu redaktorjev.

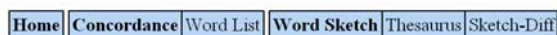
3.3.2 Drugi postopek: pretvorba kolokacij iz orodja SkE beta, funkcija Besedna skica (Word Sketch)

2_Kolokacije: iz WordSketch beta v Word

a) **Vhodna datoteka** vsebuje kolokacijske tabele s slovničnimi relacijami iz orodja SkE beta, ki jih dobimo z izdelavo besedne skice (funkcija **Word Sketch**).

Besedno skico posnamemo: **Save**, izberemo **XML**, ostale nastavitve pustimo, kot so, in pritisnemo **Save Word Sketch**.

Orodje na zaslonu prikaže datoteko, ki jo z brskalnikom shranimo še na disk: **Datoteka, Shrani stran kot**, pri **Vrsta datoteke** izberemo »Spletna stran, le XML«.



Save Word Sketch - pobrati



Slika 62: okno za shranjevanje Besedne skice v orodju SkE beta

b) V pretvorniku sprožimo drugi postopek, katerega rezultat je tabela v **.docx**:

pobrati **29468**

post_izpred	8	22.5	
oči	6	22.41	
has_obj4	8489	14.8	
šilo	218	73.54	
smetana	192	46.89	
žogica	108	43.47	
denar	488	41.03	
štopar	23	37.47	
zavarovalnina	30	37.16	
nagrada	275	37.13	
davek	134	32.12	
pridelek	93	31.93	
smet	53	31.78	
lovorika	48	31.71	
torba	52	31.01	
zlatnina	24	29.26	
žoga	69	28.85	

Slika 63: rezultat drugega postopka pretvorbe s pretvornikom (izsek tabele)

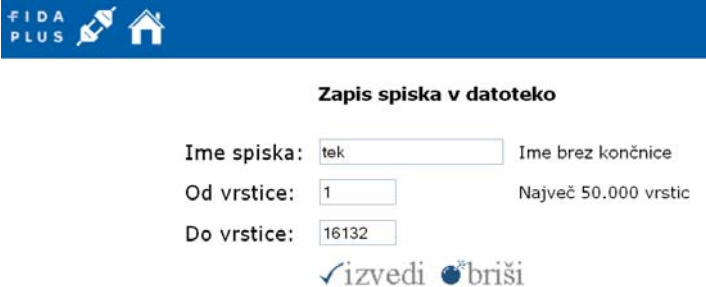




Na vrhu datoteke je navedena lema in njena frekvenca v korpusu FidaPLUS, sledijo pa leksikalni in statistični podatki iz Besedne skice. Dostopne so tudi povezave na izvirne konkordance pri posameznem kolokatorju, desni stolpec pa je namenjen vsebinskemu delu redaktorjev.

3.3.3 Tretji postopek: pretvorba konkordanc iz konkordančnika ASP32 v Excel

3_Konkordance: iz korpusa FidaPLUS v Excel

a) **Vhodna datoteka** je konkordančni niz iz FidePlus, kot ga v tekstovno datoteko posnamemo v konkordančniku ASP32 s funkcijo *Zapis spiska v datoteko*, ikona »**svinčnik**«.





FIDA PLUS  

Zapis spiska v datoteko

Ime spiska: Ime brez končnice

Od vrstice: Največ 50.000 vrstic

Do vrstice:

 izvedi  briši

Slika 64: okno za shranjevanje konkordanc v konkordančniku ASP32

Datoteko nato iz konkordančnika prenesemo na računalnik. Potrebno jo je še posneti kot XML in določiti format UTF-8.

b) V pretvorniku sprožimo tretji postopek, katerega rezultat je tabela v formatu **.xlsx**:



A	B	C	D	E	F	G	H	I	J
1	0035249..... drugo	otrdele mišice in	tek	177					
2	0035249..... drugo		tek	čez zapreke 228					
3	0035249..... drugo		tek	na dolge proge 39					
4	0035249..... drugo	dolgotrajen počasen	tek	45					
5	0035251..... knjižno	spet druga čisti in celi gnojne rane, ena zbuja	tek	in ena blaži kašelj. Uporabni deli rastlin so nekje					
6	0035251..... knjižno	3. Od rastiine, ki zbuja	tek	, se naredi poparek iz pol kavne žličke korenin na					
7	0035251..... knjižno	rane listi, sivka pospešuje prebavo cvetovi, angelika zbuja	tek	korenine, materina dušica blaži kašelj cela rastlina.					
8	0035251..... knjižno	c) srenčica vzbuja	tek	.					
9	LISA..... revialno	lahko sami seba v samo dveh minutah ponovno spravili v	tek	.					
10	LISA..... revialno	(v lekarnah), ki takoj počene cirkulacijo v	tek	.					
11	HEVNIK..... casopisno	organizatorji tekov in s katerim uspešno promovira športno-	rekreativni tek	med prebivalci Slovenije. število na novo vključenih					
12	HEVNIK..... casopisno	da bom najboljšo pripravljenost pokazala prav v Parizu. Visak	tek	bo zgodba zase," pravi Merlene Ottley, ki					
13	GORENJSKI GLAS..... casopisno	Markov	tek	na Javorč Planinsko društvo žiri Jutri, v soboto,					
14	GORENJSKI GLAS..... casopisno	soboto, 1. oktobra, prireja 6. Markov	tek	na Javorč, ki poteka v spomin na znanega alpinista					
15	GORENJSKI GLAS..... casopisno		tek	na Dražgoško goro 5D Dražgoše v nedeljo, 2.					
16	GORENJSKI GLAS..... casopisno	v nedeljo, 2. oktobra, organizira 1.	tek	na Dražgoško goro. Start bo ob 10. uri					
17	GORENJSKI GLAS..... casopisno	18. gorski	tek	na planino Javornik: absolutno: moški: 1.					
18	GORENJSKI GLAS..... casopisno	Adria fest<<, sejmu certificirane biološke ponudbe. Dober	tek	in na zdravje nam je voščil dr. Štefan Mrkač					
19	GORENJSKI GLAS..... casopisno	, kjer se tekmovalci pomerijo v štirih atletskih disciplinah (tek	na 4000 in 60 metrov, skok v daljino,					
20	GORENJSKI GLAS..... casopisno		tek	na 60 m					
21	GORENJSKI GLAS..... casopisno	GORSKI	TEK						
22	GORENJSKI GLAS..... casopisno	in naravo za prihodnje rodove. Globalizacija, to je	tek	za še večjim zaslužkom, bo pustila posledice tudi v					
23	GORENJSKI GLAS..... casopisno		TEK						
24	GORENJSKI GLAS..... casopisno	Markov	tek	na Javorč zmagovalci po kategorijah - mi. deklice:					
25	GORENJSKI GLAS..... casopisno		tek	svetovne harmonije					
26	GORENJSKI GLAS..... casopisno	Kamnik prispela mednarodna skupina tekačev, ki v okviru	tek	svetovne harmonije po Evropi spodbuja strpnost, razumevanje					
27	GORENJSKI GLAS..... casopisno	prirreditve	tek	med različnimi					
28	GORENJSKI GLAS..... casopisno	Mednarodna dobrodelna športna prirreditve	tek	svetovne harmonije se je minuli teden dotaknila tudi					
29	GORENJSKI GLAS..... casopisno		tek	Gorenjske.					
30	GORENJSKI GLAS..... casopisno		tek	svetovne harmonije poteka skozi 70 držav na vseh celinah,					
31	GORENJSKI GLAS..... casopisno		tek	svetovne harmonije je zajel tudi Kamnik.					
32	GORENJSKI GLAS..... casopisno	, kjer deluje slovenski jezuit Stanko Rozman. Dobrodelni	tek	pomeni normalno udeležbo na maratonu, le da naša skupina					
33	GORENJSKI GLAS..... casopisno	12.30 Viharne zgodbe:	tek	za življenje					
34	GORENJSKI GLAS..... casopisno	18.30 Viharne zgodbe:	tek	za življenje					

Slika 65: rezultat tretjega postopka pretvorbe s pretvornikom

V stolpcu A in B so podatki o viru konkordance, v stolpcih C, D in E pa so konkordance. Stolci od F dalje so namenjeni vsebinskemu delu redaktorjev.



4

Dictionary Production System

*Programska
oprema
za
sestavljanje
slovarjev*



4 Programska oprema za sestavljanje slovarjev: *Dictionary Production System*

4.1 Uvod

Za izdelavo leksikalne baze za slovenščino uporabljamo namensko programsko opremo podjetja IDM (<http://www.idm.fr>), t. i. *Dictionary Production System* (dalje: DPS).

V uvodnem delu najprej povzeto opišemo programsko opremo DPS in predstavimo proizvajalca, v nadaljevanju pa predstavimo ključne funkcije programa in izrazje, povezano z njim:

- zasnova programa: podatkovna baza v formatu XML, komunikacija strežnik–računalnik, programska oprema na strežniku in pri uporabniku, delo s spletno povezavo in v brezpovezavnem načinu, uporabniški profili, vrste uporabnikov
- programski vmesnik za leksikografsko delo s podatkovno bazo: delo s paketi in brez; načini za urejanje besedila, preverjanje hierarhične strukture gesla, modul za primerjanje različic gesla
- strežnik: vsebinske nastavitve projekta, določitev poteka dela (delovni postopek, postopki in koraki), razporejanje gradiv na naloge in pakete, status gesel, modul za primerjanje in potrjevanje gesel, zahtevno iskanje, izvozi
- administriranje: vzpostavitev projekta, določitev uporabniških profilov in uporabnikov, konfiguracija projekta, DTD

4.1.1 Uvodno o programski opremi DPS

DPS je program, zasnovan za specifične potrebe sestavljanja slovarjev in podobnih jezikovnih virov, zato so med ciljnim uporabniki slovaropisci, uredniki, projektni vodje, založniške hiše in druge, zlasti raziskovalne ustanove, ki upravljajo z jezikovnimi podatkovnimi bazami. Omogoča delo preko spleta ter tudi brez njega, torej s hkratno povezavo ali pa v brezpovezavnem načinu, pri čemer je možnih več režimov dela, pretežna večina strank pa se odloča za namestitev centralne jezikovne podatkovne baze v formatu XML na proizvajalčevem strežniku.

Program DPS je zasnovan tako, da uredniki dostopajo do podatkovne baze preko spleta in z njo delajo bodisi neposredno na strežniku ali pa tako, da vnaprej pripravljeno gradivo prenesejo s strežnika na svoj računalnik, ga uredijo in vrnejo v bazo, pri čemer za obdelavo uporabljajo programski vmesnik za delo v brezpovezavnem načinu. Program DPS uredništvu omogoča preprost nadzor nad gradivom – tako tehnično kot vsebinsko gledano; omogoča na primer nadzor nad posameznimi vsebinskimi segmenti, določitev faz obdelave,



določitev uporabniških pravic, razdeljevanje gradiva po različnih kriterijih, primerjave stare in nove različice gradiva, preverjanje in preusmerjanje gradiva, poleg tega pa tudi kompleksna iskanja ter obdelave na celotnem slovarskem gradivu v formatu XML. Program DPS omogoča tudi pretvorbo v PDF ter hiter izvoz/uvoz podatkov.

Med ključne prednosti programa se uvrščajo nadzor nad opravljenim delom, varnost podatkov, večuporabniško delo na daljavo, upravljanje s centralno podatkovno bazo in uporabniškimi pravicami. Menimo, da je zaradi teh lastnosti program DPS zelo primeren za sestavljanje novih jezikovnih podatkovnih baz. Predstavitvena stran programa DPS je na povezavi http://www.idm.fr/products/dictionary_writing_system/27/.

4.1.2 Profil proizvajalca DPS: podjetje IDM

IDM je francosko podjetje, ki je bilo ustanovljeno leta 1994. Osnovne storitve podjetja so razvoj programske opreme in drugih rešitev za upravljanje s podatkovnimi bazami, razvoj internetnih ter intranetnih storitev in drugih spletnih aplikacij. IDM se je uveljavil tudi kot strokovni partner na področju standardizirane programske opreme.

Uporabniki storitev IDM so založniki slovarjev, enciklopedij, imenikov, katalogov ter nekatere finančne ustanove. Program DPS trenutno uporablja preko 500 strank v Evropi in Severni Ameriki, med katerimi so tudi nekateri največji britanski založniki slovarjev in drugih jezikovnih virov.

Kontaktne podatke proizvajalca programske opreme:

Ingénierie Diffusion Multimédia
27, rue Albert Einstein
BP 117 – Champs sur Marne
77423 Marne la Vallée Cedex 2
Francija

Spletna stran: <http://www.idm.fr>
E-pošta: info@idm.fr
Telefon: (+33) (0)1 64 61 31 50



4.2 Opis programske opreme DPS

4.2.1 Uvod

Dictionary Production System (DPS) je program, zasnovan za specifične potrebe sestavljanja slovarjev. Oznaka DPS se uporablja za celoten programski paket, ki sestoji iz dveh osnovnih komponent:

DPS strežnik (*DPS Website*)

DPS vmesnik (*Entry Editor*)

Potek dela z DPS je osredinjen na Skladišče (*Repository*), kjer je shranjena vsebina podatkovne baze.

DPS strežnik je spletna aplikacija, do katere uporabniki dostopajo s standardnimi orodji za navigacijo po spletu. Na strežniku so orodja za vodenje projektov, administriranje in konfiguriranje sistema ter uporabniški dostop do podatkov.

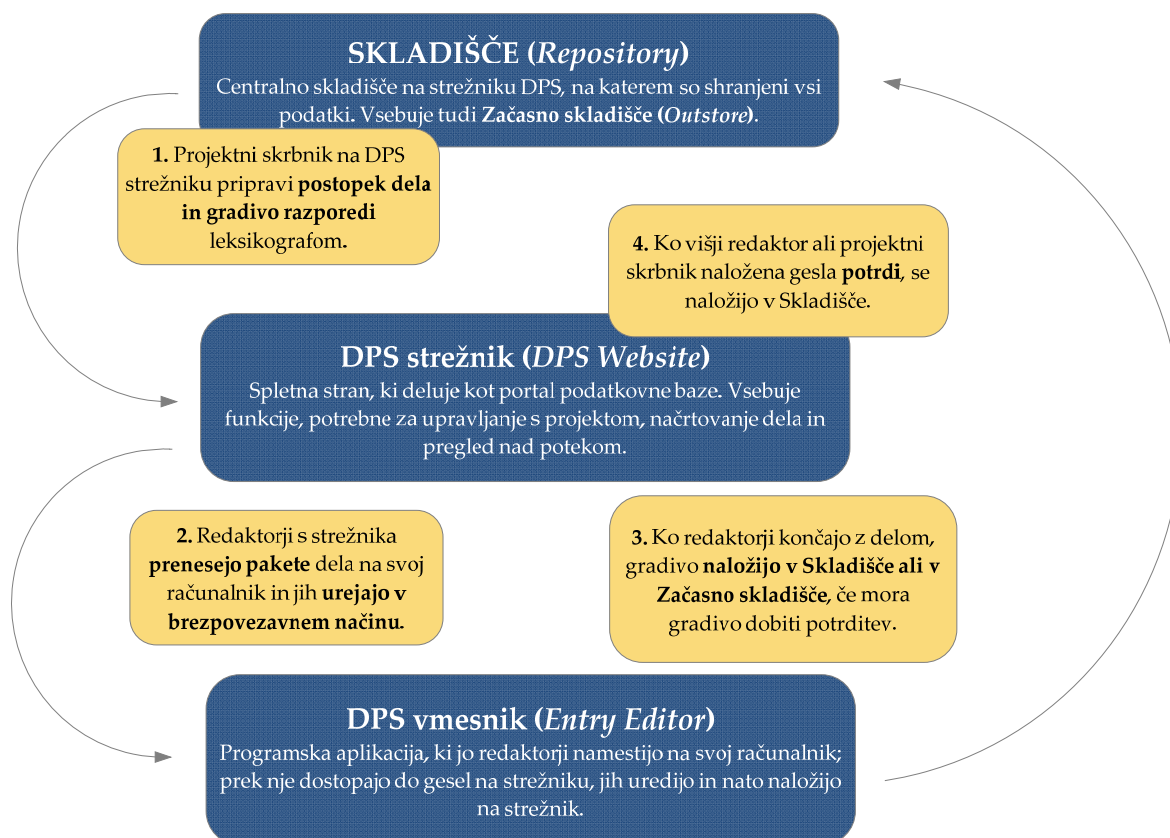
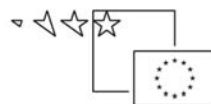
DPS vmesnik je aplikacija za uporabo v okolju Windows, ki jo uporabnik namesti na svoj računalnik. Osnovni postopek dela je, da uporabnik z DPS strežnika na svoj računalnik prenese podatke, jih obdela in vrne v Skladišče. Možno je tako delo neposredno s Skladiščem, ki je bolj prosto organiziran, ali pa s t. i. paketi (*batch*), ki omogočajo večji nadzor nad postopkom dela.

Za vse uporabnike je pomembno, da so seznanjeni s:

- celostnim delovanjem programa,
- vrstami uporabnikov,
- delovnim postopkom (*Workflow*),
- projektno konfiguracijo,
- uporabniškimi pravicami.

4.2.2 Celostno delovanje programa

Spodnji diagram prikazuje povezavo med različnimi komponentami programa DPS:



Slika 66: Program DPS: celostno delovanje programa: komunikacija med posameznimi deli programa in uporabniki

4.2.3 Vrste uporabnikov

1. Redaktorji

Redaktor je krovni izraz za vse vsebinske delavce pri projektu: leksikografe, prevajalce oziroma druge sodelavce, ki urejajo in pregledujejo vsebino; v splošnem delajo s paketi, ki so jim dodeljeni v delo. DPS strežnik uporabljajo za pregled vsebine, pridobitev svojega paketa, manjše popravke, večino dela pa opravijo v DPS vmesniku, kamor naložijo gesla z DPS strežnika, jih uredijo in naložijo nazaj na DPS strežnik.

Pri projektu SSJ je omogočeno delo do 25 redaktorjem.

2. Projektni skrbniki

Projektni skrbniki so odgovorni za vzpostavitev delovnega postopka. DPS strežnik uporabljajo za izdelavo nalog in paketov dela ter jih dodelijo redaktorjem. Nadalje spremljajo časovni in količinski potek dela in potrjujejo opravljeno delo. Projektni skrbniki morajo poznati tudi delo z DPS vmesnikom, da lahko uspešno vodijo delo redaktorjev.

Pri projektu SSJ sta dva projektna skrbnika.



3. Sistemski skrbniki

Sistemski skrbniki so odgovorni za tehnične vidike projektov v programu DPS, npr. za vzpostavitev in konfiguriranje projektov v DPS, urejanje uporabniških pravic ter procesiranje in prečiščevanje podatkov.

Pri projektu SSJ sta dva sistemska skrbnika.

4.2.4 Delovni postopek

Delovni postopek je načrt izdelave slovarskih gesel z organizacijskega vidika. V postopku so natančno določene faze izdelave gesel, t. i. postopki in koraki.

4.2.5 Projektna konfiguracija

DPS je program, ki ga je možno prilagoditi potrebam posameznega projekta. Sestavljen je iz osnovne strukture in komponent, ki jih je možno nastaviti s stališča uporabniških pravic, izgledov, razvrščanja gesel itd. Konfiguracijo izdelajo projektni in sistemski skrbniki, preden se začne vsebinsko delo pri projektu.

4.2.6 Uporabniške pravice

Uporabnikom je možno nastaviti pravice za dostop in/ali urejanje različnih delov projekta. Sistemski skrbniki imajo dostop do vseh področij DPS strežnika, uredniki pa ponavadi samo do področij, ki jim omogočajo iskanje gesel in dostop do paketov dela.



4.3 DPS vmesnik

4.3.1 Namestitev DPS vmesnika, namestitev konfiguracije, zagon DPS vmesnika in dostop do gradiva

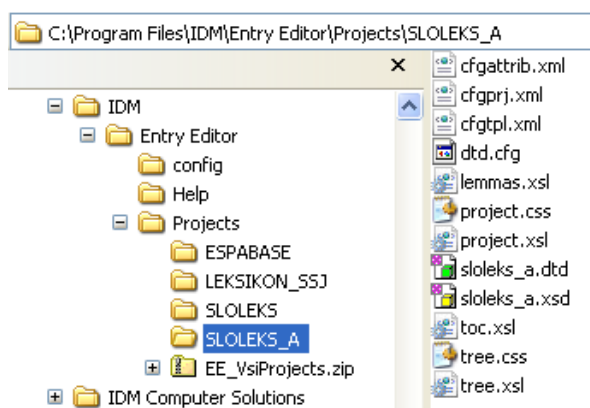
Namestitev DPS vmesnika in konfiguracije

1) Namestitev programske datoteke **.exe** z DPS vmesnikom poteka kot pri običajnih aplikacijah v okolju Windows.

2) Namestitev konfiguracije DPS vmesnika obsega dodatni postopek v mapah, kjer je nameščen DPS vmesnik (**C:\Program Files\IDM\Entry Editor**). V raziskovalni fazi konfiguracija obsega:

a) datoteko **config.xml**, s katero uporabnik zamenja istoimensko izvorno datoteko v mapi **C:\Program Files\IDM\Entry Editor\config**

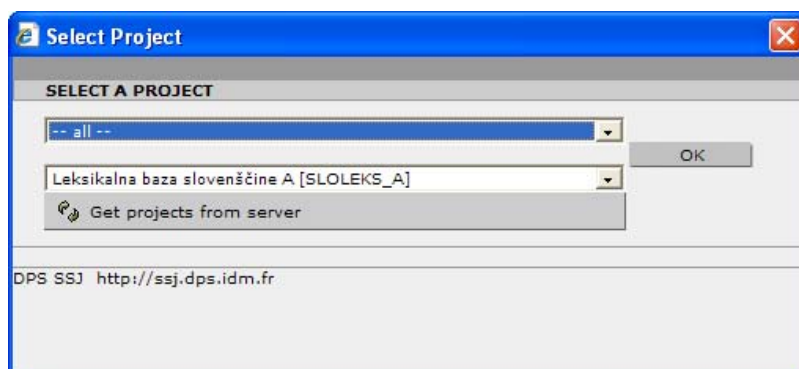
b) mapo **SLOLEKS_A**, kjer je konfiguracija za leksikalno bazo slovenščine; to uporabnik namesti v **C:\Program Files\IDM\Entry Editor\Projects**, kot prikazuje slika:



Slika 67: mesto, kamor uporabnik namesti konfiguracijo za DPS vmesnik

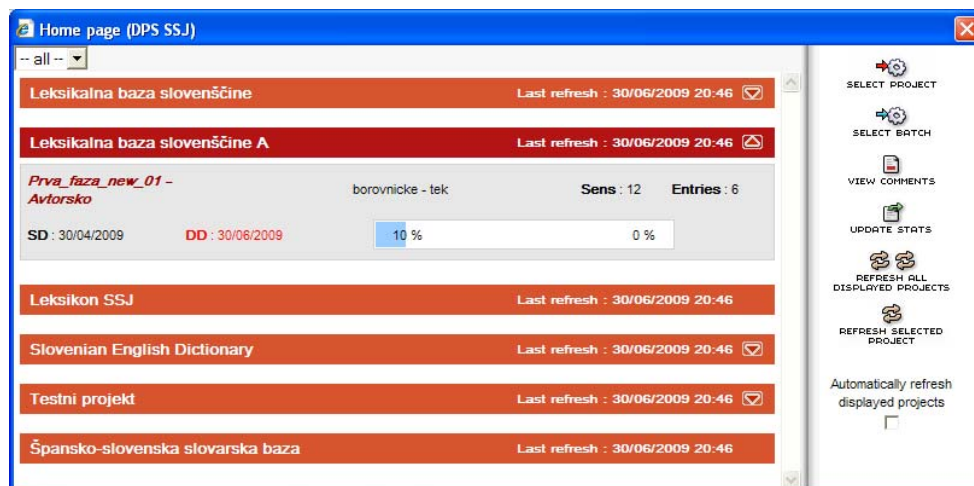
Zagon DPS vmesnika in dostop do gradiva

Ob zagonu vmesnika se pojavi okno, pri katerem je treba najprej klikniti na *Get projects from server*, nadalje vpisati geslo, ki ga uporabnik dobi pri projektnem skrbniku, za trenutek počakati, da do prijave pride, ter nato OK v spodaj prikazanem oknu:



Slika 68: DPS vmesnik: prijava na DPS strežnik

Tako se uporabnik prijavi na DPS strežnik. Dostop do podatkov, ki so uporabniku dodeljena v urejanje, je mogoč preko ikone Domov (*Home*). Uporabnik označi paket, ki mu je dodeljen, in ga prenese na svoj računalnik (dvoklik, ENTER ali *Select batch*):



Slika 69: DPS vmesnik: dostop do paketov na DPS strežniku (gumb Home)

4.3.2 Programsko okno DPS vmesnika

Celotno programsko okno je razdeljeno na štiri dele. Od leve proti desni so to:

1. **okno (*Navigation panel*)**: namenjeno dostopu do gesel v dodeljenem paketu ter za komunikacijo z DPS strežnikom
2. **okno (*Wysiwyg view*)**: namenjeno ogledu gesla brez XML elementov v izgledu za branje. Besedilo v tem oknu je možno tudi urejati. Enak izgled je tudi predviden za natis.
3. **okno (*Tree view*)**: namenjeno prikazu drevesne strukture gesla z XML elementi. To je glavno okno in je namenjeno večinskemu urejanju gesla.
4. **okno (*Attributes and Annotations*)**: namenjeno urejanju atributov pri XML elementih ter vstavljanju redaktorskih opomb k vsebini.



Gornji del 1. okna prikazuje vsebino, ki je naložena na uporabnikov računalnik – je torej v DPS vmesniku –, spodnji del 2. okna pa prikazuje vsebino na DPS strežniku.

V gornjem delu 1. okna uporabnik pred gesli, naloženimi v DPS vmesnik, vidi različne ikone – te ponazarjajo **pravice do urejanja, skladnost z DTD-jem** ter **sinhroniziranost gesel** s stanjem na DPS strežniku:

<i>ikona</i>	<i>opis ikone</i>	<i>pomen ikone</i>
	list z besedilom	Geslo je naloženo na uporabnikov računalnik in pripravljeno za urejanje.
	prečrtan znak DTD	Geslo ni skladno z DTD-jem.
	list s petkrako zvezdico	Geslo je novoustvarjeno in še ni bilo naloženo na DPS strežnik.
	prečrtan svinčnik	Gesla ni mogoče urejati, ker uporabnik zanj nima pravic ali pa je zaklenjeno s strani drugega uporabnika.

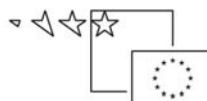
Slika 70: ikone pred iztočnico: pravica do urejanja, skladnost z DTD, sinhroniziranost z DPS strežnikom (1. okno)

Če je pri prvih dveh ikonah v gornji tabeli podlaga ikone **roza barve**, to pomeni, da uporabnik sprememb še ni naložil na DPS strežnik, na kar je potrebno **paziti pri uporabi funkcije ignoriranja sprememb** (*Lose edits*).

2. in 3. okno se sproti sinhronizirata. 1., 2. in 4. okno je možno po želji posamič izklopiti in tako prilagoditi zaslon DPS vmesnika urejanju gesla (ikone).

Gesla na uporabnikovem računalniku so shranjena v Lokalnem skladišču (*Local store*) – za leksikalno bazo pri projektu SSJ je to v C:\Documents and Settings\[uporabnik]\Application Data\DPS Entry Editor\LocalStore\DPS SSJ\SLOLEKS A.

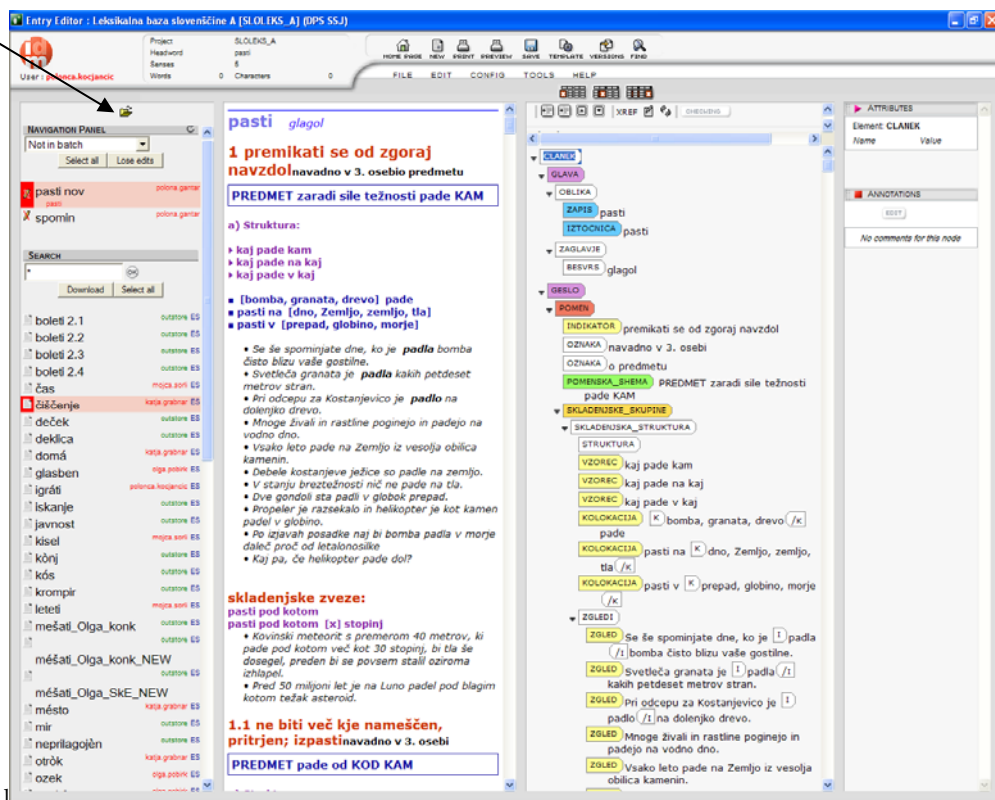
Zgoraj opisane lastnosti povzema naslednja slika:



dostop do
Lokalnega
skladišča

gesla v
uporabni-
kovem DPS
vmesniku

gesla na
oddaljenem
DPS
strežniku



1. okno

2. okno

3. okno

4. okno

Slika 71: DPS vmesnik: deli programskega okna

4.3.3 Gesla v paketih in zunaj njih


Gesla, ki so v paketih in dodeljena konkretnim redaktorjem, so za druge uporabnike zaklenjena. Pri leksikalni bazi za slovenščino je tak režim dela vzpostavljen, zato mora vse redaktorjevo delo potekati znotraj paketa.

Funkcija Zunaj paketa (*Not in batch*) je namenjena za morebiten ogled gesel drugih uporabnikov, pri čemer se na uporabnikov računalnik naložijo v režimu Samo za branje (*Read-only*). Če gesla ne bi bila razporejena v pakete (*Not in batch*), bi bila prosto dostopna iz Skladišča in bi jih bilo možno urejati neposredno po prenosu z DPS strežnika v DPS vmesnik.



4.3.4 Sprotno shranjevanje gesel, vračanje gesel na strežnik in status gesel

Sprotno shranjevanje gesla (*Save*) se nanaša na shranjevanje redaktorjevih sprememb ob

urejanju gesla. Na to, da je v geslu nekaj spremenjeno, opozori znak *Save* z rdečo senco: . Vsebinsko redaktor shrani s klikom na ikono ali s **Ctrl+S**, shrani pa se tudi samodejno, če redaktor odpre kako drugo geslo. **Pomembno:** spremembe se shranijo **na uporabnikov računalnik v Lokalno skladišče** – torej ne na DPS strežnik, zato je treba biti pazljiv pri morebitni uporabi funkcije *Lose edits*.

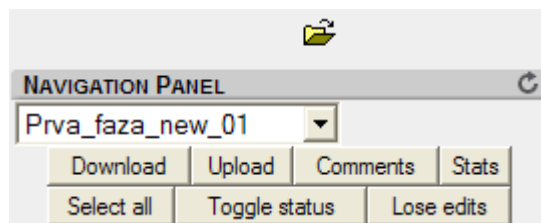
Ignoriranje sprememb (*Lose edits*): če uporabnik ne želi shraniti sprememb, lahko geslo vrne na DPS strežnik nespremenjeno. Če je geslo v paketu, gre v Začasno skladišče, če ne, pa v Skladišče in se hkrati tudi odklene.

Pomembno: s to funkcijo se geslo umakne z uporabnikovega računalnika in izgubljene so vse sprotno shranjene spremembe (Ctrl+S) od zadnjega nalaganja na DPS strežnik.

Ko redaktor prenese gesla z DPS strežnika v DPS vmesnik, ima **tri glavne možnosti upravljanja z gesli**, ki se tičejo komunikacije z DPS strežnikom:

- **nalaganje gesel** v Skladišče na DPS strežniku (*Upload entries*),
- **oddajo gesel** v postopek potrjevanja, skupaj s spremembo statusa gesla (*Submitting entries*),
- **ignoriranje sprememb**, ki omogoča, da gesla vrne v Skladišče nespremenjena (*Lose edits*).

Te in sorodne funkcije prikazuje spodnja slika:



Slika 72: DPS vmesnik, 1. okno, del z gumbi za upravljanje z gesli v paketu

Povzetek funkcij pri delu s paketi:

- 1) Prenesi z DPS strežnika (*Download*): za prenos označenih gesel z DPS strežnika v DPS vmesnik pri uporabniku.
- 2) Naloži na DPS strežnik (*Upload*): za prenos označenih gesel iz DPS vmesnika na DPS strežnik. Gumb je isti tako za sprotno nalaganje na DPS strežnik kot tudi za oddajo gesel.
- 3) Komentarji (*Comments*): prikažejo se morebitni komentarji, ki jih je projektni skrbnik dodal k izbranemu paketu.
- 4) Statistika (*Stats*): prikaže statistični izračun za izbrani paket, pri čemer morajo biti vsa gesla naložena na računalnik in vsaj enkrat predhodno naložena na DPS strežnik; statistični izračun ignorira vsa gesla, ki so bila oddana in sprejeta (*accepted*).
- 5) Izberi vse (*Select all*): za izbor vseh gesel v paketu.
- 6) Menjaj status (*Toggle status*): za menjanje statusa gesla ob oddaji.



7) Ignoriraj spremembe (*Lose edits*): za odstranitev gesla iz DPS vmesnika in Lokalnega skladišča.


Vračanje gesel na DPS strežnik (*Upload*)

Večina projektov je postavljenih tako, da je na konec vsakega koraka uvrščeno potrjevanje gesel, pri katerem ima projektni skrbnik, višji redaktor oziroma druga za to določena oseba možnost, da preveri, ali geslo ustreza merilom prehoda v naslednji korak, preden se prenese v Skladišče.


Nalaganje gesel

Ko uporabnik označi gesla in pritisne gumb *Upload entries*, se gesla s proti shranjenimi spremembami naložijo na DPS strežnik in umaknejo iz DPS vmesnika in Lokalnega skladišča na uporabnikovem računalniku. Če so v paketu, se premaknejo v Začasno skladišče in dobijo oznako *outstore*. Če niso v paketu, se prenesejo neposredno v Skladišče. Pomembno je poznati še razliko med rednim nalaganjem (*Uploading*) in oddajo gesel (*Submitting*).







Redno nalaganje gesel (*Uploading*)

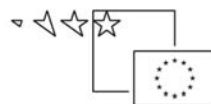
Uporabnik po presoji, a po možnosti redno, nalaga gesla na DPS strežnik, saj s tem zavaruje vsebino. Priporočljivo je denimo nalaganje kakorkoli spremenjenih gesel na DPS strežnik enkrat dnevno ali pa po večjih posegih. Pri tem se pri delu v paketu geslom menja status med belim kvadratom in *outstore* ( *outstore*) in ob nalaganju postane dostopno morebitnemu drugemu redaktorju, ki mu je dodeljen isti paket.

Oddaja gesel (*Submitting*)

Ko uporabnik geslo vsebinsko dokonča, kar pomeni, da mora ustrezati tako vsebinskim merilom, predvidenim za določen korak, kot tudi DTD-ju, geslo lahko odda. To naredi tako, da **geslu pred nalaganjem na DPS strežnik spremeni status**, in sicer z gumbom Menjaj status (*Toggle*) bel kvadratak zamenja za vprašaj na rumeni podlagi (). Gesla se prenesejo z uporabnikovega računalnika v Začasno skladišče in projektne skrbnika opozorijo nase.

Status gesla nadalje urejajo projektni skrbniki oziroma višji redaktorji na DPS strežniku (*Version manager*):

- če **geslo ustreza merilom** prehoda v naslednji korak postopka, ga s kljukico na rumeni podlagi potrdijo (*accepted*), geslo se pri tem umakne iz paketa in postane dostopno v naslednjem koraku (  ),
- če **geslo ne ustreza merilom**, vprašaj zamenjajo spet za bel kvadratak in geslo se brez opozorila vrne nazaj k redaktorju (  ).



Če povzamemo, **status gesel v paketih** (*Entry Status*) kažejo naslednji znaki, ki jih vidimo desno od iztočnice v 1. oknu:

<i>status</i>	<i>opis znaka</i>	<i>pomen statusa</i>
	bel kvadrateg	Geslo je v delu pri redaktorju (<i>blank</i>).
	vprašaj na rumeni podlagi	Geslo je končano in redaktor ga oddaja v potrditev (<i>submitted</i>).
	napis <i>outstore</i>	Geslo je v delu, redaktor pa ga je naložil v Začasno skladišče na DPS strežniku. Gre za začasno oziroma nedokončano različico gesla. Na DPS strežniku je tako geslo označeno s črko »O« na beli podlagi () in je dostopno drugim redaktorjem, ki jim je paket dodeljen (<i>outstore version</i>).
	kljukica na rumeni podlagi	Geslo je končano, sprejeto in naloženo v Skladišče (<i>accepted</i>).

Slika 73: statusi gesla (*Entry Status*)

Zgoraj opisano komunikacijo iz smeri DPS vmesnika proti DPS strežniku ponazarja naslednja slika:

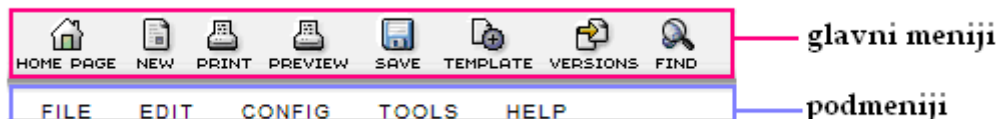


Slika 74: statusi gesla v povezavi s komunikacijo med DPS vmesnikom in DPS strežnikom



4.3.5 Glavni meniji in podmeniji v DPS vmesniku

Na vrhu okna z DPS vmesnikom imamo naslednje ikone:



Slika 75: glavni meniji in podmeniji v DPS vmesniku

Glavni meniji:

Domača stran (Homepage): z nje dostopamo do DPS strežnika, kadar želimo prenesti dodeljene pakete v DPS vmesnik. Tu tudi vidimo, kolikšen odstotek dela je že opravljen.

Novo (New): ikona, s katero ustvarimo novo geslo (v paketu ali zunaj njega). Odpre se okno Izbor šablone (*Template Picker*).

Natisni (Print): odpre se pogovorno okno z nastavitvami za natis trenutno odprtega gesla (**Ctrl+P**). Emulira se bodisi 2. bodisi 3. okno, odvisno, v katerem smo bili, ko smo izbrali ta ukaz.

Predogled (Preview): odpre okno za predogled tiskanja. Izbiramo lahko med naslednjimi možnostmi: natis paketa, natis trenutnega gesla, natis v izgledu 2. okna, natis v izgledu 3. okna, natis enega gesla na stran. Od tod lahko preidemo neposredno na tiskanje.

Shrani (Save): sprotno shranjevanje gesla (**Ctrl+S**). **Pomembno:** gesla se shranijo samo na uporabnikovem računalniku, zato pazljivo pri uporabi funkcije *Lose edits*.

Šablona (Template): odpre se Izbor šablone (*Template Picker*).

Različice (Versions): odpre se modul za primerjanje različic gesla (*Version Manager*).

Podmeniji:

Datoteka (File)

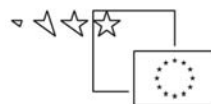
Urejanje (Edit)

Nastavitve (Config)

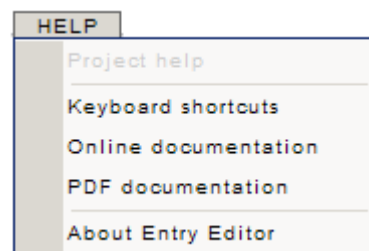
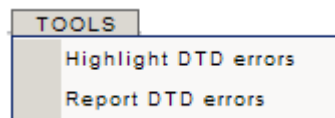
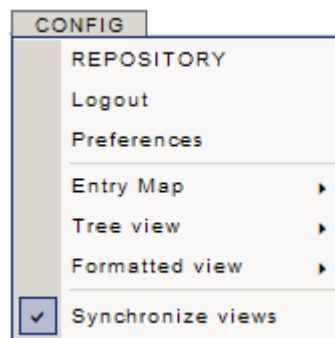
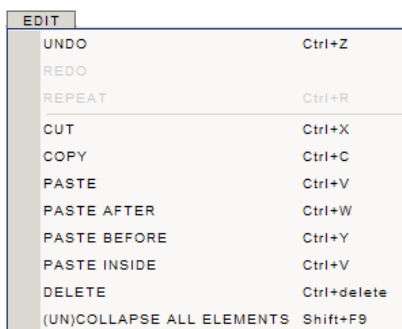
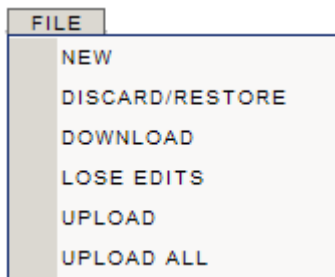
Orodja (Tools)

Pomoč (Help)

Njihove funkcije so opisane v tabeli:



Podmeniji:



Funkcije (z bližnjicami na najpogosteje uporabljene)

Datoteka:

- novo geslo
- zavrzi/obnovi
- prenesi v DPS vmesnik
- ignoriraj spremembe
- naloži na DPS strežnik
- naloži vsa gesla na DPS strežnik

Urejanje:

- razveljavi (Ctrl+Z)
- uveljavi (neaktivno)
- ponovi (neaktivno)
- izreži (Ctrl+X)
- kopiraj (Ctrl+C)
- prilepi (Ctrl+V)
- prilepi zadaj (Ctrl+W)
- prilepi spredaj (Ctrl+Y)
- prilepi element znotraj elementa (Ctrl+V)
- briši element (Ctrl+Delete)
- razširi/skrči element (F9 na trenutnem elementu, Shift+F9 na vseh istoimenskih elementih na isti ravni)

Nastavitve:

- Skladišče
- odjava
- možnosti (število gesel v 2. oknu in velikosti črk)
- izgled v 1. oknu (=Default)
- drevesna struktura (=Standard)
- formatiran izgled (=Full document)
- sinhronizacija 2. in 3. okna

Orodja:

- označi napake DTD-ja v geslu (validacija) (F4)
- razloži napake DTD-ja v označenem elementu (F3)







Pomoč:

- projektna pomoč (neaktivno)
- bližnjice na tipkovnici
- dokumentacija na spletu
- dokumentacija v PDF obliki
- O programu



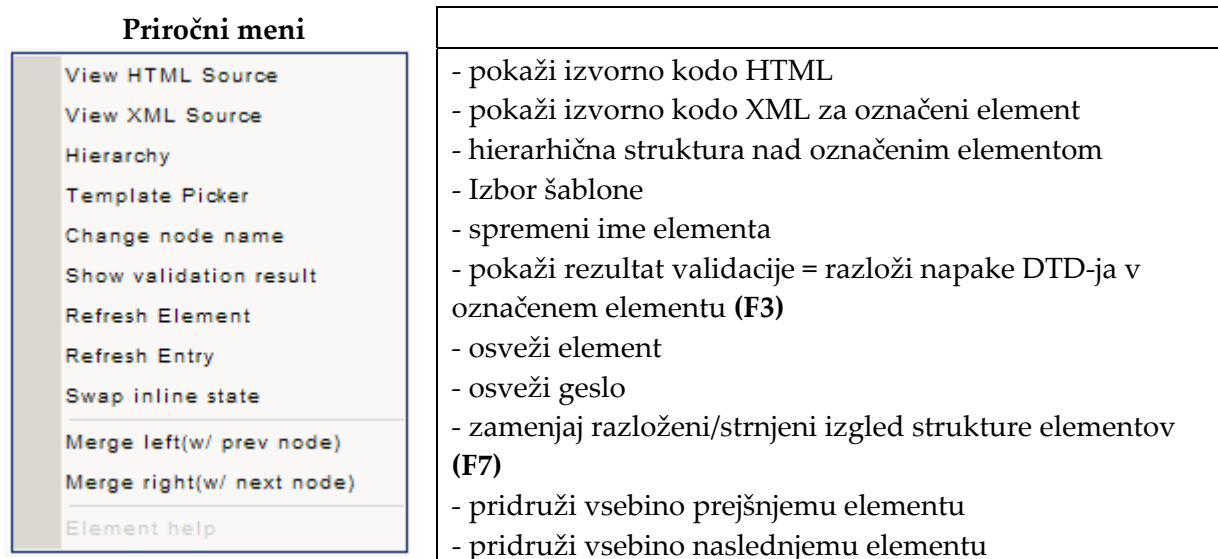
Slika 76: opis podmenijev v DPS vmesniku

4.3.6 Gumbi v drevesni strukturi (3. okno) ter priročni meni

Gumb	Ime in razlaga
	Puščica levo trenutno označeni element prestavi za eno raven više v strukturi (Ctrl+levo).
	Puščica desno trenutno označeni element prestavi za eno raven niže v strukturi (Ctrl+desno).
	Puščica gor trenutno označeni element prestavi pred prejšnji element (Ctrl+gor).
	Puščica dol trenutno označeni element prestavi za naslednji element (Ctrl+dol).
	Uredi lastnosti kazalk (<i>Edit Cross-reference properties</i>).
	Osveži kazalke (<i>Refresh Cross-reference</i>).

Slika 77: DPS vmesnik: gumbi v drevesni strukturi (3. okno)

Če na element v drevesni strukturi **kliknemo z desno miškino tipko** ali pa pritisnemo bližnjico **Shift+F10**, dobimo **priročni meni** z naslednjimi funkcijami:



Slika 78: DPS vmesnik: gumbi v priročnem meniju (desni klik)

4.3.7 Urejanje gesla

Ko imamo gesla naložena v DPS vmesnik, se lahko urejanje začne. V 2. oknu so gesla, ki so naložena na računalnik – v paketu ali zunaj njega. Privzeta nastavitve za prikaz je 8 gesel, lahko pa jo tudi spremenimo. Če gesla ne vidimo, ga izberemo v 1. oknu – kliknemo na zapis iztočnice, ne na ikono pred njim, nakar se odpre v 2. in 3. oknu in je pripravljeno za urejanje. Do večine funkcij za urejanje lahko dostopamo tudi z bližnjicami na tipkovnici, nekatere pa so v priročnem meniju (desni klik z miško).

Gesla lahko urejamo bodisi v 2. ali v 3. oknu, pri čemer se vsebina sprti sinhronizira. **2. okno** je priročno za urejanje vsebine znotraj meja elementa, ker se strukturni elementi izražajo samo preko izgleda. V 2. oknu sicer lahko elemente vstavljamo ali brišemo, ne moremo pa jih premikati ali preimenovati. Za urejanje je primarno **3. okno**, ki prikazuje drevesno strukturo z XML elementi za trenutno aktivno geslo – to je tisto, v katerem je v 2. oknu kurzor. V 3. oknu urejamo tako besedilo gesla kot tudi strukturne XML elemente.

Za delo v DPS vmesniku je pomembno poznavanje načinov urejanja, saj se zaradi kombinacije besedila in XML elementov navigacija deloma razlikuje od standardnih urejevalnikov besedil.

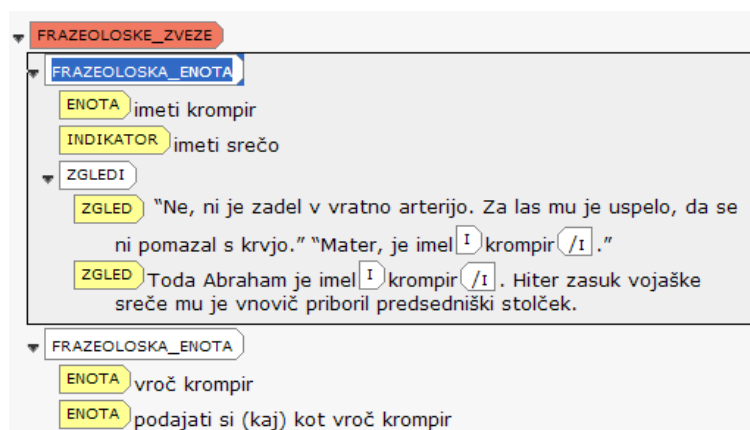
4.3.7.1 Osnovna navigacija po 3. oknu (*Editing an Entry*)

V 3. oknu imamo dva specifična režima:

- režim pomikanja po besedilu
- režim urejanja besedila



V **režimu pomikanja** smo, če kliknemo na ime elementa; okoli njega se naredi okvirček. Po elementih se pomikamo s kurzorjem gor in dol, navzdol tudi s tabulatorjem. Režim pomikanja prikazuje naslednja slika:



Slika 79: režim pomikanja po geslu v DPS vmesniku

Režim pomikanja je ključen za **urejanje XML elementov**.

Elemente pomikamo navzgor ali navzdol v strukturi bodisi z miško, z gumbi v drevesni strukturi ali pa s kombinacijo tipk **Ctrl+gor** ali **Ctrl+dol**.

Elemente pomikamo navzven ali navznoter, torej jim spreminjamo hierarhični nivo, z miško, z gumbi v drevesni strukturi ali pa s kombinacijo tipk **Ctrl+desno** (nivo nižje) ali **Ctrl+levo** (nivo višje).

Celoten element z vsebino vred brišemo s **Ctrl+Delete** (deluje kot *Backspace*). Samo element brez vsebine pa brišemo z **Alt+Delete**.

V **režim urejanja** preidemo s kurzorjem desno in začnemo urejati besedilo; kurzor skoči v prvi element, kjer je besedilo dovoljeno.

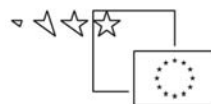
Ko smo v režimu urejanja, za prehajanje v druge elemente ne delujeta kurzorja gor in dol, temveč se vanje lahko premaknemo na spodaj opisane načine.

Najhitreje to storimo:

- s **klikom v besedilo drugega elementa** ali
- s **tabulatorjem** (režim pomikanja), nato **gor** ali **dol**, nato pa kurzor desno (režim urejanja); lahko pa tudi:
- s **klikom na element** (režim pomikanja) in kurzor desno, da gremo spet v režim urejanja,
- s kurzorjem **desno** ali s **Ctrl+desno** se pomaknemo **desno** – v naslednji element (in z **levo** oziroma **Ctrl+levo** v prejšnji element) (koristno, če smo blizu začetka ali konca elementa).

Označevanje

Če hočemo označiti celotno besedilo elementa, kjer ni mešanja elementov in besedila oz. *mixed contenta*, to naredimo tako, da v režimu urejanja najprej pritisnemo tipko **Home** (skočimo na začetek besedila), nato pa **Shift+End**. Celoten element z besedilom vred pa



označimo bodisi z miško ali pa tako, da iz režima urejanja s tabulatorjem preidemo spet v režim pomikanja po besedilu.

Več elementov hkrati lahko označimo v režimu pomikanja:

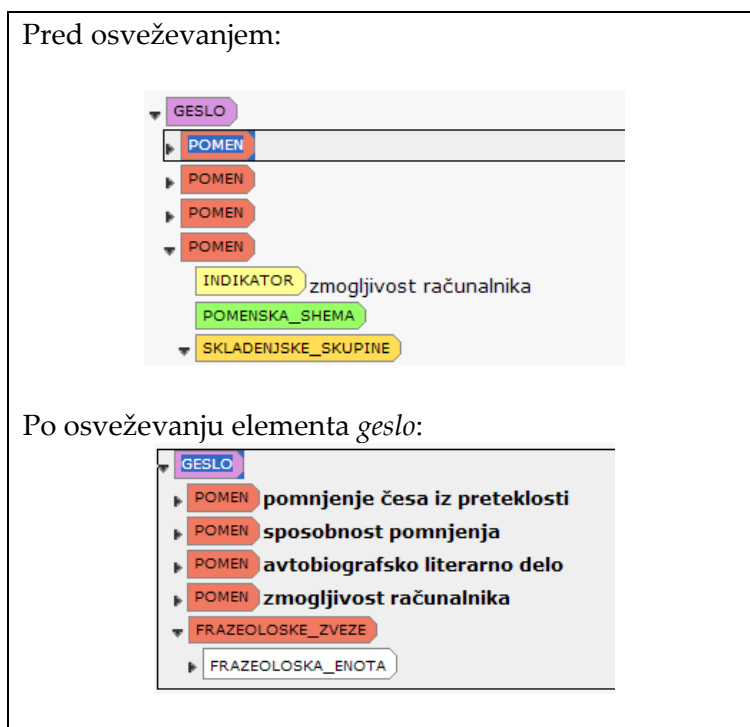
- držimo tipko **Shift** in z miško označimo zaporedne elemente (od–do, kliknemo na prvega in zadnjega),
- držimo tipko **Ctrl** in z miško označimo posamezne elemente (npr. 1., 3. in 5.).

Razširjenje/krčenje elementov

V režimu pomikanja tudi **razširimo ali skrčimo elemente** ((Un)collapse all elements), in sicer z miško tako, da pritisnemo na **trikotnik** pred elementom, ki ima podrejene elemente, oziroma z bližnjicama **F9** in **Shift+F9**: **F9** razširi ali skrči trenutni element, **Shift+F9** pa vse istoimenske elemente na isti ravni, npr. na vseh *pomenih* ali na vseh *idiomatskih enotah*.

Osveževanje elementa

Funkcijo uporabljamo za osveževanje vsebine elementa, pri leksikalni bazi za slovenščino pa jo uporabljamo tudi za **prikaz pomenske strukture gesla**, kar izboljša pregled nad geslom. To naredimo tako, da v režimu pomikanja najprej skrčimo vse *pomene* (Shift+F9), nato pa označimo element *geslo* ter z desno miško na priročnem meniju izberemo *Refresh element*:



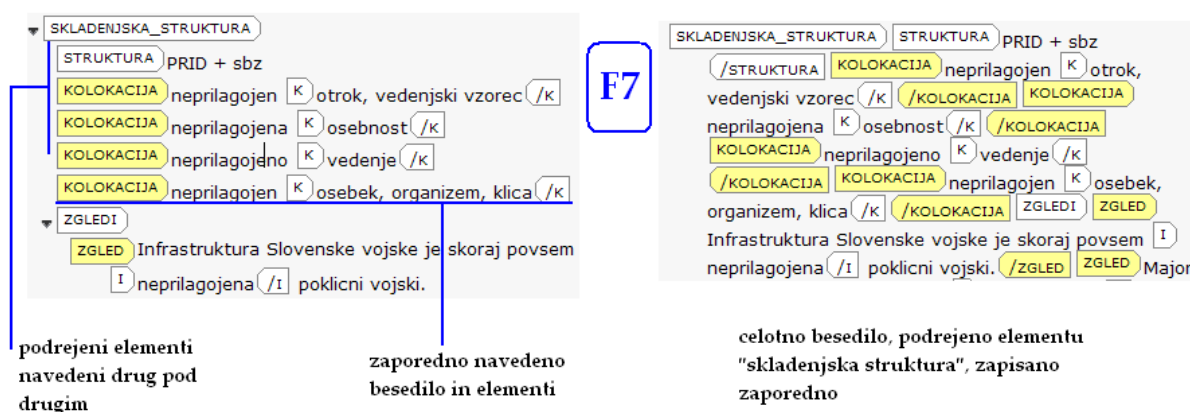
Slika 80: prikaz funkcije Osveži element (Refresh element)

Menjanje razloženega/strnjene izgleda strukture elementov (*Swap inline state*)

Elemente, pri katerih so na podrejenem mestu drugi elementi (ne pa besedilo), v drevesni strukturi 3. okna izhodiščno vidimo drugega pod drugim, vsakega v novi vrstici, podrejeni elementi pa so dodatno umaknjeni v desno.

Elementi, v katerih so dovoljeni tako elementi kot besedilo, pa so izhodiščno prikazani zaporedno oziroma v isti vrstici.

Funkcija na elemente, ki imajo podrejene druge elemente, ne pa tudi besedilo, deluje tako, da vse elemente prikaže **zaporedno oziroma tekoče**:



Slika 81: razloženi in strjnjeni izgled strukture elementov (*Swap inline state*)

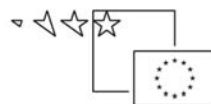
Funkcija je dostopna po dveh poteh:

- iz priročnega menija, do katerega dostopamo z desnim klikom na elementu, ki mu želimo menjati izgled,
- s tipko F7 v režimu pomikanja po besedilu, na elementu, ki mu želimo menjati izgled.

Kadar vnašamo nov element, ki ima podrejeno besedilo in elemente, npr. kolokacijo, pa funkcijo ob prehodu v režim urejanja uporabimo nujno, da lahko začnemo vnašati besedilo:



Slika 82: menjanje v strjnjeni izgled ob vnosu elementa, ki lahko vsebuje besedilo in podrejene elemente (*Swap inline state*)



4.3.7.2 Izbor šablone (*Template Picker*)

Vnos strukturnih elementov ali šablon, ki jih razumemo kot skupine elementov, v DPS vmesniku poteka s funkcijo Izbor šablone (*Template Picker*).

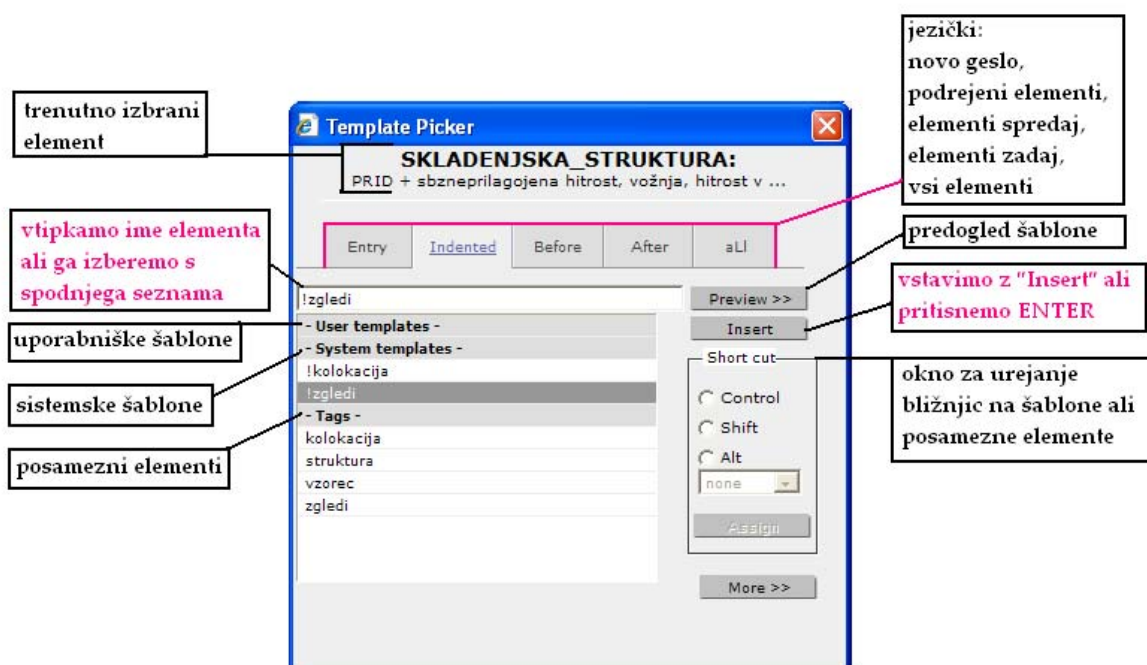
Do okna Izbor šablone dostopamo na tri možne načine:

- 1) kliknemo ikono *Template* v glavnem meniju,
- 2) s kurzorjem gremo na mesto, kjer želimo vstaviti element ali šablono, z desnim klikom na miški odpremo priročni meni in izberemo *Template Picker*,
- 3) na izbranem mestu ga pokličemo z eno od bližnjic. Glede na izbrano bližnjico se bo odprl eden od jezičkov v oknu Izbor šablone:

- | | |
|---------------|--|
| Ctrl+E | 1. jeziček: za vnos novega gesla |
| Ctrl+I | 2. jeziček: za vnos elementov, ki so podrejeni trenutno izbranemu |
| Ctrl+B | 3. jeziček: za vnos elementov, ki so pred trenutno izbranim |
| Ctrl+A | 4. jeziček: za vnos elementov, ki so za trenutno izbranim |
| Ctrl+L | 5. jeziček: za vnos kateregakoli elementa |

Tipka **Escape** okno Izbor šablone spet zapre.

Ob izbiri funkcije se odpre naslednje okno:



Slika 83: okno Izbor šablone z opisom funkcij



Osnovna funkcija okna je, da elemente, ki so na določenem mestu dovoljeni, **izberemo in vnesemo**. Izberemo jih bodisi z miško ali pa tako, da jih vtipkamo v aktivno okence za vnos. Po seznamu se lahko tudi pomikamo s kurzorjem in izbrani element vstavimo s klikom na *Insert* ali pa potrdimo s Potrdi (ENTER).

V seznamu za vnos so na voljo:

- **uporabniške šablone** (*User templates*): te uporabnik po želji ustvari sam,
- **sistemske šablone** (*System templates*): te določijo sistemski skrbniki in jih uporabnik prejme skupaj s konfiguracijo DPS vmesnika,
- **posamezni elementi** (*Tags*): seznam vseh elementov, ki so na določeni hierarhični ravni dopuščeni.

Sistemske šablone pospešijo vnos elementov, ki so po DTD-ju nujni in se na določenem mestu zagotovo pojavijo. Uporabnik šablone lahko pred vnosom lahko pogleda s klikom na gumb *Preview*.

Uporabniške šablone določi uporabnik sam, kadar želi pospešiti vnos skupin elementov, ki jih sistemske šablone ne krijejo.

Ustvarimo jih tako, da:

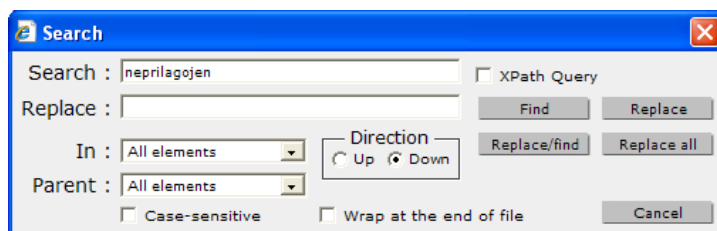
- vnesemo ali izberemo želeni element ali skupino elementov, označimo,
- pritisnemo eno od možnosti:
 - **Ctrl+T**, če želimo ustvariti šablono, ki vsebuje izbrani **element in podelemente brez besedila**,
 - **Ctrl+Shift+T**, če želimo ustvariti šablono, ki vsebuje izbrani **element in podelemente skupaj z besedilom**.
- prikaže se pojavno okno, v katerem vpišemo poljubno ime in novo šablono shranimo z OK.

Določanje bližnjic

Šablonam in elementom lahko določimo bližnjice na tipkovnici, in sicer v oknu Izbor šablone (*Template Picker*) v razdelku *Short cut*.

4.3.7.3 Išči in zamenjaj (*Search and Replace*)

Iskalno okno pokličemo s **Ctrl+F**:



Slika 84: okno Išči/zamenjaj v DPS vmesniku





Išči deluje **samo znotraj trenutno aktivnega gesla** (torej tem, ki je v 3. oknu) za pozicijo, kjer se nahaja kurzor. Naslednjo pojavitev iščemo s klikom na *Find*. Poleg standardnih nastavitev (gor/dol, upoštevanje velike/male začetnice, nadaljevanje na začetku) lahko določimo še, znotraj katerega elementa naj se iskanje izvede (*In*) ter kateri element mu je nadrejen (*Parent*). Nastavitve se shranijo.

Zamenjaj (*Replace*): vnesemo parametre in za prvo iskanje kliknemo *Find*. Nato pa bodisi pritisnemo *Replace/Find* (zamenja in poišče naslednjega), *Replace* (zamenja in ne gre naprej), *Find* (ne zamenja in poišče naslednjega) ali *Replace all* (zamenja vse).

Okno zapremo standardno. Drugače kot pri večini pojavnih oken pa je to, da **okno ne blokira možnosti urejanja** v 2. in 3. oknu, če pustimo odprtega in ga imamo neke na zaslону.

4.3.7.4 Validacija (*Highlight DTD errors, Report DTD errors*)

Z validacijo preverjamo skladnost logične strukture gesla z določili v DTD-ju. Na pravilnost gesla opozarja ikona pred njim, ko je naloženo na uporabnikovem računalniku:

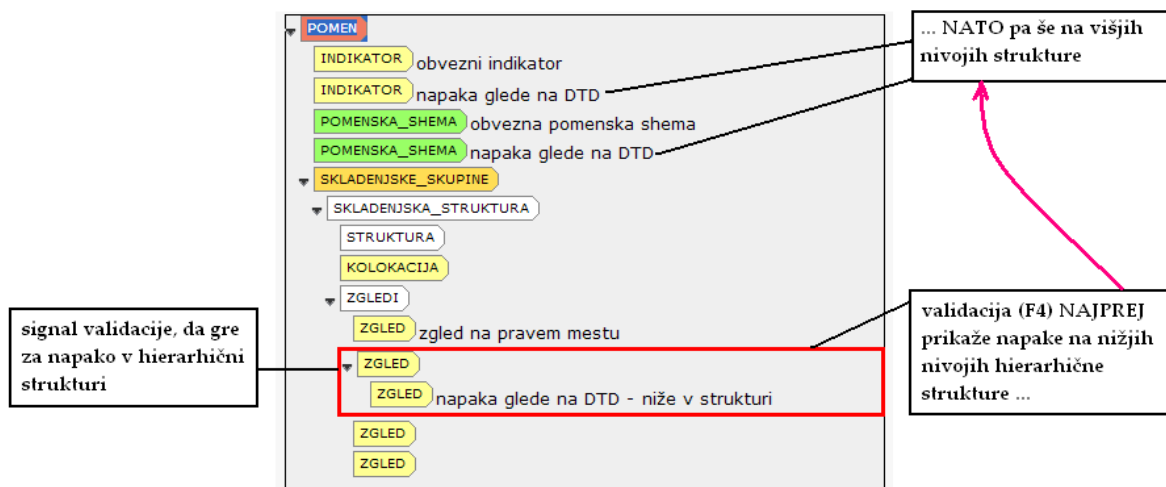
- če je geslo skladno z DTD-jem:  (list z besedilom)
- če geslo ni skladno z DTD-jem:  (prečrtan znak DTD)

Validacijo vključimo/izključimo na naslednje načine:

- v podmeniju *Tools/Highlight DTD errors*,
- s tipko **F4**.

Validacija ne kaže vseh napak po vrsti, temveč deluje po načelu **od spodaj navzgor**: če so napake globlje v strukturi, napak na višjem nivoju še ne prikaže. Prikaže jih takrat, ko so napake na nižjih hierarhičnih nivojih odpravljene. Zato je priporočljivo, da sproti preverjamo manjše izdelane strukturne dele gesla. Grafično je napaka prikazana **z rdečim okvirjem**.

Delovanje validacije je ponazorjeno na naslednji sliki:



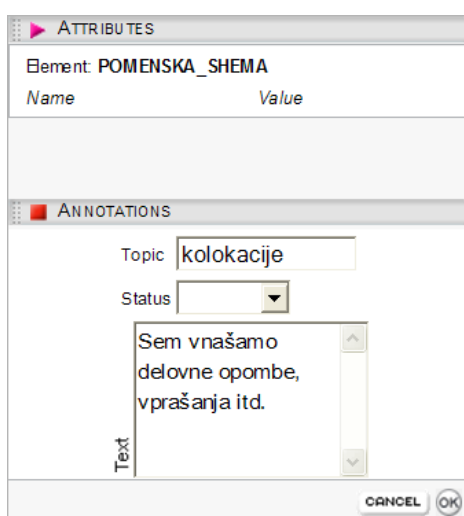
Slika 85: način delovanja validacije v DPS vmesniku

Pri tolmačenju napake si lahko pomagamo tudi z razlago napake DTD-ja v **označenem elementu**, ki je uporabna predvsem **za delo na manjših segmentih strukture**; do nje dostopamo:

- v podmeniju *Tools/Report DTD errors*,
- s tipko **F3** na označenem elementu,
- s klikom na desni miškin gumb, v priročnem meniju *Report DTD errors*.

4.3.7.5 Opombe in atributi (*Annotations, Attributes*)

Z opombami in atributi delamo v 4. oknu DPS vmesnika:



Slika 86: urejanje atributov in vnos opomb v 4. oknu DPS vmesnika



Atribute v DPS vmesniku razumemo kot tiste dele XML elementov, ki s svojimi vrednostmi spreminjajo funkcije elementa (gl. poglavje 5.2 kazalnika 5). V DTD-ju za leksikografsko delo pri leksikalni bazi za slovenščino niso določeni. Pri elementih, ki imajo atribut vpisan, se v 3. oknu izriše **rdeč trikotnik**.

Opombe lahko dodajamo kjerkoli v besedilu, lahko jih je več, lahko jih tudi brišemo. Na mestu, kjer smo vstavili opombo, se v 3. oknu izriše **rdeč kvadrček**. Vsebinsko služijo internim zapiskom, vprašanjem, dodatnim podatkom, ki jih želimo shraniti, niso pa del gesla.

4.3.7.6 Modul za primerjanje različic gesla (*Version manager*)

Modul je namenjen temu, da dobimo **pregled nad vsemi različicami gesla, ki so bile naložene na DPS strežnik**, in služi predvsem arhivskim namenom. Zaradi svoje narave ni primarnega pomena za urejanje gesla, koristen pa je, kadar želimo dobiti vpogled v preteklo dogajanje pri posameznem geslu.



Do modula dostopamo z gumbom *Versions* v glavnem meniju:

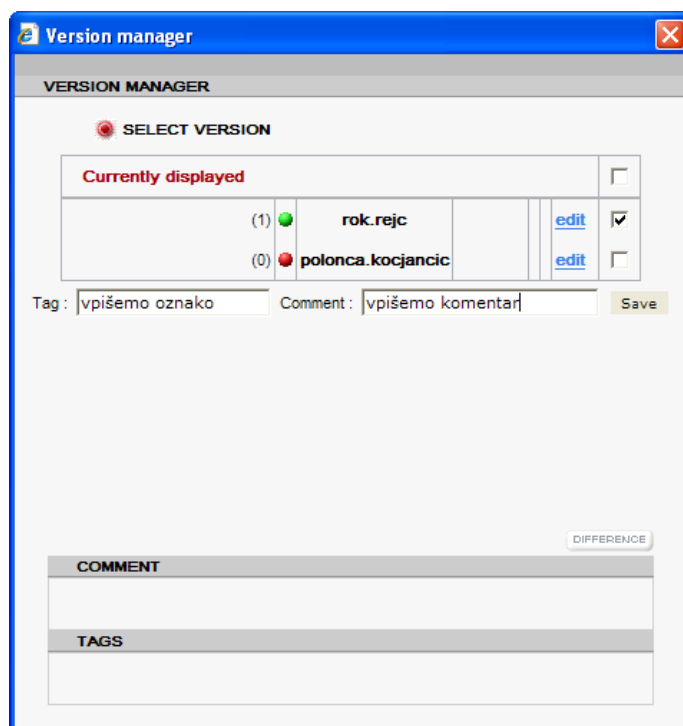
Podatki, ki jih izhodiščno o geslih dobimo, so:

- koliko različic gesla obstaja,
- kdo jih je naložil na DPS strežnik,
- ali so skladne z DTD-jem ali ne.

Modul omogoča:

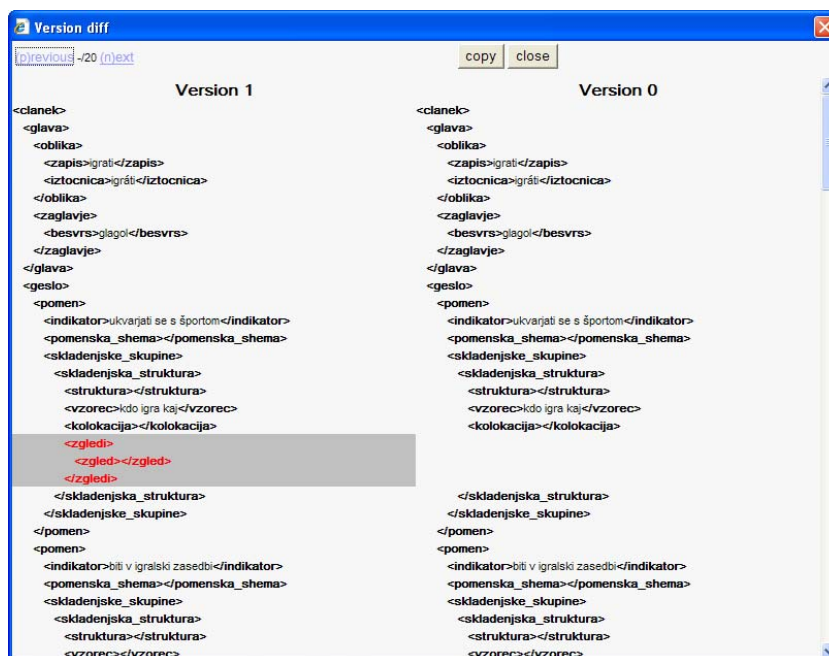
- **označevanje posamezne različice gesla**: označimo jo s kljukico ter vpišemo poljubno oznako (*Tag*) in/ali komentar (*Comment*),
- **pregled in urejanje različic gesla v DPS vmesniku**: kliknemo gumb *edit*,
- **primerjanje dveh različic gesla**: označimo dve različici in kliknemo gumb *Difference*.

Osnovno okno modula za primerjanje različic gesla:



Slika 87: osnovno okno modula za primerjanje različic gesla (Version manager) v DPS vmesniku

Osnovno okno za primerjavo različic gesla (Version diff):



Slika 88: primerjava dveh različic gesla igrati: funkcija Diffence v modulu za primerjanje različic gesla (Version manager)



Ko arhivsko različico odpremo v DPS vmesniku, je v 1. oknu označena s svojo zaporedno številko:



Slika 89: oznaka arhivske različice gesla, ko jo odpremo v DPS vmesniku

Za natančnejše podatke o spreminjanju statusa pri urejanju različic gesla napotujemo na izvorna navodila (poglavje *The Version Manager*).



4.4 DPS strežnik

Dostop do DPS strežnika za potrebe leksikalne baze za slovenščino je omogočen na naslovu <http://ssj.dps.idm.fr/web/Signon.html>. Gesla dodeljujejo projektni in sistemski skrbniki.

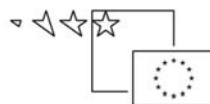
Slika 90: DPS strežnik: okno za vnos podatkov za dostop

V nadaljevanju poglavja so predstavljeni posamezni deli DPS strežnika:



Slika 91: DPS strežnik: razdelki za upravljanje s posameznimi podatkovnimi bazami (projekti) v programu DPS

- 1) Projekti (*Projects*)
- 2) Profili (*Profiles*)
- 3) Uporabniki (*Users*)
- 4) Konfiguracija (*Configuration*)
- 5) Iskanje (*Search*)
- 6) Delovni postopek (*Workflow*)
- 7) Tehnični postopki (*Processes*)
- 8) Zahtevno iskanje (*SkXml Box*)
- 9) Statistika (*Statistics*)



4.4.1 Projekti (Projects)

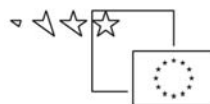
Razdelek je namenjen sistemskim nastavitvam posamezne podatkovne baze oziroma projekta v programu DPS; forma za vnos podatkov je:

Project Name	<input type="text"/>
Project Code	<input type="text"/>
Database template	<input type="text" value="▼"/>
JDBC url	<input type="text"/>
Hibernate dialect	<input type="text"/>
JDBC driver	<input type="text"/>
JDBC username	<input type="text"/>
JDBC password	<input type="text"/>
Locked	<input type="checkbox"/>
Create database	<input type="checkbox"/>

Slika 92: DPS strežnik: Projekti

4.4.2 Profili (Profiles)

Razdelek za urejanje uporabniških profilov omogoča nastavljanje pravic posameznim uporabnikom za vsak posamezni projekt. Za leksikalno bazo za slovenščino so nastavitve redaktorskega profila (*FLBrainiac*) naslednje:



<ul style="list-style-type: none"><input type="checkbox"/> Assign permissions<input checked="" type="checkbox"/> Browse the dictionary<input checked="" type="checkbox"/> Create a new entry<input checked="" type="checkbox"/> Attach files<input checked="" type="checkbox"/> Create comments<input checked="" type="checkbox"/> Add entry to a group<input type="checkbox"/> Manage groups<input checked="" type="checkbox"/> Tagging versions<input type="checkbox"/> Discard entry <input checked="" type="checkbox"/> Download entry<input type="checkbox"/> Upload entry to the database<input checked="" type="checkbox"/> Upload entry to project outstore<input checked="" type="checkbox"/> Force unlock <input type="checkbox"/> Export project<input type="checkbox"/> Import project<input type="checkbox"/> Execute scripts <input type="checkbox"/> Access to SM<input type="checkbox"/> Upload MS Project file<input type="checkbox"/> Allocate people<input checked="" type="checkbox"/> Put sign-off<input checked="" type="checkbox"/> Put final sign off<input type="checkbox"/> Upload config<input checked="" type="checkbox"/> View all batches<input checked="" type="checkbox"/> View all groups<input checked="" type="checkbox"/> View update statistics	<p>Uporabnik z nastavljenim profilom <i>FLBrainiac</i> torej lahko:</p> <ul style="list-style-type: none">- pregleduje podatkovno bazo,- ustvari novo geslo,- pripenja datoteke,- piše komentarje,- dodaja gesla v skupine,- označuje verzije gesel, - prenese geslo v DPS vmesnik,- naloži geslo v Začasno skladišče,- odklene geslo, - geslo odda (<i>sign-off</i>),- geslo potrdi (<i>final sign-off</i>),- gleda vse pakete,- gleda vse skupine,- gleda statistiko.
--	--

Slika 93: pregled uporabniških pravic pri profilu *FLBrainiac*

4.4.3 Uporabniki (Users)

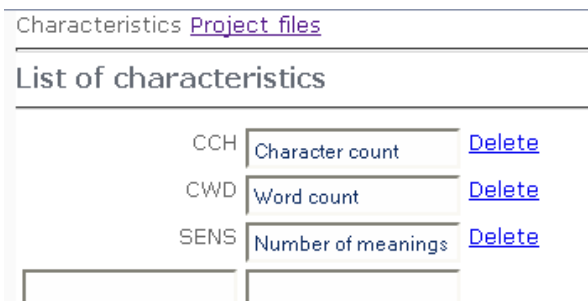
Razdelek za vnos podatkov o uporabnikih in določitev njihovih profilov v posameznem projektu, pri katerem sodelujejo:

Username	sporazumevanje	Projects	Leksikalna baza slovenščine A [SLOLEKS_A]
First name	<input type="text"/>	Profiles	FLBrainiac
Last name	<input type="text"/>		Apply
Email	<input type="text"/>	<input type="checkbox"/> Assign permissions	
Administrator	<input type="checkbox"/>	<input checked="" type="checkbox"/> Browse the dictionary	
Active	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> Create a new entry	
Change password	<input type="text"/>	<input checked="" type="checkbox"/> Attach files	
Confirm password	<input type="text"/>	<input checked="" type="checkbox"/> Create comments	
		<input checked="" type="checkbox"/> Add entry to a group	
		<input type="checkbox"/> Manage groups	
		<input checked="" type="checkbox"/> Tagging versions	
		<input type="checkbox"/> Discard entry	

Slika 94: DPS strežnik: podatki o uporabniku in uporabniških pravicah pri posameznem projektu

4.4.4 Konfiguracija (Configuration)

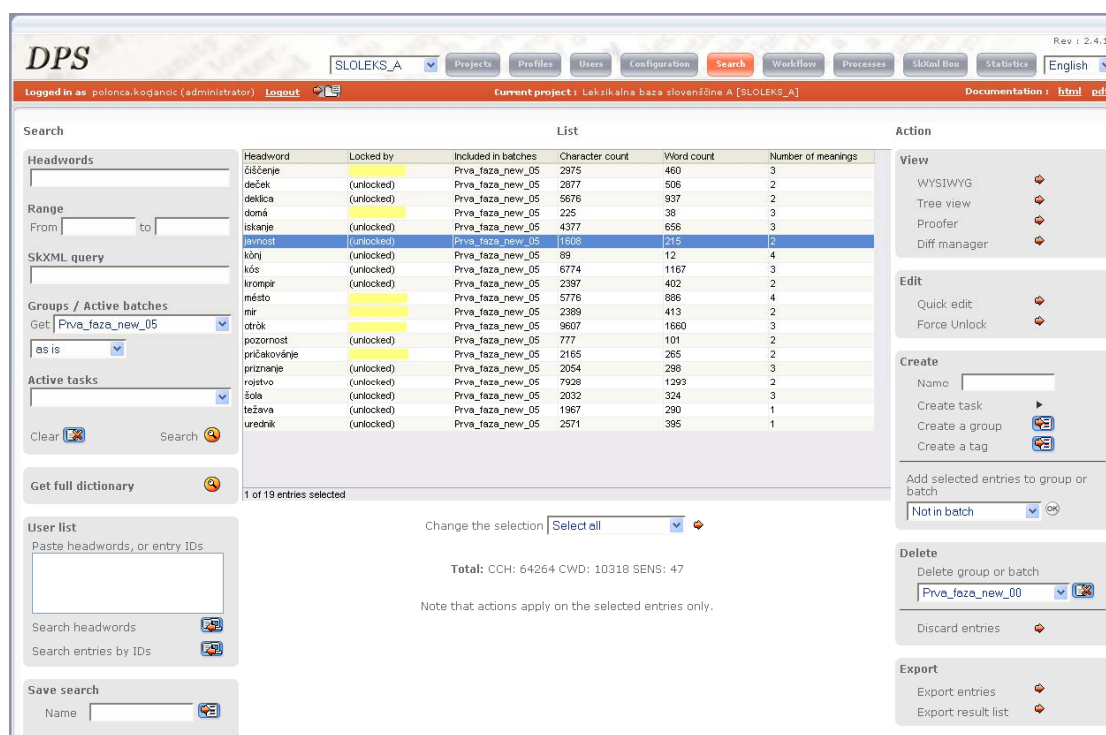
Nastavitve za izdelavo statistik:



Slika 95: DPS strežnik: konfiguracija za izdelavo statistike

4.4.5 Iskanje (Search)

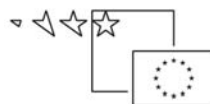
Okno za iskanje gesel na strežniku prikazuje slika:



Slika 96: DPS strežnik: Iskanje (Search)

Na levi strani zaslona so iskalne možnosti.

Osrednji del zaslona je namenjen prikazu seznama rezultatov, ki za vsako iztočnico vsebuje podatek, ali je geslo odklenjeno ali zaklenjeno, v katerem paketu je, ter koliko ima znakov, besed in pomenov.



Desni del okna je namenjen prehodu v različne postopke, ki jih opravljajo projektni skrbniki. Povzeto so to:

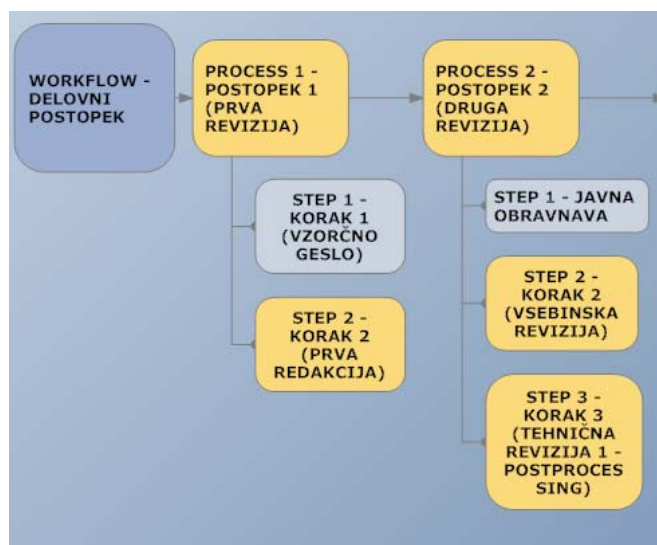
- ogled gesla (*WYSIWYG*),
- ogled drevesne strukture (*Tree view*),
- orodje za pregledovanje gesla (*Proofer*),
- orodje za primerjanje gesel (*Diff-manager*),
- hitro urejanje v DPS vmesniku (*Quick edit*),
- odklepanje (*Force Unlock*),
- ustvarjanje nalog, skupin in oznak (*Create ... /task, group, tag*),
- dodajanje gesel v pakete (*Add selected entries to group or batch*),
- brisanje skupine ali paketa (*Delete group or batch*),
- brisanje gesla (*Discard entries*),
- izvoz gesel (*Export entries*),
- izvoz seznama rezultatov (*Export result list*).

4.4.6 Delovni postopek (*Workflow*)

Ta razdelek projektnim skrbnikom omogoča:

- vzpostavitev delovnega postopka,
- določitev, katero gradivo gre v delovni postopek,
- razdelitev dela redaktorjem,
- časovni in količinski pregled nad potekom dela.

Delovni postopek (*Workflow*) je določilo, ki opredeljuje, skozi katere faze mora iti vsako geslo v podatkovni bazi. Delovni postopek je lahko samo en in se razdeli na posamezne **postopke** (*Processes*), ti pa nadalje na **korake** (*Steps*), ki so lahko obvezni ali neobvezni. Primer delovnega postopka in njegovih delov prikazuje naslednja slika:



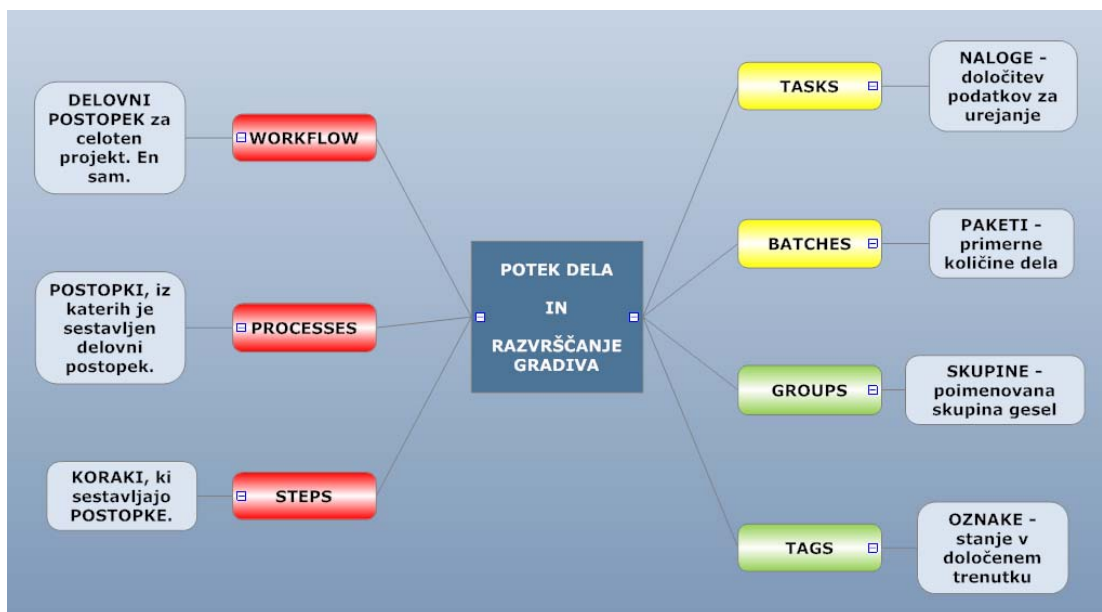
Slika 97: primer opredelitve delovnega postopka, posameznih postopkov v njem ter razdelitev na naloge



Sledi določitev, **katero gradivo gre v redatorsko urejanje** – npr. vsa gesla pri projektu ali pa samo nekatera, ter na podlagi katerih kriterijev. To so posamezne **naloge** (*Tasks*), ki jih nadalje razdelimo na količine, primerne za redatorsko obdelavo – te imenujemo **paketi** (*Batch*).

Možna je tudi izdelava skupin (*Group*) in oznak (*Tag*).

Postopek dela in razvrščanje gradiva povzema naslednja shema:



Slika 98: potek dela in razvrščanje gradiva pri projektu v programu DPS

V nadaljevanju sledi **razporeditev gradiva redaktorjem** – bodisi na podlagi števila znakov, pomenov ali gesel, lahko pa določeno količino gradiva razdelimo tudi na enake dele. Vsakemu paketu projektni skrbniki določijo redaktorja, fazo obdelave, začetni datum obdelave ter rok oddaje; po tem postopku je gradivo redaktorjem takoj na voljo (do njega dostopajo z DPS vmesnikom, gumb *Home*, *Refresh all projects*, *Select batch*).

Okno za organiziranje poteka dela in razvrščanje gradiva je naslednje:



The screenshot displays a software interface for managing batches. On the left, there are filters for 'Batch status' (Created, In progress, Done) and 'Tasks' (Prva_faza). Below these are 'Batch properties' for CR, AS, and UP. The main area is titled 'List of batches' and shows a table of batch entries. Each entry includes a batch name (e.g., Prva_faza_new_00), a task name (e.g., Avtorsko), a progress bar (e.g., 100%), and a 'Processorsko' section with a progress bar (e.g., 0%). The table also shows dates (UD, SD, DD), assigned users, and the number of entries. On the right, there is an 'Action' menu with options like 'View', 'Set batch properties', 'Modify', 'Export', and 'Process manager'. At the bottom, there is a 'Refresh all batches' button.

Slika 99: DPS strežnik: osnovno okno za organiziranje poteka dela in razvorščanje gradiva (Workflow)

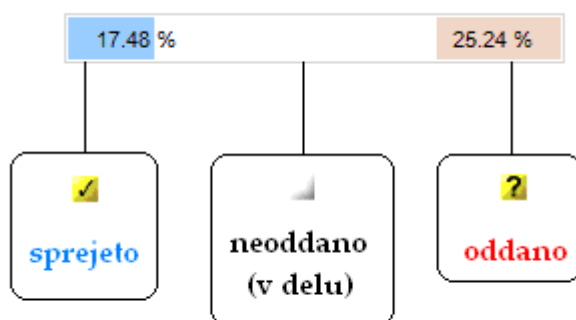
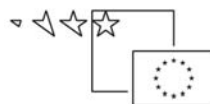
Na levi strani zaslona so iskalne možnosti pri delu s paketi in nalogami.

Osrednji del zaslona je namenjen prikazu paketov, kot so razvrščeni v posamezne postopke in/ali korake. Pri vsakem je tudi grafično prikazano, na kateri stopnji obdelave so.

Desni del okna je namenjen prehodu v različne postopke, ki jih opravljajo projektni skrbniki. Povzeto so to:

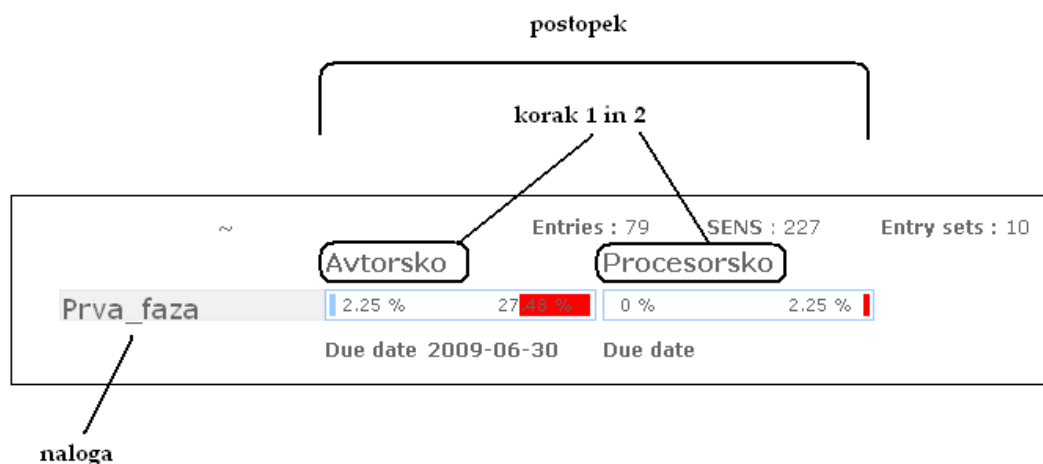
- ogled seznama zadetkov (*Go to list module*),
- ogled naloge (*Task*),
- prehod v modul za primerjanje različic na DPS strežniku (*Go to version manager*),
- določitev lastnosti paketa (*Set batch properties*: določitev redaktorja/redaktorjev, začetnega datuma obdelave, roka oddaje, možnost komentarja),
- urejanje nalog (*Modify / Remove tasks*),
- izvoz poročila o delu (*Export report*),
- prehod v urejanje delovnega postopka, postopkov in nalog (*Process manager*).

V tem razdelku DPS strežnika je omogočen tudi **pregled nad časovnim in količinskim potekom dela** tako pri posameznem paketu kot pri celotnem postopku, kar je prikazano tudi grafično. Napredek pri posameznem paketu vidi tudi redaktor ob komunikaciji z DPS strežnikom (*Home* v DPS vmesniku) – ponazarja ga **statusna pasica**:



Slika 100: prikaz poteka dela v paketu: statusna pasica

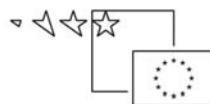
Nadzor nad količinami, obdelanimi v vseh paketih določene naloge in koraka, pa je na DPS strežniku ponazorjen takole:



Slika 101: prikaz poteka dela za vse pakete v posameznem koraku naloge (View task)

4.4.7 Tehnični postopki (Processes)

Razdelek za različna systemska opravila pri projektu vsebuje naslednje funkcije:



[Import entries...](#)
[Export...](#)
[Extract meta-data](#)
[Build cross-references table](#)
[Update all cross-references labels](#)
[Rebuild the annotation index](#)
[Update dictionary order](#)
[Manage tasks..., Run a task...](#)
[Custom BeanShell script...](#)
[Tagging the entire dictionary...](#)
[Dictionary statistics](#)
[Remove all entries](#), [Empty project](#)

[Running processes](#)

Slika 102: DPS strežnik: tehnični postopki za delo s podatkovno bazo

4.4.8 Zahtevno iskanje (SkXml Box)

Zahtevno tekstovno iskanje po celotni bazi je omogočeno v t. i. SkXml Box-u. Možno je iskati tako po vsebini kot v kombinaciji z XML elementi; gesla, ki ustrezajo navedenim pogojem, se izpišejo kot konkordance z najdenim pogojem na sredini. Omogočeno je tudi neposredno urejanje gesla. Okno za nastavitve zahtevnih iskanj prikazuje slika:

SkXml Box Evidence finder Evidence finder Tutorial Close session Rev: 1.0.0.11316 English

Logged in as: polonca.kocjancic (Polonca Kocjančič) Personal information July 1, 2009 5:09:44 PM

Search query: Go Close query builder

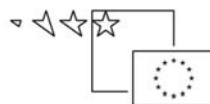
Remove this section

Search string:

<input checked="" type="radio"/> Exact word	<input checked="" type="radio"/> Contained in	tag	<input type="text" value="zveza"/>
<input type="radio"/> Word containing	<input type="radio"/> Entire content of	attribute	<input type="text" value="Choose attribute"/>
<input type="radio"/> Word beginning	<input type="radio"/> Beginning	value	<input type="text"/>
<input type="radio"/> Word ending	<input type="radio"/> Ending		
<input type="checkbox"/> Case sensitive	<input checked="" type="radio"/> within	tag	<input type="text" value="skladska_zveza"/>
<input type="checkbox"/> Accent sensitive	<input type="radio"/> as child of	attribute	<input type="text" value="Choose attribute"/>
		value	<input type="text"/>

And And not Or

Slika 103: DPS strežnik: okno za nastavitve zahtevnega iskanja (SkXml Box); primer iskanja: element <zveza>, ki je podrejen elementu <skladenjska zveza>, izpis od <zveza> dalje (znak %)



Prikaz rezultatov iskanja:

The screenshot shows the SkXml Box Evidence finder interface. The search query is '<skladnjska_zveza/><zveza'. The results are displayed in a table with columns for DocID, Inbel, and xml_0. The table shows 209 matches in 22 documents. The results are sorted by relevance, and the first few rows are visible. The interface also includes a search bar, a 'Query builder' button, and a 'Results layout' dropdown menu.

DocID	Inbel	xml_0
297	čas	< zveza > uvožnja na čas </ zveza >
297	čas	< zveza > čas in zastlonka </ zveza >
297	čas	< zveza > zastlonka in čas </ zveza >
297	čas	< zveza > za kratek/krajši čas </ zveza >
297	čas	< zveza > za dolgo/dlje/daljši čas </ zveza >
297	čas	< zveza > po dolgem/daljšem času </ zveza >
297	čas	< zveza > sadrži čas </ zveza >
297	čas	< zveza > imeti (kaj) ves čas pred očmi </ zveza >
297	čas	< zveza > največji kdo vseh časov </ zveza >
297	čas	< zveza > </ zveza >
297	čas	< zveza > čas trajanja </ zveza >
297	čas	< zveza > čas pred </ zveza >
297	čas	< zveza > čas po </ zveza >
297	čas	< zveza > čas od česa do česa / čas od ... do ... </ zveza >
297	čas	< zveza > čas od </ zveza >

Slika 104: DPS strežnik: prikaz rezultatov zahtevnega iskanja (SkXml Box), izgled Pretty Printed Xml

Natančna sintaksa iskalnega jezika v zahtevnem iskanju ter nastavitve točke izpisa (matchpoint), ki jo označuje znak za odstotek (%), je opisana v izvornih navodilih.

4.4.9 Statistika (Statistics)

Razdelek je namenjen prikazu podatkov

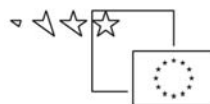
- o številu posameznih elementov v podatkovni bazi ter številu znakov, ki jih vsebujejo,
- o številu posameznih XML poti ter številu znakov, ki jih vsebujejo.

Izsek iz statistike prikazujeta naslednji sliki:

XML tags [Export statistics](#)

XML Tag name	Number of occurrences	Number of characters
b	5	33
besvrs	73	508
clanek	73	903358
enota	341	8325
frazeosloska_enota	268	68926
frazeosloske_zveze	51	68329
geslo	73	899715
glava	73	1820
i	2463	16005

Slika 105: DPS strežnik: statistika, podatki o elementih



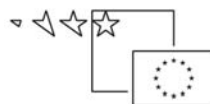
XML paths

[Export statistics](#)

XML path	Number of occurrences	Number of characters
/clanek	73	903358
/clanek/geslo	73	899715
/clanek/geslo/frazeoloske_zveze	51	68329
/clanek/geslo/frazeoloske_zveze/frazeoloska_enota	264	68329
/clanek/geslo/frazeoloske_zveze/frazeoloska_enota/enota	318	7705
/clanek/geslo/frazeoloske_zveze/frazeoloska_enota/indikator	270	9114
/clanek/geslo/frazeoloske_zveze/frazeoloska_enota/zgled	4	5465

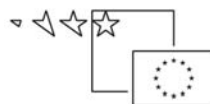
Slika 106: DPS strežnik: statistika, podatki o XML poteh

Statistiko je možno tudi izvoziti v tekstovni obliki.

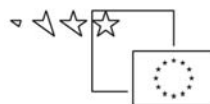


4.5 Izrazje, povezano s programom DPS

<i>angleško</i>	<i>slovensko</i>
Batch	paket
Change node name	spremeni ime elementa
Comments	Komentarji
Configuration	Konfiguracija
Dictionary Production System (DPS)	DPS
Download	Prenesi z DPS strežnika
DPS Website	DPS strežnik
DTD (Document Type Definition)	DTD (definicija tipa dokumenta)
Editor	redaktor
Entry Editor	DPS vmesnik
Group	skupina
Highlight DTD errors	označi napake DTD-ja v geslu (validacija)
Homepage	domača stran DPS strežnika
Local store	Lokalno skladišče
Lose edits	Ignoriraj spremembe
Merge left	pridruži vsebino prejšnjemu elementu
Merge right	pridruži vsebino naslednjemu elementu
Not in batch	Zunaj paketa
Outstore	Začasno skladišče
Process (pri poteku dela)	postopek
Processes (za sistemske skrbnike)	Tehnični postopki
Profiles	Profili
Project manager	projektni skrbnik
Projects	Projekti
Refresh Element	osveži element
Refresh Entry	osveži geslo
Report DTD errors	razloži napake DTD-ja v označenem elementu
Repository	Skladišče
Search	Iskanje
Select all	Izberi vse
Show validation result	pokaži rezultat validacije = razloži napake DTD-ja v označenem elementu
SkXml Box	Zahtevno iskanje
Statistics	Statistika
Stats	Statistika
Step	korak
Submitting entries	Oddaja gesel
Swap inline state	zamenjaj razloženi/strnjeni izgled strukture elementov
System administrator	sistemske skrbnik



System templates	sistemske šablone
Tag	oznaka
Task	naloga
Template	šablona
Template picker	Izbor šablone
Toggle	Menjaj status gesla
(Un)collapse all elements	razširi/skrči element
Upload	Naloži na DPS strežnik
Upload entries	Nalaganje gesel na DPS strežnik
User templates	uporabniške šablone
Users	Uporabniki
Versions	različice (gesla)
View HTML Source	pokaži izvorno kodo HTML
View XML Source	pokaži izvorno kodo XML za označeni element
Workflow	delovni postopek



5

Format zapisa in DTD

*pri
leksikalni bazi
za slovenščino*



5 Format zapisa in DTD pri leksikalni bazi za slovenščino

5.1 XML: razširljiv označevalni jezik (*Extensible Markup Language*)

Leksikalna baza za slovenščino je zapisana v računalniškem jeziku XML, ki med drugim omogoča opisovanje strukturiranih podatkov. Format XML je zasnovan na ideji ločitve vsebinskega in oblikovnega označevanja besedil. Vsak del besedila, ki ga želimo ločiti bodisi zaradi vsebinskih ali oblikovnih razlogov, označimo s poljubnim elementom in tako dobimo nadzor nad logično strukturo določenega besedila. Pri delu z jezikovnimi podatkovnimi bazami je ta nadzor ključen za urejen videz gesel.

Več o XML: <http://www.w3schools.com/xml/default.asp>.

5.2 Osnovni pojmi in interpretiranje DTD-ja

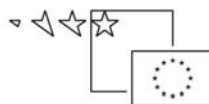
Koda (*Tag*) je osnovni gradnik jezika XML in ima predpisan zapis z ostrimi oklepaji:

- začetna koda, ki označuje začetek označenega besedila, npr. <clanek>
- končna koda, ki označuje konec označenega besedila, npr. </clanek>

Element (*Element*) je struktura, sestavljena iz začetne in končne kode ter informacije med njima (npr. <i>dihati</i>). Lahko imajo attribute in lahko vsebujejo druge elemente ali besedilo (#PCDATA) ali oboje. Elementi lahko vsebujejo poljubne črke in druge znake, ne smejo vsebovati presledkov, se začeti s cifro ali ločilom, prav tako pa ne neposredno z zaporedjem "xml". Elementi so lahko tudi prazni (*empty elements*).

DTD oziroma definicija tipa dokumenta (*Document Type Definition*, **.dtd**) je standardiziran dokument, ki ureja hierarhične odnose med posameznimi XML elementi za vsebinske dele gesla v leksikalni bazi. V njem so določene vse lastnosti elementov in hierarhični odnosi med njimi. Glede na ta pravila si elementi nato sledijo v besedilu. V leksikalni bazi za slovenščino je na najvišji ravni t. i. korenski element <baza></baza >, ki vsebuje posamezne članke oziroma gesla.

Preverjanje ali validacijo XML strukture regulira **validacija**, ki je vgrajena v program DPS in opisana v 4. poglavju kazalnika 5. Namenjena je interpretiranju in preverjanju pravilnosti logične strukture.



Operatorji so dveh vrst:

BINARNI: ti določajo razmerja med elementi ali skupinami elementov v oklepajih, ki se pojavljajo levo in desno od operatorja:

<i>znak</i>	<i>pomeni</i>
,	elementa ali skupini si morata slediti v zapovedanem vrstnem redu
&	elementa ali skupini si lahko sledita v poljubnem vrstnem redu
	lahko se pojavi le en element ali skupina na levi in desni od operatorja

UNARNI: določajo, kolikokrat se določen element na nekem mestu lahko pojavi:

ZNAK	POMENI
+	element se mora pojaviti vsaj enkrat in se lahko ponavlja
?	element se lahko pojavi le enkrat ali nobenkrat
*	element se lahko pojavi večkrat ali nobenkrat

Unarni operatorji imajo prednost pred binarnimi.

Primer definicije elementa <pomen>:

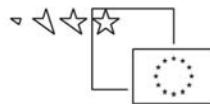
```
<!ELEMENT pomen (indikator, oznaka*, pomenska_shema, skladskejske_skupine?, skladskejske_zveze?, podpomen*, stalne_zveze?) >
```

Definicija elementa se začne z ostrim oklepajem in klicajem, sledi t. i. rezervirano ime ELEMENT, presledek in ime elementa (*pomen*). Sledi presledek in navedba podrejenih elementov, dovoljenih v elementu *pomen*. Definicija elementa se konča z ostrim zaklepajem. Navedbo v definiciji elementa preberemo takole:

- na prvem mestu je element <indikator>, ki je nujen in se *mora* pojaviti točno enkrat,
- *lahko* sledi *en ali več* elementov <oznaka>
- *nujno* sledi *en* element <pomenska_shema>,
- *lahko* sledi *en* element <skladskejske_skupine>, temu enako <skladskejske_zveze>,
- *lahko* sledi *en ali več* elementov <podpomen>,
- *lahko* sledi *en* element <stalne_zveze>.

Atributi (*Attributes*)

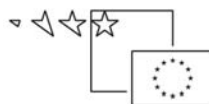
Atribute uporabljamo zaradi možnosti podrobnejšega kvalificiranja nekega elementa z vrednostmi, ki so določene v definiciji atributa. Atributi so strukturni gradniki XML-a, ki s svojimi vrednostmi spreminjajo funkcije elementa. Vse vrednosti atributov morajo biti med narekovaji. Atributi so v začetni kodi elementa in so zapisani npr. takole:
<s kat="ena"> - najprej torej ime elementa, presledek, ime atributa, enačaj in vrednost atributa med narekovaji.



5.3 DTD pri leksikalni bazi za slovenščino

DTD za leksikografsko urejanje leksikalne baze za slovenščino je naslednji (leksbaza-14-LEX.dtd, 28. 4. 2009):

```
<!ELEMENT baza (clanek)+ >
<!ELEMENT clanek ((nadgeslo, nadgeslo+) | (glava, geslo)) >
<!ELEMENT glava (oblika, zaglavje) >
<!ELEMENT oblika (zapis, iztocnica) >
<!ELEMENT zapis (#PCDATA) >
<!ELEMENT iztocnica (#PCDATA) >
<!ELEMENT zaglavje (besvrs, oznaka*, kazalka?) >
<!ELEMENT besvrs (#PCDATA) >
<!ELEMENT oznaka (#PCDATA) >
<!ELEMENT kazalka (#PCDATA) >
<!ELEMENT nadgeslo (glava, geslo) >
<!ELEMENT geslo (pomen+, frazeoloske_zveze?, idiomatske_zveze?) >
<!ELEMENT pomen (indikator, oznaka*, pomenska_shema, skladijske_skupine?,
skladijske_zveze?, podpomen*, stalne_zveze?) >
<!ELEMENT stalne_zveze (stalna_zveza+) >
<!ELEMENT stalna_zveza (zveza, indikator, oznaka*, struktura, kolokacija*, zgledi) >
<!ELEMENT indikator (#PCDATA | pr)* >
<!ELEMENT pr (#PCDATA) >
<!ELEMENT pomenska_shema (#PCDATA) >
<!ELEMENT skladijske_skupine (skladijska_struktura+) >
<!ELEMENT skladijska_struktura (struktura, vzorec*, kolokacija*, zgledi) >
<!ELEMENT struktura (#PCDATA | r)* >
<!ELEMENT vzorec (#PCDATA | r)* >
<!ELEMENT kolokacija (#PCDATA | k)* >
<!ELEMENT r (#PCDATA) >
<!ELEMENT k (#PCDATA) >
<!ELEMENT zgledi (zglied+) >
<!ELEMENT zglied (#PCDATA | i)* >
<!ELEMENT i (#PCDATA) >
<!ELEMENT enota (#PCDATA) >
<!ELEMENT skladijske_zveze (skladijska_zveza+) >
<!ELEMENT skladijska_zveza (zveza+, vzorec*, kolokacija*, zgledi) >
<!ELEMENT zveza (#PCDATA | k)* >
<!ELEMENT frazeoloske_zveze (frazeoloska_enota+) >
<!ELEMENT frazeoloska_enota (enota+, indikator, oznaka*, zgledi?) >
<!ELEMENT idiomatske_zveze (idiomatska_enota+) >
<!ELEMENT idiomatska_enota (enota+, indikator, oznaka*, zgledi?) >
<!ELEMENT podpomen (indikator, oznaka*, pomenska_shema, skladijske_skupine?,
skladijske_zveze?, stalne_zveze?) >
```



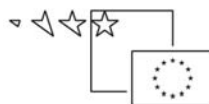
5.4 Primer gesla

5.4.1 Izvoz iz programa DPS v obliki XML

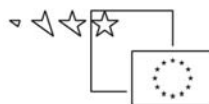
Za ilustracijo zapisa leksikalne baze za slovenščino predstavljamo vzorčni izsek gesla *pasti*:

```
<?xml version="1.0" encoding="utf-8" ?>
<dps-data project="SLOLEKS_A" guid="583c10bfdbd326ba:-13deba50:121541de8f3:-7ffe--
7eae3529-9c5a-4742-903c-1fc9aa464361">
<clanek xmlns="urn:SLOLEKS" xmlns:d="urn:SLOLEKS">
  <glava>
    <oblika>
      <zapis>pasti</zapis>
      <iztocnica>pásti</iztocnica>
    </oblika>
    <zaglavje>
      <besvrs>glagol</besvrs>
    </zaglavje>
  </glava>

  <geslo>
    <pomen>
      <indikator>premikati se od zgoraj proti navzdol</indikator>
      <pomenska_shema>PREDMET zaradi sile težnosti pade KAM</pomenska_shema>
    </pomen>
    <skladenjske_skupine>
      <skladenjska_struktura>
        <struktura>sbz1 GBZ rbz/na sbz4/v sbz4</struktura>
        <vzorec>kaj pade kam</vzorec>
        <vzorec>kaj pade na kaj</vzorec>
        <vzorec>kaj pade v kaj</vzorec>
      </skladenjska_struktura>
      <kolokacija><k>bomba, granata, drevo</k> pade </kolokacija>
      <kolokacija>pasti na <k>dno, Zemljo, zemljo, tla</k></kolokacija>
      <kolokacija>pasti v <k>prepad, globino, morje</k>
      </kolokacija>
    </skladenjske_skupine>
    <zgledi>
      <zgled>Se še spominjate dne, ko je padla bomba čisto blizu vaše gostilne.</zgled>
      <zgled>Svetleča granata je padla kakih petdeset metrov stran.</zgled>
      <zgled>Pri odcepu za Kostanjevico je padlo na dolenjko drevo.</zgled>
      <zgled>Mnoge živali in rastline poginejo in padejo na vodno dno.</zgled>
      <zgled>Vsako leto pade na Zemljo iz vesolja obilica kamenin.</zgled>
      <zgled>Debele kostanjeve ježice so padle na zemljo.</zgled>
      <zgled>V stanju breztežnosti nič ne pade na tla.</zgled>
      <zgled>Dve gondoli sta padli v globok prepad.</zgled>
      <zgled>Propeler je razsekalo in helikopter je kot kamen padel v globino.</zgled>
      <zgled>Po izjavah posadke naj bi bomba padla v morje daleč proč od
      letalonosilke.</zgled>
      <zgled>Kaj pa, če helikopter pade dol?</zgled>
    </zgledi>
  </skladenjska_struktura>
</skladenjske_zveze>
```



= <skladenjska_zveza>
 = <zveza>pasti pod kotom <k>x</k> stopinj</zveza>
 = <zgledi>
 <zgled>Kovinski meteorit s premerom 40 metrov, ki pade pod kotom več kot 30 stopinj, bi tla še dosegel, preden bi se povsem stalil oziroma izhlapel.</zgled>
 <zgled>Pred 50 milijoni let je na Luno padel pod blagim kotom težak asteroid.</zgled>
 </zgledi>
 </skladenjska_zveza>
 </skladenjske_zveze>
 = <podpomen>
 <indikator>predmet ni več kje nameščen ali pritrjen</indikator>
 <pomenska_shema>PREDMET pade od KOD</pomenska_shema>
 = <skladenjske_skupine>
 = <skladenjska_struktura>
 <struktura />
 <vzorec>kaj pade s/z česa</vzorec>
 <vzorec>kaj pade iz česa</vzorec>
 = <kolokacija>pasti z <k>mize</k></kolokacija>
 = <kolokacija>pasti iz <k>okvirja</k></kolokacija>
 = <zgledi>
 <zgled>Če prevrnemo stekleno vazo, da pade z mize na tla, se razbije.</zgled>
 <zgled>Okensko steklo je zacvililo in padlo iz razpadajočega okvirja.</zgled>
 </zgledi>
 </skladenjska_struktura>
 </skladenjske_skupine>
 </podpomen>
 ± <podpomen>
 /.../
 </pomen>
 ± <pomen>
 ± <pomen>
 ± <pomen>
 ± <pomen>
 ± <frazeoloske_zveze>
 = <frazeoloska_enota>
 <enota>(kaj) (ne) pade na pamet/misel/glavo/um (komu)</enota>
 <indikator>kdo si izmisli kaj presenetljivega, čudnega, nenavadnega</indikator>
 = <zgledi>
 <zgled>še vedno ostaja vprašanje, koliko je tisto, kar ti pade na pamet, res iskreno</zgled>
 </zgledi>
 </frazeoloska_enota>
 ± <frazeoloska_enota>
 /.../
 ± <frazeoloska_enota>
 ± <frazeoloska_enota>
 </frazeoloske_zveze>
 ± <idiomatske_zveze>
 = <idiomatska_enota>
 <enota>Jabolko ne pade daleč od drevesa.</enota>
 <indikator />
 = <zgledi>
 <zgled>Kot kaže, pa jabolko ne pade daleč od drevesa, saj je danes skoraj v vsaki hiši v Butajnovi kdo, ki zna prav ubrano raztegovati meh</zgled>



<zgled>Saj poznate ljudski pregovor, da jabolko ne pade daleč od drevesa. Ded mi je vcepil ljubezen do konj, ki me ne bo nikoli minila</zgled>

</zgledi>

</idiomatska_enota>

+<idiomatska_enota>

+<idiomatska_enota>

+<idiomatska_enota>

+<idiomatska_enota>

+<idiomatska_enota>

</idiomatske_zveze>

</geslo>

</clanek>

</dps-data>

5.4.2 Izvoz iz programa DPS v obliki PDF oziroma predogled tiskanja

Glava gesla *pasti* in prvi pomen s skladijskimi zvezami:

pasti *glagol*

1 premikati se od zgoraj proti navzdol navadno v 3. osebi;

PREDMET zaradi sile težnosti pade KAM

a) Struktura: sbz1 GBZ rbz/na sbz4/v sbz4

- ▶ kaj pade kam
- ▶ kaj pade na kaj
- ▶ kaj pade v kaj

- [bomba, granata, drevo] pade
- pasti na [dno, Zemljo, zemljo, tla]
- pasti v [prepad, globino, morje]

- Kaj pa, če helikopter pade dol?
- Se še spominjate dne, ko je **padla** bomba čisto blizu vaše gostilne.
- Svetleča granata je **padla** kakih petdeset metrov stran.
- Pri odcepu za Kostanjevico je **padlo** na doljenko drevo.
- Mnoge živali in rastline poginejo in padejo na vodno dno.
- Vsako leto pade na Zemljo iz vesolja obilica kamenin.
- Debele kostanjeve ježice so padle na zemljo.
- V stanju breztežnosti nič ne pade na tla.
- Dve gondoli sta padli v globok prepad.
- Propeler je razsekalo in helikopter je kot kamen padel v globino.
- Po izjavah posadke naj bi bomba padla v morje daleč proč od letalonosilke.

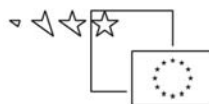
skladijske zveze:

pasti pod kotom

pasti pod kotom [x] stopinj

- Kovinski meteorit s premerom 40 metrov, ki pade pod kotom več kot 30 stopinj, bi tla še dosegel, preden bi se povsem stalil oziroma izhlapel.
- Pred 50 milijoni let je na Luno padel pod blagim kotom težak asteroid.

Slika 107: glava gesla *pasti* in prvi pomen s skladijskimi zvezami, predogled v programu DPS



Podpomen 1.2 glagola *pasti*:

1.2 izgubiti, izpustiti navadno v 3. osebi

PREDMET pade ČLOVEKU KAM |ali| MU pade od KOD

a) Struktura: sbz1 GBZ sbz3 (iz sbz2) v sbz4/na sbz4/pod sbz4

- ▶ kaj pade komu v kaj
- ▶ kaj pade komu na kaj
- ▶ kaj pade komu pod kaj
- ▶ kaj pade komu iz česa

- [pištola, cigareta] pade
- pasti v [jarek, jašek]
- pasti na [tla]
- pasti pod [mizo]
- pasti iz [rok, roke, ust]
- pasti iz [žepa]

- Vžigalice so ji padle v cestni jarek in so zdaj vse premočene.
- Fantku je plastični tovarnjaček padel v jašek.
- Pištola mu je med prerivanjem padla na tla.
- Če vam hrana pade na tla, na katerih so mikroorganizmi, je lahko okužena v petih sekundah.
- Vse kaže, da mi je padlo pod mizo deset šillingov.
- Poberem knjigo, ki mi je padla iz rok in nadaljujem branje.
- Sneguljčici zastrupljeni košček jabolka pade iz ust in dekle oživi.
- Med begom mu je ukradena denarnica padla iz žepa
- Nemara bi zadostovalo, če rečem: Nekaj vam je padlo na tla. Ali pa preprosto: Nekaj ste izgubili.

Slika 108: podpomen 1.2 glagola *pasti*, predogled v programu DPS

Frazeološke zveze pri glagolu *pasti*:

FRAZEOLOŠKE ZVEZE:

(kaj) (ne) pade na pamet/misel/glavo/um (komu)

kdo si izmisli kaj presenetljivega, čudnega, nenavadnega

- še vedno ostaja vprašanje, koliko je tisto, kar ti pade na pamet, res iskreno

kaj (ne) pade na kraj pameti (komu)

kaj ne pade niti na kraj pameti (komu)

kot grožnja, opozorilo

- Da ti ne bi slučajno padlo na kraj pameti, da bi še kdaj parkiral na MOJEM prostoru !

(kaj) pade v vodo

propasti, neuspeti

- Podražitev otroškega varstva je padla v vodo

(kaj) pade iz glave (komu)

pozabiti kaj

- Oprosti, tvoj priimek mi je padel iz glave . "

zastor/zavesa pade

predstava (festival, gledališče, film) se zaključi

- Maribor postane zaspana provinca, takoj ko padejo festivalski zastori .
- In ko zastor pade, se oglasi gromek aplavz, z vseh strani
- Zavesa je spet padla in na odru so hiteli postavljati novo sceno

zastor/zavesa pade

dogodek, proces itd. se konča/zaključi

- Letošnji zastor v Planici je padel, ostalo pa je bolj klavrno zakulisje
- Seveda pa s tem v zgodbi Toncity še ni padel zadnji zastor .
- Tako, zavesa je padla, volilne kampanje je konec .

(izgledati, videti) kot bi (kam) padla bomba

oči padejo iz jamic (komu)

začuditi se, občudovati

- Oči bi ji kmalu padle iz jamic, ko je prebrala, kaj je pisalo

Slika 109: frazeološke zveze pri glagolu *pasti*, odlomek iz predogleda v programu DPS



Idiomatske zveze pri glagolu *pasti*:

IDIOMATSKE ZVEZE:

Jabolko ne pade daleč od drevesa.

- Kot kaže, pa jabolko ne pade daleč od drevesa, saj je danes skoraj v vsaki hiši v Butajnovi kdo, ki zna prav ubrano raztegovati meh
- Saj poznate ljudski pregovor, da jabolko ne pade daleč od drevesa . Ded mi je vcepil ljubezen do konj, ki me ne bo nikoli minila

Kdor visoko leta, nizko pade.

- Kdor visoko leta, nizko pade ! « je pregovor, ki bi ga lahko v vrstah Krima Ete Neutro Robertsa vzeli za svojega .

Kocka je padla odločilo se je

- Novinarsko poročilo z naslovom Kocka je padla ? govori o odločitvi za prenovo štadiona na podlagi načrta ZIL, ki predvideva povečanje števila sedežev za blizu 18.000 ljudi

Okostnjak, ki pade iz omare

- Lastninjenje zavarovalnic je še en okostnjak, ki mi je padel iz omare na glavo (kot privatizacija bank), « je dejal minister

kdo pade v jamo, ki jo je kopal sam/sebi komu/drugemu

- Nikar, nikar, drugače bodo vse tri padle v jamo, ki so jo same izkopale .
- v najslabšem pa vohun pade v jamo, ki jo je kopal drugemu .
- politika je Ljubljancanom obljubila nov štadion, zdaj pa so padli v jamo, ki so jo skopali sami sebi .
- Najhujše pa je, da so Palestinci sami padli v jamo, ki so jo kopali drugim .

pobirati, dobiti itd. drobtine/drobtinice, ki padejo s (čigave/katere) mize

dobiti to, kar ostane komu, ki je pomembnejši, boljši

- Ni vzroka, da bi pobirali drobtinice, ki drugim padejo z mize .
- lobiste, ki jim je le do lastne promocije, nemara celo do tega, da pozobajo kakšno drobtinico, ki bi padla s Holdingove mize

Slika 110: idiomatske zveze pri glagolu *pasti*, predogled v programu DPS



6

Viri in

spletne povezave



6.1 Viri

Atkins, S, Rundell, M. (2008): *Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Baroni, M., Kilgarriff, A., Pomikálek, J., Rychlý, P. (2006a): "WebBootCaT: instant domain-specific corpora to support human translators". V: *Proceedings of EAMT 2006*, Oslo, 247-252.

Baroni, M., Kilgarriff, A., Pomikálek, J., Rychlý, P. (2006b): "WebBootCaT: a web tool for instant corpora". V: *Proceedings / XII Euralex International Congress*, Alessandria: Edizioni dell'Orso, 123-131.

Kilgarriff A., Rundell, M. (2002): "Lexical Profiling Software and its lexicographic applications: a case study". V: *Proceedings EURALEX 2002*, Copenhagen, str. 807-818.

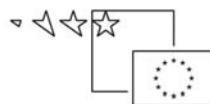
Kilgarriff, A., Rychlý, P., Smrz, P., Tugwell, D. (2004). "The Sketch Engine" V: *Proceedings / XI Euralex International Congress*, Lorient: Université de Bretagne-Sud, str. 105-116.

Kilgarriff, A., Husak, M., McAdam, K., Rundell, M., Rychlý, P. (2008): "Automatically Finding Good Dictionary Examples in a Corpus". V: Bernal, E., DeCesaris, J (ur.): *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Barcelona: Documenta Universitaria: Institut universitari de lingüística aplicada, Universitat Pompeu Fabra, str. 425-432.

Krek, S., Kilgarriff, A. (2006): "Slovene Word Sketches". V: Erjavec, T., Žganec Gros, J. (ur.): *Proceedings of the 5th Slovenian and 1st International Languages Technology Conference*. Ljubljana, Slovenia, str. 62-67.

Rundell, M. (ur.) (2002): *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan.

Rundell, M. (ur.) (2007): *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan.



6.2 Spletne povezave

DPS: prijava na strežnik	http://ssi.dps.idm.fr/web/Signon.html
FidaPLUS	http://www.fidaplus.net/
IDM: domača stran podjetja	http://www.idm.fr
IDM: predstavitevna stran DPS	http://www.idm.fr/products/dictionary_writing_system/27/
IMS Corpus Workbench	http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/
IMS Corpus Workbench: sintaksa iskanj	http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/COPSyntax.html
Oblikoskladenjske oznake za slovenščino: celotna dokumentacija	http://nl.ijs.si/ME/V3/msd/html/
Oblikoskladenjske oznake za slovenščino: seznam	http://www.sketchengine.co.uk/tagsets/slovene.html
Pomoč: .net Framework	http://www.microsoft.com/downloads/en/default.aspx
Pomoč: MS Office filter	http://www.microsoft.com/downloads/details.aspx?FamilyId=941B3470-3AE9-4AEE-8F43-C6BB74CD1466&displaylang=en
Pomoč: Sketch Engine: odložišče (Mozilla Firefox)	http://trac.sketchengine.co.uk/wiki/ProblemSolving/OneClick
Sketch Engine	http://www.sketchengine.co.uk/
Sketch Engine Beta	http://beta.sketchengine.co.uk/
Sketch Engine: izvorna navodila	http://trac.sketchengine.co.uk/wiki/SkE/DocsIndex
Sketch Engine: izvorna navodila	http://trac.sketchengine.co.uk/wiki/SkE/DocsIndex
Sketch Engine: sintaksa CQL iskanj in slovničnih relacij	http://trac.sketchengine.co.uk/wiki/SkE/CorpusQuerying
Sketch Engine: sintaksa iskanj v CQL	http://trac.sketchengine.co.uk/wiki/SkE/CorpusQuerying
Sketch Engine: statistike v orodju (<i>Statistics used in the Sketch Engine</i>)	http://trac.sketchengine.co.uk/wiki/SkE/DocsIndex
XML W3schools	http://www.w3schools.com/xml/default.asp