

УНИВЕРЗИТЕТ У БЕОГРАДУ  
ФИЛОЛОШКИ ФАКУЛТЕТ

Мр Небојша М. Васиљевић

АУТОМАТСКА ОБРАДА ПРАВНИХ ТЕКСТОВА  
НА СРПСКОМ ЈЕЗИКУ

докторска дисертација

Београд, 2014.

UNIVERSITY OF BELGRADE  
FACULTY OF PHILOLOGY

Nebojša M. Vasiljević MSc

**AUTOMATIC PROCESSING OF LEGAL TEXT  
IN SERBIAN LANGUAGE**

doctoral dissertation

Belgrade, 2014

# Комисија

Ментор: Др Цветана Крстев, редовни професор,  
Универзитет у Београду, Филолошки факултет

Чланови комисије: Др Александра Вранеш, редовни професор,  
Универзитет у Београду, Филолошки факултет

Др Божо Ђорић, редовни професор,  
Универзитет у Београду, Филолошки факултет

Др Добросав Миловановић, ванредни професор,  
Универзитет у Београду, Правни факултет

Др Душко Витас, ванредни професор,  
Универзитет у Београду, Математички факултет

Датум одбране: \_\_\_\_\_

# Подаци о докторској дисертацији

## „Аутоматска обрада правних текстова на српском језику”

### Сажетак

Предмет истраживања у овој дисертацији су посебна језичка правила у текстовима прописа на српском језику која се могу изразити методама рачунарске лингвистике, са циљем да се на теоријском нивоу створи основа за развој софтверских алата за обраду текстова прописа и да се испитају конкретне могућности имплементације. Истраживање посебно обухвата:

1. опис општих језичких особености правног језика на основу припремљеног корпуса и расположивих језичких ресурса;
2. прецизан опис методама рачунарске лингвистике посебних језичких правила која се користе у прописима;
3. имплементацију синтаксне анализе текста прописа, као почетне фазе за посебне обраде;
4. израду експерименталних софтверских алата заснованих на претходном, као и анализу могућности примене таквих и сличних решења.

У овом истраживању се користе методе засноване на правилима, насупрот статистичким методама које је такође могуће користити у обради природног језика. Коришћена је методологија локалних граматика представљених коначним трансдукторима у форми синтаксних графова и уз подршку електронских речника.

Као технолошка основа за подршку наведеној методологији је коришћен систем *Unitex* који је развијен у *Laboratoire d'Automatique Documentaire et Linguistique (LADL)* у оквиру Универзитета Марн-ла-Вале, као и електронски речници развијени у оквиру Групе за језичке технологије Универзитета у Београду.

Одговарајућим методама су посебно детаљно описане упућујуће фразе и структура правног акта. На основу тога је израђен софтверски алат који за

дати чисти текст прописа формира приказ текста тако да се прегледно изражава препозната структура и мрежа упућивања, омогућава навигација кроз упућивања и приказују се идентификовани пропусти у структури и упућивањима.

Током истраживања је развијен и одређен број софтверских алата у сврху практичног коришћења у истраживачком раду и детаљи имплементације тих алата су представљени у посебном поглављу дисертације, као допринос резултатима истраживања.

## **Остали подаци о дисертацији**

**Кључне речи:** рачунарска лингвистика, обрада природног језика, језик прописа, српски језик, локалне граматике, електронски речник

**Научна област:** филолошке науке

**Ужа научна област:** рачунарска лингвистика

**УДК број:** 81'322(043.3)=163.41

# **Information on the Doctoral Dissertation**

## **Automatic Processing of Legal Text**

### **in the Serbian Language**

#### **Abstract**

The subject of the research presented in this dissertation concerns specific language rules in legislative texts in the Serbian language that can be expressed by computational linguistics methods with the goal to build foundations, on a theoretical level, for the development of legislative text software tools, and to examine concrete implementation possibilities. The research specifically covers:

1. description of the general lingual characteristics of legal language based on the prepared corpus and the available language resources;
2. precise description, by computational linguistic methods, specific language rules that are used in legislation;
3. legislative text parser implementation, as a common initial phase of more specific processing;
4. development of experimental software tools based on the above, as well as an analysis of implementation possibilities of these and similar solutions.

Rule based methods are used in this research, as opposed to statistical methods that also can be used in natural language processing. The methodology used is based on local grammars represented by finite transducers in the form of syntax graphs, and on electronic dictionaries.

Technology base that supports this methodology is the Unitex system developed at Laboratoire d'Automatique Documentaire et Linguistique (LADL) within the University of Paris-Est Marne-la-Vallée, as well as electronic dictionaries developed by the Human Language Technology Group at the University of Belgrade.

Referencing phrases and the structure of legal act are described in particular details using appropriate methods. Based on that, the software tool is developed which

takes a plain text of legislation and produces clear presentation of the recognised legislation structure and the network of references, with navigation through references and identified fails in the structure and references.

During research, a number of software tools were developed for the purpose of practical use in the research activities and implementation details of these tools are presented in the last chapter of the dissertation, which is also considered to be a contribution to the research results.

## **Other Information on the Dissertation**

**Keywords:** computational linguistics, natural language processing, legislative language, Serbian language, local grammars, electronic dictionary

**Scientific area:** philology science

**Scientific subarea:** computational linguistics

**UDC number:** 81'322(043.3)=163.41

# Садржај

<b>1</b>	<b>Увод</b>	<b>1</b>
1.1	Предмет и циљ истраживања . . . . .	2
1.2	Методологија . . . . .	6
<b>2</b>	<b>Правни текст и природно-језичка обрада</b>	<b>8</b>
2.1	Природно-језичка анализа текста . . . . .	10
2.1.1	Изворни текст у специфичном формату . . . . .	12
2.2	Обрада након природно-језичке анализе . . . . .	13
2.3	Друга истраживања о природно-језичкој обради правних текстова	17
2.4	Расположиви алати за обраду природног језика . . . . .	19
<b>3</b>	<b>Корпус</b>	<b>22</b>
3.1	Претходна обрада корпуса . . . . .	24
3.1.1	Преузимање лексичког материјала са веб-сајта . . . . .	25
3.1.2	Конверзија у чист текст . . . . .	26
3.1.3	Спајање и библиографска анотација . . . . .	28
3.1.4	Претходна обрада у оквиру <i>Unitex</i> -а . . . . .	30
3.2	Основни статистички показатељи корпуса . . . . .	34
3.3	Примена електронских речника . . . . .	38
3.4	Лексичке особености законских текстова . . . . .	44
3.5	Непознате речи . . . . .	52
3.6	Именовани ентитети . . . . .	55
<b>4</b>	<b>Упућујуће фразе</b>	<b>59</b>
4.1	Упућивање и упућујућа фраза . . . . .	60



4.2	Цитирање одредби прописа у оквиру овог рада . . . . .	63
4.3	Елементарне форме упућивања . . . . .	63
4.3.1	Упућивање на акт . . . . .	63
4.3.2	Упућивање на делове акта . . . . .	65
4.3.3	Суфикси у нумерацијама код измена и допуна прописа . . . . .	68
4.3.4	Повратно упућивање . . . . .	70
4.3.5	Просто једноструко упућивање . . . . .	71
4.3.6	Просто вишеструко упућивање . . . . .	71
4.3.7	Просто упућивање у општем случају . . . . .	72
4.3.8	Ниво простог и повратног упућивања . . . . .	72
4.4	Затворене и отворене упућујуће фразе . . . . .	73
4.5	Почетни и крајњи ниво упућујуће фразе . . . . .	74
4.6	Слагање упућујућих фраза . . . . .	75
4.6.1	Конструкција отворених упућујућих фраза . . . . .	76
4.6.2	Конструкција затворених упућујућих фраза . . . . .	77
4.7	Разлике у нормативној пракси . . . . .	78
4.7.1	Историјске разлике у нормативној пракси . . . . .	78
4.7.2	Разлике у савременој нормативној пракси . . . . .	80
4.8	Препознавање упућујућих фраза у корпусу . . . . .	85
4.9	Препознавање неисправних упућујућих фраза . . . . .	89
<b>5</b>	<b>Логичка структура</b>	<b>93</b>
5.1	Основни елементи логичке структуре прописа . . . . .	93
5.2	Поглед на логичку структуру прописа прилагођен локалним гра- матикама . . . . .	94
5.3	Структура блока изнад члана . . . . .	97
5.4	Унутрашња структура члана . . . . .	100
5.5	Подела на пасусе . . . . .	103
5.6	Синтаксни графови . . . . .	104
<b>6</b>	<b>Синтаксна анализа</b>	<b>105</b>
6.1	Лексичке етикете у синтаксној анализи . . . . .	108
6.2	Каскадна примена трансдуктора . . . . .	109

<b>7</b>	<b>Контролни документ</b>	<b>113</b>
7.1	Основна техничка и технолошка опредељења . . . . .	114
7.2	Изглед и функционалност . . . . .	115
<b>8</b>	<b>Израђени софтверски алати</b>	<b>120</b>
8.1	Претходна обрада корпуса . . . . .	121
8.1.1	Преузимање лексичког материјала са веб-сајта . . . . .	122
8.1.2	Обрада преузетог лексичког материјала . . . . .	124
8.2	Додатне статистичке анализе корпуса . . . . .	127
8.3	Статистике препознатих именованих ентитета . . . . .	130
8.4	Преклопљене и вишедокументске конкорданце . . . . .	131
8.5	Генерисање контролног документа . . . . .	132
8.5.1	Конструисање мреже упућивања . . . . .	134
8.5.2	Ажурирање <i>DOM</i> структуре . . . . .	136
8.6	Додатне могућности <i>Unitex</i> -а . . . . .	138
8.6.1	Генерисање битмапираних слика за графове . . . . .	138
8.6.2	Унапређења уређивача графова . . . . .	140
8.6.3	Остала унапређења . . . . .	140
8.7	Транскодовање графова . . . . .	142
<b>9</b>	<b>Закључак и даљи рад</b>	<b>146</b>
9.1	Даљи рад . . . . .	148
	<b>Литература</b>	<b>153</b>
<b>A</b>	<b>Синтаксни графови</b>	<b>161</b>
A.1	Графови који се као подграфови користе на разним местима . . .	161
A.2	Упућујуће фразе . . . . .	162
A.3	Фразе сличне упућујућим . . . . .	170
A.4	Логичка структура прописа . . . . .	171
A.4.1	Блок изнад члана . . . . .	171
A.4.2	Унутрашња структура члана . . . . .	173
A.4.3	Почетни део акта . . . . .	175

<b>Б CasSys формат датотеке</b>	<b>176</b>
Б.1 Садржај датотеке <i>propisi.csc</i> . . . . .	176
<b>В XSLT спецификација</b>	<b>177</b>
В.1 Садржај датотеке <i>propis.xsl</i> . . . . .	177
<b>Г Списак закона у корпусу</b>	<b>179</b>
<b>Биографија аутора</b>	<b>215</b>

# Поглавље 1

## Увод

Језик прописа (закона и подзаконских аката), као специфичан подјезик у оквиру административног функционалног стила, обично се сматра изузетно формалним сегментом језика са врло ограниченим репертоаром филолошких и синтаксичких структура. Са оваквим својствима, он потенцијално представља природно-језички сегмент подеснији за информатичке експерименте од других, мање формализованих подјезика. С једне стране, њега одликује прецизно утврђен логички изглед: текст закона је подељен у чланове, ставове, тачке, па се стиче утисак да ће чак и читалац који нема правничко образовање лако пронаћи одредбу према броју члана, става или тачке. Проблеми који се иначе јављају у аутоматској обради природно-језичког материјала могу изгледати лакше решиви у подјезику прописа због његовог (привидно) ограниченог вокабулара, а посебно, због тежње ка једнозначности, суженог скупа употребљених глагола и избегавања употребе заменица. С друге стране, и поред тога што су језичке конструкције у овом подјезику понекад сличне формалним језицима, оне су прожете слободом природно-језичког изражавања, тако да оно што на први поглед представља поједностављену структуру, заправо може унети додатан степен сложености, јер задржава наслеђену сложеност природног језика и истовремено уводи специфична методолошка правила за израду прописа.

Изградња формализованог модела овог подјезика може имати низ корисних примена како за ауторе законских текстова, тако и за најшири круг кори-

сника. Ниво развијености оваквих модела, као и укупан степен информатичке подршке за српски језик је релативно низак, за разлику од информатички боље подржаних језика [1].

У [1] је дат упоредан приказ нивоа подршке језичким технологијама за 30 европских језика према следећим критеријумима:

- Обрада говора
- Машинско превођење
- Граматичка анализа
- Ресурси

Упоредни приказ из [1] показује да се српски језик налази у групи заједно са другим језицима са малим бројем говорника као што су хрватски, словеначки и словачки и да су сви ти језици рангирани далеко иза заступљенијих и информатички боље подржаних језика као што су немачки и француски. Очекивано, по свим критеријумима се као најбоље подржан издваја енглески језик.

## **1.1 Предмет и циљ истраживања**

Предмет истраживања у овој дисертацији су посебна језичка правила у текстовима прописа на српском језику која се могу изразити методама рачунарске лингвистике.

Циљ истраживања је да се на теоријском нивоу створи основа за развој софтверских алата за обраду текстова прописа и да се испитају конкретне могућности имплементације. Посебна пажња се посвећује могућности примене у процесу израде прописа.

Истраживање обухвата анализу посебних језичких правила на изабраном корпусу прописа полазећи од званичних докумената који дају оквире за та правила, затим исказивање синтаксних правила и семантичких модела коришћењем одговарајућих формализама, на томе базирану имплементацију експерименталних софтверских решења, као и анализу резултата добијених

помоћу тих софтверских решења и анализу практичне употребљивости таквих решења.

Текст се, као предмет информатичке обраде, може сагледавати на различите начине. У зависности од степена потребног формализованог знања о језику, може се говорити о његовој природно-језичкој обради или само о обради текста [2], [3]. Ако се у обради текста у значајној мери користе ресурси који садрже формализована природно-језичка својства, онда је реч о обради природног језика [4], [5].

Области обухваћене овим терминима треба разликовати од области уређивања и форматирања текста, где се термин „апликација за обраду текста” користи у значењу апликације за уређивање и форматирање текста. За прве апликације за уређивање текста је коришћен термин *text editor*, што је уобичајена терминолошка конструкција за апликације којима се уређује неки садржај. Те апликације су имале могућност уређивања просте секвенце карактера распоређене у линије текста без додатних опција форматирања. Термин *text editor* се и даље користи за апликације за уређивање обичне текстуалне датотеке, док је за апликације које имају додатне могућности форматирања текста, када су се појавиле, скован термин *text processor* да би се нагласила разлика у односу на прве. У том случају се *processing* (обрада) односи на обраду која служи да се постигне жељени графички изглед (форматирање) текста, што је само један специфичан случај обраде текста.

Текст који је предмет обраде може бити изражен на формалном или природном језику [6]. Примери формалних језика су програмски језици и поједини формати за размену података (који су често засновани на генерализованим формалним језицима као што је *SGML*<sup>1</sup> и његов наследник *XML*<sup>2</sup>). Свака веб страна коју гледамо је описана формалним језиком *HTML*, мада већина корисника на то не мора да обраћа пажњу, јер је то за њих нешто што се дешава иза сцене. Пример формалних језика у широкој употреби које корисници непосредно користе су језици *SMS*<sup>3</sup> порука којима се захтева нека аутоматизована

<sup>1</sup>Standard Generalized Markup Language (ISO 8879:1986)

<sup>2</sup>Extensible Markup Language <http://www.w3.org/XML/>

<sup>3</sup>*SMS* је скраћеница од *Short Message Service*, што означава услугу кратких текстуалних порука у мобилној телефонији.

услуга, попут захтева за провером стања на текућем рачуну банке, који би могао да изгледа:

*ста број-рачуна ознака-валуте*

Предмет овог истраживања је обрада природног језика, мада у неким елементима језик прописа има стриктност форме која подсећа на формалне језике.

И када је текст изражен на природном језику, његова обрада може бити усмерена на откривање природно-језичких својстава, али и на техничку структуру укупног текста изражену елементима који нису непосредно повезани са природно-језичким својствима, као што су подела на поглавља или пасусе. Овај аспект организације текста називамо логичка структура или изглед текста (енгл. *logical layout*).

У оквиру овог истраживања испитују се својства оба наведена нивоа организовања текста (логичког и природно-језичког), укључујући и њихове интеракције.

У оквирима многих професија су успостављена посебна правила језичког изражавања када је у питању професионална тематика, чиме су формирани језици посебне намене или подјезици у Херисовом смислу [7], [8], а један такав пример је управо правни језик [9]. Нормативна акта одликује специфична хипертекстуална структура којом се преко елемената логичке структуре текста реферише његов природно-језички садржај, као и лексичке специфичности у односу на стандардни језик.

За енглески језик и англо-саксонску правну традицију постоји бројна литература која обрађује тему правног језика, као на пример [9]–[13]. Избор литературе која се бави језичким аспектима правних аката у правном систему Србије, укључујући бившу Југославију, је очекивано скромнији, а ту бисмо пре свега истакли Зборник радова са научног скупа „Право и језик” из 2006. године и у оквиру њега радове „Кратка историја Српског правног језика” [14], где је дат историјски приказ језичких особености значајних правних аката и периода законодавне праксе, „Језик у праву” [15], где се анализирају разни варијетети језика у праву, почевши од поделе на језик права и правнички језик, „Глосе о језику у праву” [16], где се у девет тачака (глоса) износи виђење разли-

читих аспеката правног језика и „Језик закона – карактеристике и родна перспектива” [17], где је дат систематски приказ карактеристика језика закона са посебним освртом на питање родне сензибилности. Исте године је објављен и рад „Српски правни језик” [18] у коме је дат преглед развоја српског правног језика. У књизи „Језик права” [19] објављеној 1989. године аутор разликује четири слоја језика права: први слој обухвата елементе општег језика у неизмењеном значењу, други слој обухвата елементе општег језика којима је нормативним дефиницијама проширено или сужено значење, трећи слој обухвата правне термине, а четврти стручне термине из других области.

Поред анализе карактеристичних појавних облика општих језичких својстава у корпусу прописа, ово истраживање обухвата анализу и формализацију правила за писање правних аката [20], [21]. У том погледу су посебно разматрана два званична документа: акт Законодавног одбора Народне скупштине „Јединствена методолошка правила за израду прописа” [22], који стандардизује садржај, форму, језик, стил и начин писања прописа и сличан акт Владе Републике Србије „Методологија за израду подзаконских прописа” [23].

Прописи имају прецизну логичку структуру (организовани су у главе, чланове, тачке и слично), специфична правила форматирања, посебне начине коришћења писма, начина изражавања бројева и датума, изражавања унакрсних референци (упућивања), измена и допуна и слично. Поред тога прописи имају специфичну терминологију укључујући општу правну терминологију, као и термине утврђене појединим прописима. Ове структуре у правним текстовима на српском до сада нису биле детаљно формализоване, нити информатички обрађиване.

Као пример који осликава специфичне елементе правног језика навешћемо завршну одредбу Закона о државним службеницима:

Овај закон ступа на снагу 1. јула 2006. године, изузев одредаба члана 158. ст. 1. и 3, члана 164. став 1, чл. 165-167, члана 176, члана 177. ст. 1. и 2, члана 178. и члана 188, које ступају на снагу осмог дана од дана објављивања овог закона у „Службеном гласнику Републике Србије”.

Цитирана одредба указује на сложеност задатка издвајања језичких инфор-



мација из правне формулације, као и на повезаност логичке структуре (изражене у упућивању на друге одредбе закона) са природно-језичким садржајем закона. Овакве исказе у стандардном језику срећемо изузетно ретко.

Циљ истраживања се може разложити на више нивоа:

1. опис општих језичких особености правног језика на основу припремљеног корпуса и расположивих језичких ресурса (поглавље 3);
2. прецизан опис методама рачунарске лингвистике посебних језичких правила која се користе у прописима (поглавља 4 и 5);
3. имплементација синтаксне анализе текста прописа, као почетне фазе за посебне обраде (6);
4. израда експерименталних софтверских алата заснованих на претходном, као и анализа могућности примене таквих и сличних решења (поглавље 7).

Поред наведеног су у поглављу 8 приказани додатни софтверски алати који су развијени за потребе истраживања.

## 1.2 Методологија

Обради природно-језичког текста се обично приступа на два дијаметрално супротна начина: један приступ се заснива на коришћењу статистичких метода, а други на коришћењу метода заснованих на правилима.

Употреба статистичких метода је једноставнија и присутнија у пракси [24] јер већ на основу ограниченог лингвистичког разумевања и великог скупа примера, постижу одређене резултате у аутоматској обради. Главно ограничење статистичких метода јесте да не продубуљују разумевање језичких конструкција, а разноврсност језика је таква да релативно лако долазимо до случајева који нису довољно слични постојећим примерима.

Методе засноване на правилима дају одличне резултате онда када постоје добро описана правила. Међутим, потребан скуп правила може бити изузетно

обиман и ако на располагању немамо претходно израђена правила на задовољавајућем нивоу, ресурси потребни за израду правила могу превазићи могућности пројекта у коме нам је потребна природно-језичка обрада.

Све у свему, избор методе зависи од природе проблема: методе засноване на правилима дају боље резултата онда када постоје услови да се успешно примене, док за све остале случајеве имамо на располагању статистичке методе, укључујући и могућност њихове комбинације.

На пример, када обрађујемо бројеве (записане арапским или римским цифрама или словима) или датуме, тада свакако треба да користимо методе засноване на правилима јер за описивање свих случајева није потребан превише велики број правила. Бројеви и датуми нису погодни за обраду статистичким методама, јер сваки скуп примера садржи само мали сегмент свих могућих бројева и датума.

Проблем препознавања на ком језику је одређен текст је, на пример, погодан за статистичке методе, јер је довољно посматрати погодно изабране статистичке показатеље, без дубље језичке анализе.

С обзиром на циљеве овог истраживања, који обухватају анализу специфичних језичких конструкција у правном језику, определили смо се за методе засноване на правилима, а конкретно су изабране методе засноване на примени коначних аутомата и електронских речника [25].

Коришћена је методологија локалних граматика представљених коначним трансдукторима [26], [27], односно одговарајућим графовима уз подршку електронских речника у *LADL*<sup>4</sup> формату [4], [25]. Као основни алат који подржава наведену методологију коришћен је *Unitex* [28], [29] који је такође развијен у оквиру *LADL* лабораторије.

*LADL* формат електронских речника је развијен првобитно за француски језик [30], а касније и за велики број европских језика, укључујући и српски ћириличним и латиничним писмом. У овом истраживању су коришћени претходно развијени електронски речници за српски језик [31], [32]. Више детаља о електронским речницима дато је у одељку 3.3.

---

<sup>4</sup>*LADL* је скраћеница за *Laboratoire d'Automatique Documentaire et Linguistique* која је део Универзитета *Marne-la-Vallée*, <http://ladl.univ-mlv.fr/>

## Поглавље 2

# Правни текст и природно-језичка обрада

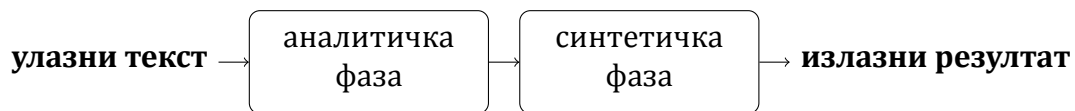
Потреба за природно-језичком обрадом правних текстова се у пракси јавља у различитим случајевима у којима се правна акта примењују, израђују или су предмет истраживања, као што је:

1. претрага правних текстова у циљу проналажења релевантних одредби прописа, примера судске праксе, елемената уговора или других типова аката;
2. истицање и издвајање битних чињеница, појмова и односа у оквиру правног текста у циљу ефикаснијег сагледавања садржаја правног акта или групе правних аката;
3. анализе укупних показатеља и односа на већем скупу аката;
4. подршка у изради правних аката у складу са обавезним и уобичајеним правилима за одговарајућу врсту акта, што укључује усмеравање у току уноса текста и проверу израђеног текста;
5. аутоматска и полуаутоматска анотација текста када је начин анотације задат у оквиру успостављеног информационог система или утврђеним стандардима за размену докумената;
6. анализа и конструкција правне аргументације;

## 7. подршка вишејезичности и превођење.

Међу наведеним групама случајева се препознају кључне подобласти обраде природног језика и вештачке интелигенције, као што су: претрага текста (доминантна у тачки 1), екстракција информација (тачка 2, 3 и 5), аутоматска коректура (тачка 4), аутоматско превођење (тачка 7) и аутоматско резоновање (тачка 6). При томе је за све случајеве заједничко да је на почетку обраде неопходна одређена природно-језичка анализа, док преостала обрада може укључивати методе из разних области да би се одговорило на захтеве конкретног случаја.

Без обзира на сврху обраде, сваки процес обраде текста садржи неке основне елементе. У процесу обраде текста се на основу улазног текста формира излазни резултат, а сам процес се може поделити на аналитичку и синтетичку фазу (слика 2.1).



Слика 2.1: Основни елементи процеса обраде текста

У аналитичкој фази је обрада усредсређена на извођење информација из улазног текста, док је у синтетичкој фази обрада усредсређена на формирање излазног резултата на основу изведених информација. На пример, ако је циљ обраде одређивање учесталости појављивања појединих врста речи, онда је у аналитичкој фази потребно препознати шта су речи у тексту и одредити врсту сваке од њих, а затим се у синтетичкој фази врши бројање речи по врстама и формира се излазни резултат где су избројане вредности наведене у прикладном формату.

У датом примеру аналитичка фаза у целини има природно-језички карактер, али је у општем случају могуће да се након природно-језичке анализе настави анализа другим методама, као на пример семантичка анализа методама вештачке интелигенције.

Природно-језичка анализа је почетна фаза сваке обраде у којој је улазни текст изражен на природном језику, под претпоставком да су природно-језичка

својства текста од значаја.

Природно-језичка обрада се може појавити и у синтетичкој фази уколико излазни резултат треба да буде на природном језику. На пример, у коришћењу разних софтверских апликација, банкомата, говорних аутомата и сл., свакодневно срећемо поруке у којима се појављује променљива бројна вредност иза које стоји именица, па ту именицу треба граматички ускладити са бројном вредношћу, што није посебно сложен проблем, али неспорно припада обради природног језика. Ипак, главни пример обраде чији је резултат природни језик је аутоматско превођење.

Дакле, ако обрада почиње природно-језичком анализом улазног текста или се завршава синтезом текста на природном језику као резултатом обраде, тада можемо говорити о природно-језичкој обради. Потреба за природним језиком увек потиче од интеракције са човеком, па је тешко замислити потребу за природно-језичком обрадом унутар неког рачунарског програма, а да се природни језик не појављује нити на улазу нити на излазу.

## **2.1 Природно-језичка анализа текста**

За почетак ћемо претпоставити да је улазни текст представљен као чиста секвенца карактера, што је најнижи стандардизован ниво апстракције у коме се текст представља у рачунару, а на крају ћемо продискутовати и случај када је изворни текст у неком другом формату (поделељак 2.1.1).

Због саме чињеница да је улазни текст најчешће на знатно нижем нивоу апстракције од нивоа на коме размишљамо о односу улаза и резултата обраде, неопходно је да у аналитичкој фази подигнемо апстракцију до нивоа на коме можемо ефикасно да сагледамо однос са жељеним резултатом.

Природно-језичка анализа обично има више корака где се у сваком кораку за по једну степеницу подиже ниво апстракције. При томе подизање нивоа апстракције обично значи придруживање додатних информација које дају могућност погледа са вишег нивоа апстракције, али без занемаривања информација на нижим нивоима апстракције које су наслеђене из претходних корака, односно из улазног текста. На пример, уколико се у тексту појави реч „столицом”,

можемо закључити да је у питању инструментал јединице столица и таква информација представља један виши ниво апстракције од просте ниске карактера који чине реч. Лему и морфосинтаксичка својства која смо одредили ћемо придружити као додатне информације, без занемаривања информација на нижем нивоу апстракције. То значи да ћемо у наредним корацима обраде и даље моћи да приступимо нисци карактера, на пример да бисмо проверили да ли је прво слово велико због препознавања именованих ентитета или да бисмо анализирали поделу на слоге због препознавања риме у песми.

Код обраде формалних језика смо слободнији да занемаримо одређене детаље са нижег нивоа апстракције јер имамо затворен скуп добро дефинисаних правила укључујући и стриктно дефинисану семантику, тако да слободно занемарујемо све што не утиче на коначно значење, па можемо једноставније да дефинишемо посебне моделе за одређене нивое апстракције. На пример, приликом интерпретације програма на неком програмском језику у одређеном кораку у аналитичкој фази обраде текста програма можемо занемарити разлику између  $(a+b)+c$  и  $a+(b+c)$ , или разлику између  $a+6+4$  и  $a+10$ , јер у оба примера имамо семантички еквивалентне случајеве.

Дакле, за природно-језичку обраду текста је карактеристично да се кораци аналитичке фазе углавном своде на придруживање додатних информација резултату дотадашње обраде, односно улазном тексту.

Први корак природно-језичке анализе је готово увек подела на лексичке јединице (токене) са идејом да појединачне речи стају у оквиру појединих токена. На тај начин је секвенца карактера трансформисана у секвенцу токена. За сваки токен се обично памти ниска карактера који га чине, па су тако све информације из улазног текста задржане.

Типични наредни кораци су сегментација реченица и придруживање морфосинтаксичких и семантичких својстава токенима који су препознати као речи у одређеном граматичком облику, за шта се обично користи морфолошки речник са придруженим одговарајућим својствима. Даље се могу препознати изрази, фразе, именовани ентитети, поједине синтаксне конструкције, специфична логичка структура текста итд.

Сваки од ових корака можемо да посматрамо као специфичну анотацију

текста, то јест придруживање додатних обележја деловима текста. Анотација се може вршити на разним нивоима, као што су библиографске јединице, елементи логичке структуре текста, реченице, делови реченице и речи.

Током природно-језичке анализе текста обично користимо одређен начин представљања аотираног текста над којим у сваком кораку вршимо додатну анотацију. Начин представљања аотираног текста и начин како се задају поједини кораци обраде зависе и од избора алата за обраду природног језика.

Као основни алат за обраду природног језика у овом истраживању користимо *Unitex*, што смо већ истакли у одељку 1.2. Начин примене система *Unitex* за природно-језичку анализу описан је у одељцима 6.1 и 6.2 на случају анализе текста закона, где видимо да систем *Unitex* користи формат који се назива етикетираним текстом, у коме су речи имплицитно аотиране путем електронског речника, за сегментацију реченица се користи ознака  $\{S\}$ , за додатно аотирање се користе лексичке етикете (изворно намењене експлицитној аотацији на нивоу речи, али се користе и шире), а природно-језичка анализа у више корака изводи се каскадом коначних трансдуктора које корисник задаје у облику синтаксних графова.

За избор алата је поред његових општих карактеристика посебно важно постојање квалитетне подршке за одређен језик, што се огледа у расположивим језичким ресурсима, као што су електронски морфолошки речници и решења за неке често потребне кораке аналитичке обраде, као што је препознавање именованих ентитета. У овом истраживању су заједно са системом *Unitex* коришћени језички ресурси за српски језик развијени у оквиру Групе за језичке технологије Универзитета у Београду [31], [32]. Добра подршка за српски језик представљала је кључну предност система *Unitex* при избору алата за обраду природног језика, од којих су неки представљени у одељку 2.4.

### **2.1.1 Изворни текст у специфичном формату**

Текст над којим желимо да вршимо обраду често нам је доступан у изворном облику који није чисти текст. То може бити формат одређеног програма за уређивање текста, веб-страница (*HTML* формат) или већ обележен текст у неком

систему и за неке потребе.

У таквим случајевима је потребно извршити претходну обраду, са циљем:

1. издвајања чистог текста са елиминасањем свих елемената формирања и додатних обележја или
2. формирања аотираног текста у облику прилагођеном алату који користи се за обраду природног језика, где се поред издвојеног чистог текста остављају и додатне информације из изворног формата које могу бити корисне у даљој обради.

Оба типа претходне обраде се једноставно интегришу у раније описан процес природно-језичке анализе.

## **2.2 Обрада након природно-језичке анализе**

Као што смо већ истакли, у свим групама случајева природно-језичке обраде правног текста које су набројане на почетку овог поглавља обрада почиње природно-језичком анализом, током које се текст аотира, а након тога се обрада наставља на различите начине. Наравно, ни фаза природно-језичке анализе није увек иста, већ се у овој фази само решавају слични проблеми. У тој фази се придружују информације које су од значаја за даљу обраду: погодне су да се по њима текст даље претражује, садрже или указују на чињенице, појмове и односе које желимо истаћи, користе се у правилима које треба да испуњава правни акт и сл.

У првој групи случајева природно-језичке обраде правног текста, где је циљ претрага текста, следећи корак може бити индексирање аотираног текста ради ефикаснијег претраживања или се може претраживати без додатног индексирања. Затим следи само претраживање где се за задати упит формира резултат претраживања, као крајњи резултат обраде. Уколико желимо да понављамо претраживање над истим текстом, али са измењеним упитом, тада не морамо понављати фазу анализе текста, јер се текст није променио.

Један од начина приказивања резултата претраге је да се поред пронађених делова текста приказује леви и десни контекст, што је форма која се често ко-



ристи у лингвистичким истраживањима и назива се конкорданцама (пример можете видети у одељку 4.8).

У другој и трећој групи случајева, који се могу сврстати у екстракцију информација, већ у фази природно-језичке анализе се настоји да се изведу информације које су што ближе крајњем циљу обраде.

У неким случајевима се главни посао екстракције информација завршава у фази природно-језичке анализе и у даљој обради преостаје да се синтетише резултат у задатом облику, а природа и сложеност тог дела обраде зависи од задатог облика крајњег резултата.

У нешто сложенијем случају је потребно додатно анализирати везе међу деловима текста, рачунати статистике и слично. Говоримо о случајевима када је за имплементацију додатне обраде потребно имплементирати одређене врсте алгоритама, али без сложене семантичке обраде, односно без примене метода вештачке интелигенције. Иза тога, као и обично, остаје потреба синтезе крајњег резултата у задатом облику.

Постоје случајеви у којима је потребно реализовати напредну семантичку обраду коришћењем посебних метода и алата, што у мањој или већој мери заглази у примену метода вештачке интелигенције. У таквом случају се за презентацију знања могу користити посебни језици за које постоје развијени алати за семантичку обраду, а један од најчешће коришћених је *OWL 2 Web Ontology Language* [33].

На основу досадашњег разматрања можемо рећи да се даља обрада аотираног текста који је резултат природно-језичке анализе може сврстати у један од три нивоа:

1. непосредна синтеза крајњег резултата;
2. додатна анализа без напредне семантичке обраде;
3. додатна напредна семантичка обрада.

Код провере израђених правних аката у природно-језичкој анализи је потребно препознати све елементе на које се односе правила, укључујући и случајеве у којима постоје синтаксне неправилности. Након тога је потреба додатна

анализа која, на пример, може обухватити проверу упућивања и нумерисања, а што све заједно обично не укључује напредну семантичку обраду. Пример из овог рада је формирање контролног документа који помаже у провери текста прописа (поглавље 7).

Када је задатак обраде специфична анотација, у неким случајевима се у природно-језичкој анализи може добити текст аотиран свим обележјима потребним за крајњу анотацију које је још евентуално потребно пресложити у задати крајњи облик. У зависности од природе информација које је потребно придружити тексту, обрада може захтевати додатну анализу, при чему се уместо напредне семантичке обраде може предвидети одређена интеракција са корисником путем које би се добиле додатне информације које се не могу довољно једноставно екстраховати из текста (тада говоримо о полуаутоматској обради).

Анализа и конструкција правне аргументације улази у област аутоматског резоновања што представља подобласт вештачке интелигенције. Веза између права и вештачке интелигенције није нова, њена историја је, према [34], дуга бар 40 година, када су Бученен и Хедрик [35] изнели могућност моделирања правног резоновања, посебно за давање правних савета, анализу закона и правну аргументацију. У међувремену је та могућност постала реалност те су коришћењем различитих техника вештачке интелигенције (доказивање теорема, експертски системи, логичко програмирање) развијени системи који решавају различита питања правне аргументације – на пример, откривање грешака у правном акту у време његовог настанка да би се пронашла правила која су супротстављена или двосмислена [36]. Неки од тих система су већ у деведесетим годинама постали оперативни и почели су се користити у локалним и државним институцијама, на пример у Холандији и Аустралији.

Аутоматско превођење представља посебну област у оквиру обраде природног језика у чији развој се улажу значајни ресурси, али још увек са ограниченим крајњим резултатима. Развој система за аутоматско превођење има још дужу историју. Према [37] прва размишљања и покушаји датирају још из педесетих година XX века. У међувремену, уз један кратак период који је уследио после објављивања познатог *ALPAC* извештаја [38] када је истраживање у

овој области на једно време спласнуло, област машинског превођења је остала веома активна што је и дало бројне позитивне резултате. Приступи које се примењују су такође бројни и, у основи, се деле на системе засноване на правилима, на коришћењу статистичких законитости и на хибридне системе који интегришу оба приступа. Премда се сви слажу да су ови системи још увек далеко од савршених, њихова слободна доступност, посебно кроз машине за претрагу (*Google* и *Microsoft*), је показала колико такви системи могу бити корисни, што значи да сви до сада уложени и будући напори у правцу развоја система за аутоматско превођење имају смисла. То посебно важи за вишејезичну Европу где број језичких парова за које треба обезбедити превод на дневној основи (парламентарна заседања, записници, и сл.) драматично расте [39]. Овако велики број језичких парова удружен с такође значајним бројем различитих система захтева прецизне мере за евалуацију и поређење, каква је на пример *BLEU* метрика [40].

Ова важна област обраде природних језика налази своју примену и у обради правних текстова, на пример за превод упита претраживања вишејезичних база текстова [41]. У области аутоматског превођења правних текстова посебно је значајан пример потребе Европске комисије да аутоматизује превођење између 24 званична језика Европске уније. Средином 2010. године је започет пројекат *Machine Translation Service by the European Commission*<sup>1</sup> у оквиру кога се имплементира нов систем за аутоматско превођење за потребе институција Европске уније и државних администрација земаља чланица. Завршетак пројекта се очекује током 2014. године, а од 2013. године је у фази пробне употребе. Овај пројекат користи систем *Moses*[42] као технолошку основу за аутоматско превођење статистичким методама.

Као и за све друге специјализоване домене, предуслов за развој аутоматског превођења правних текстова је постојање вишејезичних корпуса у том домену [43].

---

<sup>1</sup>[http://ec.europa.eu/isa/actions/02-interoperability-architecture/2-8action\\_en.htm](http://ec.europa.eu/isa/actions/02-interoperability-architecture/2-8action_en.htm)

## 2.3 Друга истраживања о природно-језичкој обради правних текстова

Истраживање и развој у области примене језичких технологија у обради правних текстова мотивисани су потребама у адвокатури, у раду судија, у процесу израде, доношења и објављивања прописа, приликом извршавања закона од стране надлежних органа, али и у свим другим пословима у којима је потребно посебно водити рачуна о правним питањима.

Потреба за природно-језичком обрадом правних текстова је нарочито изражена у прецедентним правним системима попут америчког, канадског и британског, где су извор права и све претходне одлуке судова (енгл. *case law*), што не значи да анализа судске праксе није од значаја и у земљама са другачијим правним системима.

Потреба проналажења и ефикасног сагледавања потенцијално релевантних предмета из судске праксе била је покретач истраживања у области природно-језичке обраде правних текстова и још увек спада међу најчешће циљеве истраживања у овој области.

Један од првих система који је за потребе претраге имплементирао екстракцију информација из правних аката је *FLEXICON (Fast Legal EXpert Information CONsultant)* који је развијен крајем осамдесетих година у оквиру Правног факултета Универзитета у Британској Колумбији (Канада) [44]. Систем је био заснован на анализи појављивања кључних речи из речника правних израза.

Поменућемо и систем *SALOMON* који је развијен средином деведесетих година и који аутоматски формира сажетке белгијских кривичних предмета користећи статистичке методе. *SALOMON* је развијен у оквиру Међународног центра за право и информационе технологије Католичког универзитета у Лувену (Белгија) [45]. Иста истраживачка група је учествовала у још неким пројектима са сличном темом: пројекат *MOSAIC* који је се бавио истраживањем модела проналажења судских одлука на основу природно-језичких информација у структурираним и неструктурираним текстовима [46] и пројекат *ACILA* који се бави детекцијом аргументације у разним врстама текстова, а пре свега у правним текстовима [47].

Међу новијим резултатима у овој области издвојићемо рад *An Automatic System for Summarization and Information Extraction of Legal Information* [41]. У раду је описан систем који омогућава екстракцију и сажети приказ информација из пресуда канадских судова у области имиграционог права, интелектуалне својине и пореза. Аутоматски сажетак пресуде се сачињава од реченица које су у целини преузете из пресуде, за разлику од ручно израђеног сажетка који може да садржи и преформулисане реченице. Систем обухвата и аутоматско превођење екстрахованих информација са енглеског на француски или обрнуто, што је потребно јер су оба језика у званичној употреби у Канади. Систем такође омогућава претрагу како метаподатака и сажетака, тако и целокупних текстова пресуде, при чему се на упит у претрази примењује аутоматско превођење, тако да се претрага врши и по француском и по енглеском тексту.

Основна технолошка и методолошка опредељења у истраживању из овог рада имају сличности са решењем описаним у претходно наведеном раду, а посебно у делу који је обрађен у поглављима 6 и 7 овог рада. У оба случаја се за природно-језичку анализу користи *Unitex*, након чега се у једној фази обраде користи *XSLT*<sup>2</sup> технологија и на крају се обрада врши у програмском језику опште намене, што је у наведеном раду *C#*, а у овом истраживању *Java*, чиме се у оба случаја крајњи резултат производи у *HTML* формату. Једна од разлика је да се у наведеном раду аотира обележјима у *XML* облику, док се у овом истраживању аотација врши лексичким етикетама које се касније преводе у *XML*. Наравно, у два рада се решавају различити проблеми и ово истраживање је знатно мањег обима, тако да у [41] има још пуно других технолошких детаља, укључујући и посебан систем за аутоматско превођење статистичким методама између француског и енглеског језика.

Од радова у којима срећемо сличне проблеме онима које решавамо у овом раду, за почетак ћемо истаћи рад *Automated detection of reference structures in law* [48] где се детаљно описује структура упућивања у холандским правним актима, слично као што смо у овом раду описали упућивања у српским прописима у поглављу 4. При томе смо се ми више бавили упућивањима унутар правног

---

<sup>2</sup>*Extensible Stylesheet Language Transformations*, <http://www.w3.org/Style/XSL/>

акта, док су се холандске колеге више бавиле упућивањем на друга акта, али се могу препознати и одређене сличности у приступу. На пример, холандски рад препознаје комплетна и некомплетна упућивања (*complete and incomplete references*), што одговара затвореним и отвореним упућивањима у овом раду.

Други пример рада који се бави препознавањем упућивања је *A Rule-based Parsing Approach for Detecting Case Law References in Italian Court Decisions* [49]. С обзиром да у том раду у питању упућивање на предмете из судске праксе структура упућивања је нешто другачија од упућивања на одредбе прописа.

## 2.4 Расположиви алати за обраду природног језика

Као и у многим другим областима, у почетку су за потребе истраживања у области природног језика развијани засебни софтверски системи који су се користили у појединим истраживањима или у истраживањима одређене групе аутора или институције. Временом су се неки од тих система развили у алате који су доступни ширем кругу истраживача и крајњих корисника. Са друге стране је у новим истраживањима потребно на брз и ефикасан начин имплементирати делове система који се заснивају на добро познатим решењима, како би се тежиште активности пренело на оригиналне аспекте. Због тога се у истраживањима у области обраде природног језика све више користе алати који представљају технолошку платформу за специфична решења у тим истраживањима. Један од таквих алата је већ више пута поменут *Unitex* који користимо и у овом раду, а на овом месту ћемо дати кратак приказ још неких расположивих алата. Сви овде наведени алати се дистрибуирају као софтвер отвореног кода.

Прво ћемо представити *GATE*<sup>3</sup> (*General Architecture for Text Engineering*) који се развија под окриљем Универзитета у Шефилду (Енглеска). *GATE* са једне стране представља алат са богатим могућностима у обради природног језика, а са друге стране, од чега и потиче његов назив, представља платформу која им-

---

<sup>3</sup><http://gate.ac.uk>

плементира модуларну архитектуру са неколико добро дефинисаних типова компоненти као што су компоненте за приступ језичким ресурсима, за обраду, за индексирање и слично. Одређен број компоненти је саставни део система *GATE* што му даје могућност да се користи као готов алат, а такође постоји могућност коришћења независних компоненти или израде сопствених. У процесу обраде текста централно место има аотиран текст а сваки корак обраде се обавља уз помоћ неке од компонента за обраду (*Processing Resource*), док се спецификација целог процеса обраде назива апликацијом.

*GATE* је имплементиран на програмском језику *Java*, а модуларна архитектура се заснива на добро дефинисаним апликативним интерфејсима (*API*) у *Java*-и. Интеграција са системом *GATE* је могућа за готово све познатије алате за обраду природног језика који су имплементирани на програмском језику *Java* као што је *OpenNLP* или *Stanford NLP*, а *GATE* је такође прилагођен могућности да у целини буде интегрисан као део друге апликације. Захваљујући флексибилности и могућностима интеграције, као и значајном скупу могућности за обраду природног језика, *GATE* је прилично заступњен у решењима која се користе у пракси. За систем *GATE* постоји подршка за енглески, шпански, кинески, арапски, бугарски, француски, немачки, хинди, италијански, себуано, румунски и руски, са различитим нивоом подржаних функционалности за разне језике.

Систем *UIMA*<sup>4</sup> (*Unstructured Information Management Architecture*) се развија под окриљем Софтверске фондације Апач (*The Apache Software Foundation*), а раније је развијан у оквиру компаније *IBM*. Систем *UIMA* се у најкраћем може описати као имплементација модуларне архитектуре попут оне код система *GATE*. Систем имплементира истоимени *OASIS* стандард,<sup>5</sup> а поред *Java*-е подржава компоненте у *C++* програмском језику. Предности система *UIMA* су заснованост на стандарду усвојеном од стране респективне независне организације и могућност интеграције са алатима који су имплементирани на *C* или *C++* програмском језику. Са друге стране систем *UIMA* представља само инфра-

---

<sup>4</sup><http://uima.apache.org/>

<sup>5</sup>*OASIS Unstructured Information Management Architecture (UIMA)* <http://www.oasis-open.org/committees/uima>

структурни оквир у комбинацији са којим треба користити алате за поједине задатке обраде природног језика, а који имају могућност интеграције са овим системом. На првом месту међу таквим алатима је *OpenNLP* који је такође пројекат Софтверске фондације Апач.

Систем *OpenNLP*<sup>6</sup> подржава већину уобичајених задатака обраде природног језика као што је токенизација, морфосинтаксичко обележавање, препознавање именованих ентитета, синтаксна анализа и слично. Систем такође подржава одређене методе машинског учења. Систем је развијен на програмском језику *Java*, а постоји подршка за енглески, немачки, холандски, шпански, португалски и самијски језик, са различитим скупом функционалности које су подржане за поједине језике.

Систем познат као *Stanford NLP*<sup>7</sup> је развијен у оквиру Групе за обраду природног језика Универзитета Стенфорд (САД). Развијен је у програмском језику *Java* и подржава бројне уобичајене задатке у обради природног језика, а од језика су подржани само енглески и кинески.

На крају ћемо поменути и систем *NLTK*<sup>8</sup> (*Natural Language Toolkit*) који је имплементиран у програмском језику Пајтон (*Python*) и одликује га једноставност са којом можете да почнете са прављењем програма за обраду природног језика, те не само да је погодан на учење већ има и практичну вредност, због чега је присутан у многим пројектима. Систем *NLTK* подржава енглески језик, док је подршка за друге језике ограничена.

---

<sup>6</sup><http://opennlp.apache.org>

<sup>7</sup><http://www-nlp.stanford.edu/software/>

<sup>8</sup><http://www.nltk.org/>



## Поглавље 3

### Корпус

У лингвистици се под корпусом, у најширем смислу, подразумева лингвистички материјал намењен истраживању језика [50]. За потребе овог истраживања је формиран корпус прописа сачињен од закона Републике Србије. С обзиром да је овај корпус намењен специфичном лингвистичком истраживању, он се сврстава у специјализоване корпуре, за разлику од општих корпуса који се формирају да би представљали основу за произвољно будуће истраживање [51, стр. 24].

Корпус се састоји од законских текстова у електронском облику који су доступни на веб-сајту Народне скупштине Републике Србије,<sup>1</sup> (слика 3.1). У оквиру аката које је донела и на свом веб-сајту објавила Народна скупштина, извршен је избор за корпус на следећи начин:

1. изабрани су само текстови закона, с обзиром да Народна скупштина поред закона доноси и друга акта (одлуке, резолуције, итд);
2. изостављени су закони о ратификацији (потврђивању) међународних споразума зато што обично садрже текст споразума на страном језику, а и сам текст међународног споразума не прати обавезно домаћа методолошка правила за израду прописа;
3. изостављени су закони који садрже обимне табеле, попут закона о буџету;

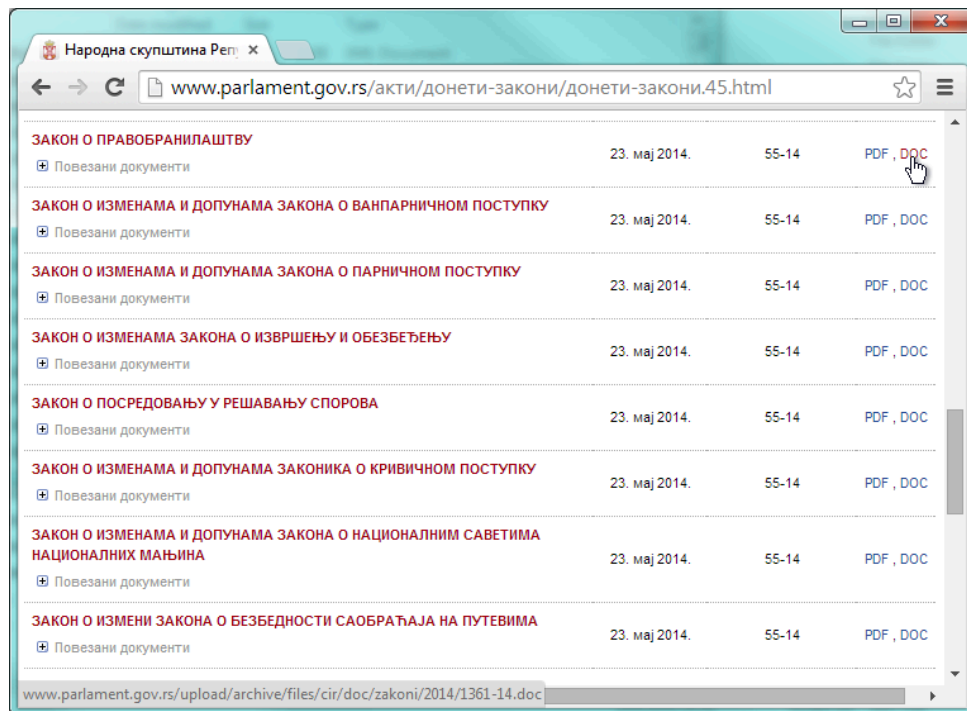
---

<sup>1</sup><http://www.parlament.gov.rs/>

4. изостављени су текстови закона код којих су постојали технички проблеми, попут тога да датотека на коју упућује хиперлинк није присутна, није у одговарајућем формату или кодном распореду;
5. изабрани су закони донети 2005. године и новији, јер се код текстова закона од пре 2005. године у већој мери појављују технички проблеми.

Корпус приказан у овом раду обухвата законе који су донети до краја маја 2014. године и садржи 681 законски текст.

Определили смо се да корпус прописа заснујемо на законима зато што се, због природе процеса израде и доношења закона, у законима доследније примењују методолошка правила него што је то случај са подзаконским актима.



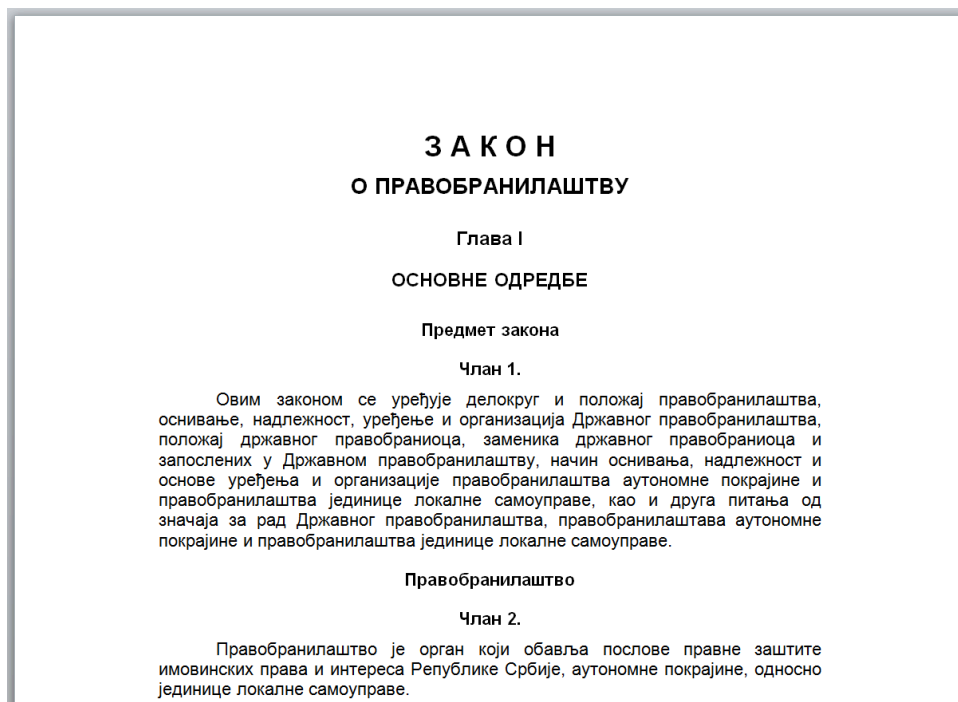
Слика 3.1: Веб-страница Народне скупштине са донетим законима

Корпус неселективно обухвата вишегодишњи период савремене законодавне праксе, са изузетком одређених екстрема. Изостављање појединих законских текстова суштински не ремети репрезентативност узорка, осим уколико се посматрају показатељи који су у непосредној вези са критеријумом изостављања. Пример таквог показатеља би могло бити присуство непознатих речи које су значајно присутније у међународним споразумима, али управо

зато нисмо ни укључили међународне споразуме јер речи у тексту споразума на страном језику нису предмет овог истраживања.

### 3.1 Претходна обрада корпуса

Претходна обрада корпуса представља низ радњи којима се лингвистички материјал из изворног облика у коме је доступан преводи у облик који је погодан за претраге и анализе коришћењем одговарајућих софтверских алата. За претраге и анализе корпуса у овом истраживању користимо *Unitex* [28], [29], о коме је било речи у одељку 1.2.



Слика 3.2: Текст закона отворен у програму *Microsoft Word*

У нашем случају изворни лингвистички материјал чине текстови закона објављени на веб-сајту Народне скупштине. На слици 3.1 је приказана веб-страница са делом листе донетих закона. За сваки закон постоји могућност да се датотека са текстом закона преузме у *PDF*<sup>2</sup> или *DOC*<sup>3</sup> формату. Међу законима који

<sup>2</sup>Portable Document Format <http://www.adobe.com/pdf/>

<sup>3</sup>Microsoft Word Binary File Format <http://msdn.microsoft.com/en-us/library/office/cc313153.aspx>

су излистани на слици 3.1 при врху можемо приметити Закон о правобранилаштву. Ако са веб-сајта преузмемо одговарајућу датотеку у *DOC* формату (датотека *1361-14.doc*) и отворимо је у програму *Microsoft Word*, видећемо текст у облику који је приказан на слици 3.2.

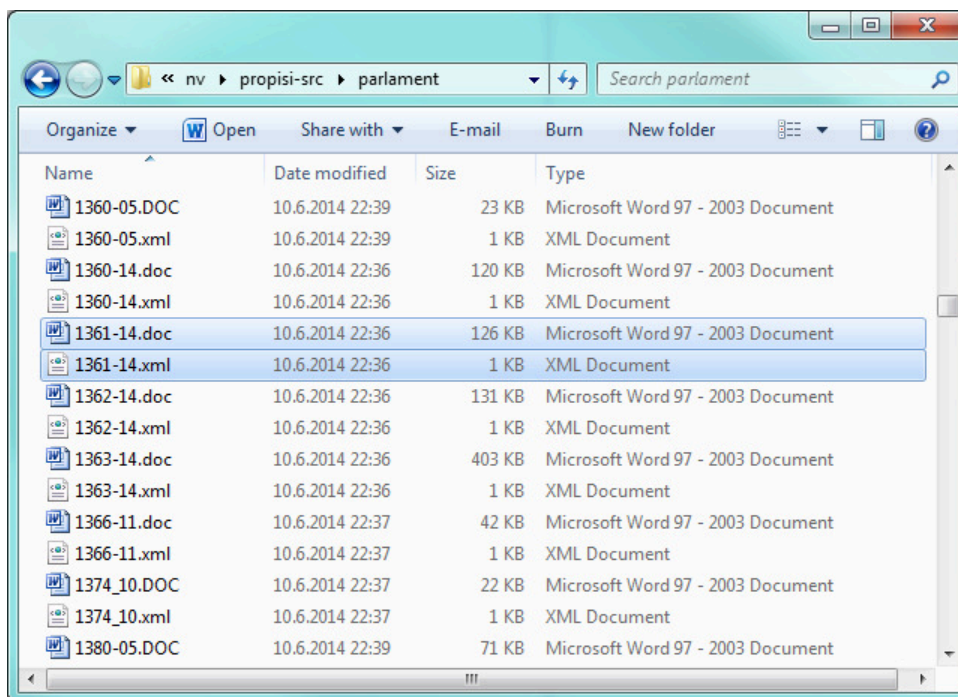
У овом истраживању је спроведена претходна обрада корпуса прописа која обухвата следеће кораке:

1. преузимање са веб-сајта датотека у *DOC* формату, као и додатних података о закону (пре свега наслова који су приказани на веб-страни);
2. конверзија из *DOC* формата у чист текст;
3. спајање текстова закона у једну датотеку у формату чистог текста уз анотацију на библиографском нивоу (означавање шта припада ком закону);
4. претходна обрада у оквиру *Unitex*-а.

За прва три корака претходне обраде је коришћен софтверски алат посебно израђен за ту намену. О детаљима самог алата ће бити више речи у одељку 8.1, а у наставку ћемо додатно описати сврху сваког корака у претходној обради.

### **3.1.1 Преузимање лексичког материјала са веб-сајта**

У овом кораку претходне обраде се датотеке са текстовима закона преузимају са веб-сајта Народне скупштине и смештају у директоријум на локалном рачунару. Додатни подаци о закону, који су приказани на одговарајућој веб-страни, смештају се у посебну датотеку у *XML* формату.



Слика 3.3: Лексички материјал преузет са веб-сајта Народне скупштине

На слици 3.3 је приказан садржај директоријума са преузетим лексичким материјалом. Две истакнуте датотеке одговарају Закону о правобранилаштву. Садржај датотеке *1361-14.doc* је приказан на слици 3.2, док је садржај датотеке *1361-14.xml*, која представља додатне податке о закону, приказан на слици 3.4. Од додатних података чувају се наслов, број службеног гласника у коме је закон објављен, датум доношења закона и основа назива датотеке.

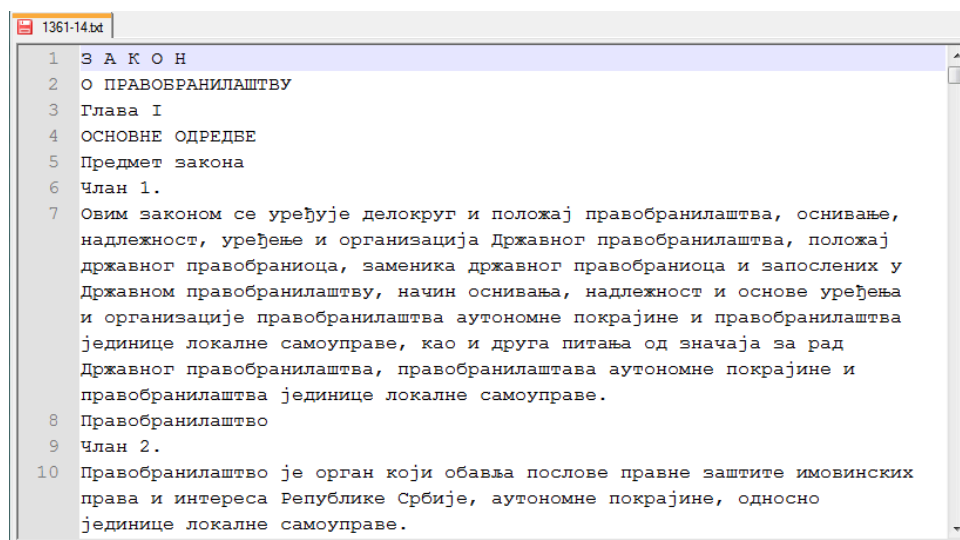
```
<?xml version="1.0" encoding="UTF-8"?>
<law>
  <title>Закон о правобранилаштву</title>
  <gazete>55-14</gazete>
  <date>23. мај 2014.</date>
  <fileNameBase>1361-14</fileNameBase>
</law>
```

Слика 3.4: Додатни подаци о закону (приказ XML-а у Internet Explorer-у)

### 3.1.2 Конверзија у чист текст

Када датотеку приказану на слици 3.2 конвертујемо у чисти текст и отворимо у програму за уређивање чистог текста, добијамо приказ попут оног на

слици 3.5.



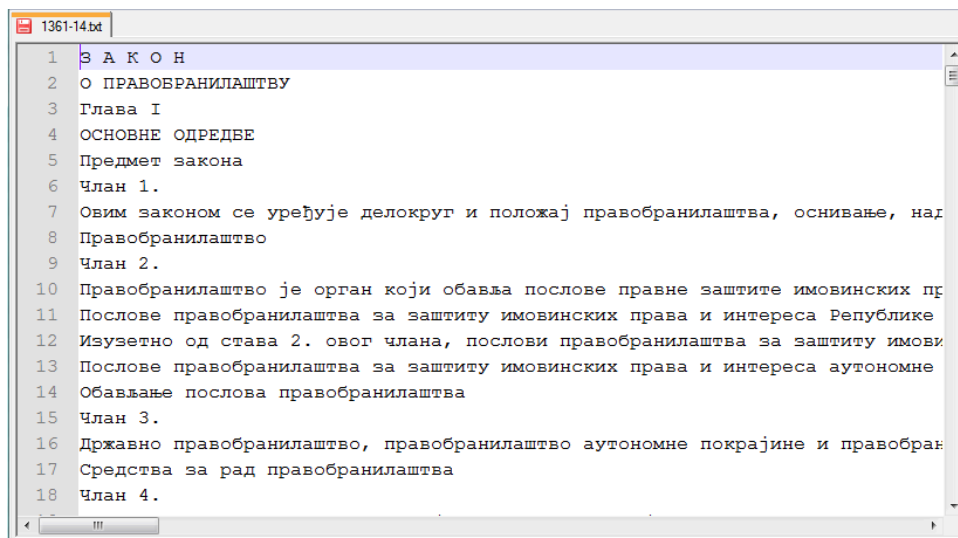
Слика 3.5: Закон у облику чистог текста (приказ у програму *Notepad++*)

Датотека у формату чистог текста (коју називамо и текстуалном датотеком) је проста секвенца карактера без додатних одређења изгледа појединих карактера, начина формирања пасуса и слично. Подела на редове одређена је карактером који означава прелазак у нови ред.<sup>4</sup> На слици 3.5 са леве стране су приказани редни бројеви редова текстуалне датотеке. Први (и једини) став члана 1. је заправо један ред текстуалне датотеке (седми), али је, због дужине, приказан преломљено. Дакле, на слици 3.5 треба разликовати преласке у нови ред који потичу од карактера за нови ред садржаних у текстуалној датотеци (када следи нови редни број са леве стране) од прелазака у нови ред који су додати да би се текст приказао у ширини прозора програма за уређивање текста, то јест, када се на месту где је приказан прелазак у нови ред не налази карактер за нови ред.

Уколико у програму за уређивање текста (у овом случају *Notepad++*) искључимо приказивање дугих редова као преломљених, добићемо изглед као на слици 3.6. У том случају се сваки ред текстуалне датотеке приказује у једном реду (без преламања), али се зато дужи редови приказују у ширини знатно ве-

<sup>4</sup>Карактер за нови ред нема своју графичку репрезентацију као што је имају карактери који представљају слова, цифре и знаке interpunkције, већ се у приказу текста карактер за нови ред манифестује преласком у нови ред.

ћој од ширине прозора, па се у прозору види само део таквих редова, због чега је на дну прозора укључена трака за хоризонтално померање (хоризонтални клизач), да би корисник могао да помера приказ садржаја и у хоризонталном правцу.



Слика 3.6: Приказ чистог текста без додатног преламања других редова

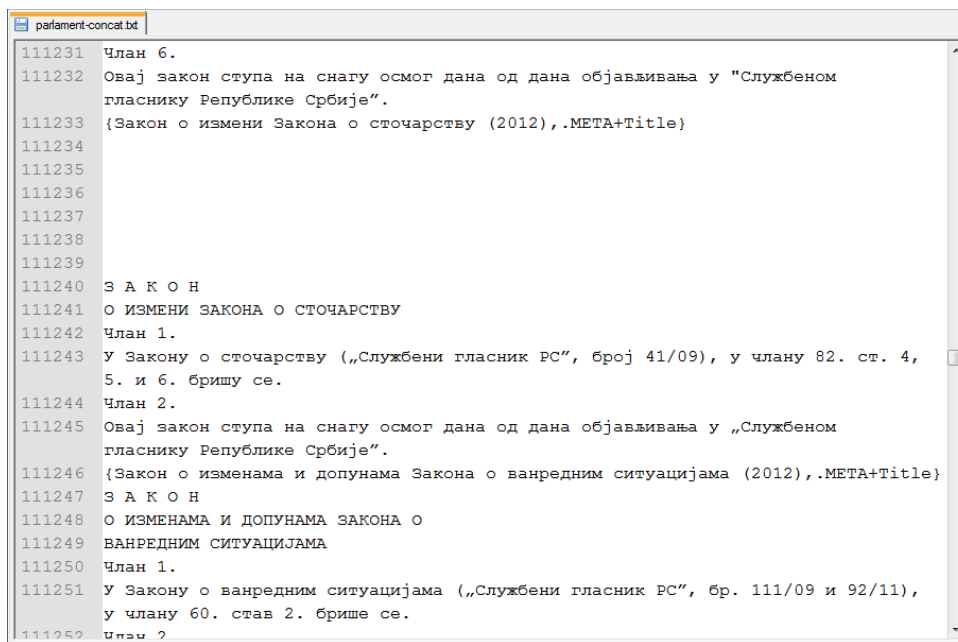
Дакле, приликом конверзије у чист текст нисмо вршили преламање других редова, што значи да је сваки пасус из *DOC* формата конвертован у један ред текстуалне датотеке. Ово је логичан избор, с обзиром да преламање текста у циљу приказа у одређеној ширини ступца треба решавати приликом формирања текста за потребе жељеног начина приказа, а не у самом садржају текстуалне датотеке.

### 3.1.3 Спајање и библиографска анотација

Обрада корпуса путем система *Unitex* претпоставља да цео корпус, односно део корпуса који се у једном тренутку обрађује, смештен у једну текстуалну датотеку. Због тога смо текстове свих закона спојили у једну велику текстуалну датотеку (величине око 40 мегабајта).

Ако бисмо текстове већег броја закона само спојили такве какви су изгубили бисмо податак о томе шта припада ком закону. Можда бисмо такав податак и могли касније да реконструирамо језичком анализом са одређеном веро-

ватноћом тачности, али је једноставније и логичније да тај податак оставимо сачуван у експлицитном облику. Стога испред текста сваког појединачног закона додајемо анотацију (посебну ознаку) у којој је садржан податак о називу закона чији текст на том месту почиње. Резултат спајања и анотације видимо на слици 3.7.



```
111231 Члан 6.
111232 Овај закон ступа на снагу осмог дана од дана објављивања у "Службеном
гласнику Републике Србије".
111233 {Закон о измени Закона о сточарству (2012),.META+Title}
111234
111235
111236
111237
111238
111239
111240 З А К О Н
111241 О ИЗМЕНИ ЗАКОНА О СТОЧАРСТВУ
111242 Члан 1.
111243 У Закону о сточарству („Службени гласник РС“, број 41/09), у члану 82. ст. 4,
5. и 6. бришу се.
111244 Члан 2.
111245 Овај закон ступа на снагу осмог дана од дана објављивања у „Службеном
гласнику Републике Србије“.
111246 {Закон о изменама и допунама Закона о ванредним ситуацијама (2012),.META+Title}
111247 З А К О Н
111248 О ИЗМЕНАМА И ДОПУНАМА ЗАКОНА О
111249 ВАНРЕДНИМ СИТУАЦИЈАМА
111250 Члан 1.
111251 У Закону о ванредним ситуацијама („Службени гласник РС“, бр. 111/09 и 92/11),
у члану 60. став 2. брише се.
111252 Члан 2.
```

Слика 3.7: Спојен корпус са библиографским анотацијама

На слици се могу приметити две анотације:

- {Закон о измени Закона о сточарству (2012),.META+Title} и
- {Закон о изменама и допунама Закона о ванредним ситуацијама (2012),.META+Title}.

Анотације су у форми *Unitex*-ових лексичких етикета. О самим лексичким етикетама ће бити нешто више речи касније у раду (одељак 6.1), а овде ћемо истаћи да смо форму лексичких етикета употребили да обележимо почетак текста закона и при томе специфицирамо назив закона и годину доношења. *Unitex* овакву лексичку етикету третира као целовиту лексичку јединицу у којој препознаје ознаке *META* и *Title*, али садржај лево од тих ознака *Unitex* не анализира.

За анотацију се користе додатни подаци о закону који су у претходним корацима предобrade сачувани у одговарајућим *XML* датотекама (слика 3.3 и 3.4).



Неколико празних редова испод прве лексичке етикете значи да се у изворној *DOC* датотеци на том месту налази неколико празних пасуса за које можемо претпоставити да су уметнути ради визуелног позиционирања назива закона на првој страни, а садржајно немају функцију. Елиминисање узастопних размака, укључујући карактере за нови ред, је предмет наредног корака претходне обраде.

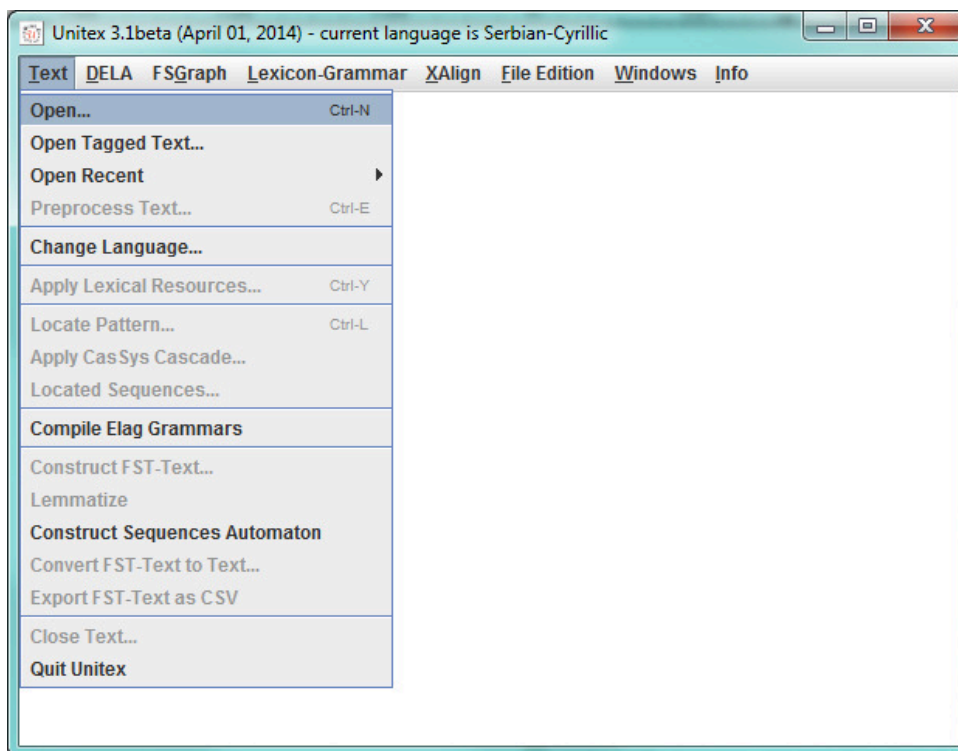
У претходној обради смо анотацију вршили само на библиографском нивоу (главне јединице текста), док анотацију на нижим нивоима, као што су логичке целине у оквиру текста или лексичке јединице (токени), нисмо вршили у оквиру претходне обраде корпуса. Унутар појединачног закона је остављен чист текст чија је даља анализа предмет овог истраживања, укључујући и препознавање логичких јединица (поглавље 5), при чему та даља анализа не припада претходној обради корпуса.

### **3.1.4 Претходна обрада у оквиру *Unitex*-а**

Текст са слике 3.7 се даље обрађује уз помоћ *Unitex*-а, при чему и у оквиру *Unitex*-а постоји корак претходне обрада.

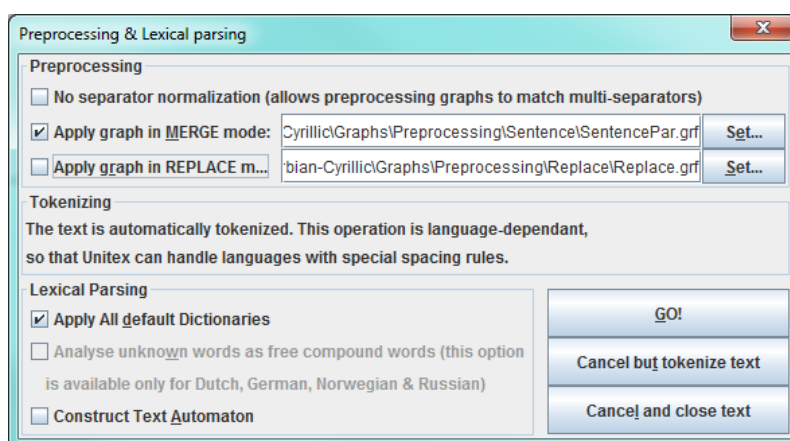
Када у *Unitex*-у желимо да обрађујемо нови текст, он треба да је претходно припремљен у формату чистог текста у складу са *Unicode* стандардом за кодирање карактера [52] у 16-битној (*UTF-16*) или 8-битној (*UTF-8*) форми кодирања. У овом истраживању смо се определили за коришћење *UTF-8* форме кодирања.

У оквиру *Unitex*-овог графичког корисничког интерфејса обраду новог текста започињемо тако што у главном менију изаберемо подмени *Text* и затим опцију *Open* (слика 3.8).



Слика 3.8: Отварање новог текста за обраду

Након што изаберемо одговарајућу текстуалну датотеку (у нашем случају то је датотека са слике 3.7 која носи назив *parlament-concat.txt*) и потврдимо да желимо да се изврши претходна обрада (што је подразумеван случај), отвориће нам се дијалог са опцијама за претходну обраду (слика 3.9).



Слика 3.9: Дијалог са опцијама за претходну обраду

Претходна обрада у оквиру *Unitex*-а обухвата:

1. нормализацију сепаратора, што значи елиминисање узастопних размака, односно прелазака у нови ред;
2. супституцију уобичајених скраћеница попут „тј.“, „итд.“ и „сл.“ њиховим пуним облицима „то јест“, „и тако даље“ и „слично“ (ову операцију смо искључили у нашем случају јер желимо да задржимо аутентичан изглед текста и у тим детаљима);
3. сегментацију на реченице уметањем посебног токена  $\{S\}$  на крај сваке реченице;
4. токенизацију (поделу на лексичке јединице);
5. примену електронских речника.

Како полазна датотека носи назив *parlament-concat.txt*, *Unitex* ће резултат претходне обраде сместити у нову датотеку *parlament-concat.snt* и поддиректоријум *parlament-concat\_snt* са додатним скупом датотека.

Датотека *parlament-concat.snt* је добијена применом прва три корака претходне обраде са претходно наведене листе, док ће преостала два корака формирати додатне датотеке у поддиректоријуму *parlament-concat\_snt*. На слици 3.10 видимо како изгледа датотека *parlament-concat.snt* отворена у оквиру *Unitex*-а.

```

parlament-concat.snt (C:\Users\nebojsa\Documents\unitex\Serbian-Cyrillic\Corpus\inv)
205348 sentence delimiters, 7981446 (62292 diff) tokens, 3290405 (61558) simple forms, 364662 (10) digits
3260507 occurrences (190878 DLF entries) simple words, 4373 occurrences (9515 DLC entries) compound words, 29061 occurrences (5765 ERR lines) unknown words
Члан 6. {S}
Овај закон ступа на снагу осмог дана од дана објављивања у "Службеном гласнику Републике
Србије". {S}
{Закон о измени Закона о сточарству (2012), .МЕТА+Title}{S}
З А К О Н {S}
О ИЗМЕНИ ЗАКОНА О СТОЧАРСТВУ {S}
Члан 1. {S}
У Закону о сточарству („Службени гласник РС“, број 41/09), у члану 82. ст. 4, 5. и 6. бришу
се. {S}
Члан 2. {S}
Овај закон ступа на снагу осмог дана од дана објављивања у „Службеном гласнику Републике
Србије“. {S}
{Закон о изменама и допунама Закона о ванредним ситуацијама (2012), .МЕТА+Title}{S}
З А К О Н {S}
О ИЗМЕНАМА И ДОПУНАМА ЗАКОНА О {S}
ВАНРЕДНИМ СИТУАЦИЈАМА {S}
Члан 1. {S}
У Закону о ванредним ситуацијама („Службени гласник РС“, бр. 111/09 и 92/11), у члану 60.
став 2. брише се. {S}
Члан 2. {S}
Члан 61. брише се. {S}

```

Слика 3.10: Изглед текста после претходне обраде

С обзиром да је на слици 3.10 приказано исто место у тексту као на слици 3.7, приликом поређења се може јасно уочити нормализација сепаратора (изостанак узастопних прелазака у нови ред) и сегментација реченица (уметнуто  $\{S\}$ ).

Сегментација реченица се врши применом коначних трансдуктора које ко-рисник може мењати, чиме правила за сегментацију реченица прилагођава специфичним потребама. У овом истраживању смо користили измењена пра-вила за сегментацију реченица тако да се  $\{S\}$  налази на крају сваког пасуса, односно реда у текстуалној датотеци, што значи да се уместо сегментације ре-ченица заправо врши сегментација пасуса. Разлог измене правила за сегмен-тацију реченица је специфична структура текста прописа, што је детаљније аргументовано у одељку 5.5.

Након сегментације реченица, *Unitex* врши токенизацију, односно поделу на лексичке јединице. Токени (лексичке јединице) могу бити:

- секвенца узастопних слова;
- појединачни карактери (све што нису слова нити витичасте заграде се токенизује као појединачни карактер);
- посебне врсте токена који су оивичени витичастим заградама, а што укључује ознаку  $\{S\}$  за крај реченице и лексичке етикете (попут оних које користимо за анотацију).

Токенизован текст је представљен помоћу неколико датотека унутар под-директоријума *parlament-concat\_snt*.

Примена електронских речника се своди на формирање извода из свих електронских речника које желимо да применимо, при чему се издвајају само одреднице које се појављују у датом тексту и на тај начин се добија изведен речник датог текста, који се смешта у оквиру поддиректоријума *parlament-concat\_snt*. Приликом даље обраде се више не приступа директно примењеним речницима, већ само изведеном речнику датог текста. Можемо сматрати да је је применом електронских речника извршена анотација на нивоу речи тако што су речима придружене додатне информације из електронског речника,

само што та анотација није експлицитно изражена у основној текстуалној датотеци (у овом случају *parlament-concat.snt*), већ је то учињено додавањем речника датог текста. О електронским речницима ће бити више речи у одељку 3.3.

## 3.2 Основни статистички показатељи корпуса

*Unitex* у оквиру претходне обраде израчунава и одређене статистичке показатеље. На слици 3.10 можемо приметити да су при врху прозора записани неки од тих бројева. Дакле, овај корпус садржи 7.981.446 токена, међу којима има 62.292 различитих. Од тога је 3.290.405 појављивања секвенци слова (61.558 различитих) и 364.662 појављивања цифара (међу којима се појављује свих 10 цифара).

*Unitex* сваку цифру токенизује посебно, тако да из основне статистике не видимо колико има бројева (секвенци узастопних цифара). То се, наравно, може одредити другим методама (функција претраживања у *Unitex*-у): бројеви се појављују 221.930 пута, међу којима је 1018 различитих. Највећи број у ознаци члана је 608 у Законику о кривичном поступку из 2011. године, што значи да се међу 1018 различитих бројева већина њих појављује у ознакама чланова.

Једнокарактерски токени који нису ни слова ни цифре ни размаци појављују се 629.588 пута, а међу њима је 56 различитих. Ту спадају знаци интерпункције, али и разни други карактери који се појављују у тексту, попут знака за проценат, за једнакост, рачунске операције, а с обзиром да се користи *Unicode* скуп карактера, појављују се и карактери за промил, грчко слово омега, ознака за евро, карактери који представљају цифре у експоненту (код квадратних и кубних метара), слова са означеним акцентом (као у примеру „одбије да дâ сагласност”), итд.

Скуп карактера који може да се појави у тексту закона није специфициран званичним методолошким правилима, нити је на други начин стандардизован. Како је Законом о објављивању закона и других прописа и аката (Сл. гласник РС бр. 45/13) прописано да електронски облик Службеног гласника постаје званично издање Службеног гласника од 1. јануара 2014. године (члан 37),

као и да се електронски облик издаје у *PDF* и *HTML* формату, при чему је званично електронско издање у *PDF* формату (члан 26), то специфична питања језичке стандардизације која се отварају коришћењем електронског облика постају додатно актуелна, а међу тим питањима је и стандардизација скупа знакова који се користе у језичком изражавању до нивоа прецизног одређења у односу на *Unicode* скуп карактера.

Ради једноставности, секвенце узастопних слова ћемо у статистикама звати речима (подразумева се да то нису увек речи српског језика), секвенце узастопних цифара бројевима, размаком ћемо звати и карактер за размак и карактер за нови ред (карактер за табулацију се приликом нормализације сепаратора претвара у карактер за размак или се брише), а једнокарактерске токене који нису ни слова, ни цифре ни размаци зваћемо знаковима.

Списак закона у корпусу са основним статистичким показатељима за текст сваког закона дат је у прилогу Г.

Најдужи закон у корпусу је Законик о кривичном поступку из 2006. године који садржи 79.003 речи, 4.287 бројева и 13.778 знакова, а најкраћи је Закон о измени закона о сточарству из 2012. године који има 41 реч, 8 бројева и 18 знакова.

Закони у корпусу у просеку имају 4.832 речи (948 различите), 326 бројева (67 различитих), 925 знакова (10 различитих) и 5.126 размака. Примећујемо да је број појављивања размака сличан укупном броју појављивања речи и бројева (5.158). Разлог томе је што се између два узастопна појављивања речи и бројева најчешће појављује један размак, без обзира да ли се такође појављују и знаци интерпункције, док се у релативно ретким случајевима између два појављивања речи и бројева не појављује ни један размак (на пример „72–75”) или се појављују два размака (најчешћи случај је када се између два размака нађе цртица, на пример на почетку алинеје или код одвајања делова реченице цртицом).

Најфреквентнији токен у Корпусу је размак (3.490.762 појављивања), а затим следе тачка (222.934) и запета (181.002). За њима је најфреквентнија реч „и” са укупно 139.248 појављивања.

Листа 20 најфреквентнијих речи (токена у облику секвенце слова) је дата у

табели 3.1. Напомињемо да се у овој статистици разликују мала и велика слова, тако да у оквиру 35.237 појављивања токена „Члан” није урачунато 6.972 појављивања токена „члан”. Ових 20 речи се јавља укупно 869.820 пута што представља 26,44% од укупног броја појављивања речи у корпусу. Речи са фреквенцијом 1, које се обично називају хапаксима, има 19.855 или 32,25% од укупног броја различитих речи.

Табела 3.1: Најфреквентније речи у корпусу

Редни број	СрпКор2013		Корпус закона	
	Фреквенција	Корпусна реч	Фреквенција	Корпусна реч
1	4.330.865 (3,54%)	<b>и</b>	139.248 (4,23%)	<b>и</b>
2	4.103.542 (3,36%)	<b>је</b>	93.809 (2,85%)	<b>у</b>
3	3.513.009 (2,87%)	<b>у</b>	71.636 (2,18%)	<b>се</b>
4	3.261.285 (2,67%)	<b>да</b>	70.109 (2,13%)	<b>за</b>
5	2.107.336 (1,72%)	<b>се</b>	49.992 (1,52%)	<b>на</b>
6	1.751.270 (1,43%)	<b>на</b>	39.019 (1,19%)	<b>је</b>
7	1.381.402 (1,13%)	<b>за</b>	38.067 (1,16%)	<b>да</b>
8	1.258.361 (1,03%)	су	37.396 (1,14%)	овог
9	919.922 (0,75%)	<b>од</b>	36.724 (1,12%)	или
10	779.469 (0,64%)	<b>са</b>	35.316 (1,07%)	<b>од</b>
11	740.476 (0,61%)	а	35.237 (1,07%)	Члан
12	650.144 (0,53%)	<b>који</b>	34.130 (1,04%)	<b>из</b>
13	612.218 (0,50%)	не	31.685 (0,96%)	члана
14	517.105 (0,42%)	<b>о</b>	31.026 (0,94%)	о
15	505.643 (0,41%)	ће	26.953 (0,82%)	<b>са</b>
16	501.473 (0,41%)	<b>из</b>	22.787 (0,69%)	<b>који</b>
17	474.120 (0,39%)	то	22.295 (0,68%)	односно
18	460.343 (0,38%)	као	18.544 (0,56%)	закона
19	445.434 (0,36%)	што	18.392 (0,56%)	става
20	417.727 (0,34%)	би	17.455 (0,53%)	које

На левој страни у табели 3.1 наведене су речи према учесталости у Кор-

пусу савременог српског језика, верзија СрпКор2013 [50], [51], [53]. Овај корпус садржи 122.255.064 речи, а 20 најучесталијих речи се укупно појављују 28.731.144 пута, што је 23.50% од укупног броја појављивања речи у корпусу. Подебљаним словима су обележене речи које се јављају у оба корпуса. Речи у левој колони, које су изостављене у десној, јављају се углавном непосредно после 20. речи у Корпусу законских текстова, што указује да високо фреквентни слојеви језика у целини (посматрано у корпусу савременог српског језика) задржавају високу фреквенцију и у Корпусу закона.

Са друге стране, одређене речи су карактеристично чешће присутне у прописима и ту пре свега треба истаћи групу речи „овог”, „члан”, „члана”, „закона” и „става” које се појављују у фразама којима се упућује на одредбе истог или другог прописа. Упућујуће фразе представљају најучесталију језичку конструкцију карактеристичну за подјезик прописа и њима смо посветили поглавље 4 овог рада.

Од осталих речи из високофреквентних слојева корпуса закона, за све осим за реч „односно” можемо рећи да су високофреквентне и у језику у целини, само што се нису нашле међу првих 20. У Корпусу савременог српског језика реч „односно” је са 63.938 (0,05%) појављивања на 128. месту по учесталости. Речца „односно” се у прописима користи у функцији сличној везницима „и” и „или” у разним случајевима када се навођењем везника „и” односно „или” не би на одговарајући начин исказало намеравано значење, што значи да се речца „односно” често користи у значењу „и/или” или „то јест”. Генерално можемо рећи да се „односно” користи да означи релацију чије значење логично произилази из контекста, као у примеру: „временски жиг је званично време придружено електронском документу или групи електронских докумената, којим се потврђује садржај електронског документа у то време, *односно* садржај сваког документа у групи” (из Закона о електронском документу). Наравно, речца „односно” се може употребити и за избегавање одређености, као у примеру из члана 3. став 2. Закона о психоактивним и контролисаним супстанцама:

2) психотропна супстанца је свака супстанца биолошког, *односно* синтетичког порекла која се налази на Списку, у складу са Конвенцијом



о психотропним супстанцама („Службени лист СФРЈ”, број 40/73), *односно* супстанца која делује примарно на централни нервни систем и мења мождане функције, због чега се мења перцепција, расположење, свест и понашање;

У наведеном примеру прво појављивање речце *односно* може бити замењено везником „или” јер супстанца може бити биолошког или синтетичког порекла. Очито је да се аутор ове одредбе осећао сигурније ако употреби речцу *односно*, али је тиме избегао да коришћењем везника „или” буде одређенији. Међутим, у овом случају крајње значење није изгубило одређеност, тако да једино можемо рећи да је стил изражавања можда непотребно бирократизован, јер у свакодневном језику не бисмо употребили реч *односно* у сличној ситуацији.

Друго појављивање речи *односно* је типичан пример задржане неодређености јер је направљен спој два начина дефинисања: административног (да ли се супстанца налази на одређеном списку) и стручног (да ли супстанца има одређене карактеристике исказане стручном терминологијом). Међутим, за евентуалну супстанцу која се по стручним критеријумима може оценити као психоактивна, али се још увек не налази на Списку, остало је неодређено да ли је у смислу наведеног закона сматрамо психоактивном супстанцом или не. То значи да у контексту у коме се наводи дефинисан појам, у случају када је питање које смо истакли релевантно, треба прецизирати значење (нпр. „психоактивна супстанца са Списка” или „има својства психоактивне супстанце”).

Особеност корпуса закона је, поред учестале употребе бројева, и изузетно велики број интерпункцијских и специјалних карактера, као и карактера из других, пре свега, латиничних алфавета. Токена сачињених од слова енглеског алфавета има укупно 8464 од чега 1032 различитих.

### 3.3 Примена електронских речника

Електронске речнике у *LADL*-формату (које смо помињали у одељку 1.2) чине редови облика:

$$w_t, w_l.K + Syn.Sem:(mgc)^* \quad (3.1)$$

где је  $w_t$  облик речи из текста,  $w_l$  – лема,  $K$  – ознака врсте речи<sup>5</sup>,  $SynSem$  – скуп синтаксичких и семантичких обележја додељених леми, а (потенцијално празан) низ  $mgc$  – низ морфосинтаксичких категорија које описују везу између  $w_t$  и  $w_l$ . Систем електронских речника се састоји из две основне компоненте: из речника простих речи  $DLF$  (ниски алфаветских симбола између два сепаратора) и речника композита<sup>6</sup>  $DLC$ . Речник простих речи, преко параметара  $K$ ,  $SynSem$  и  $mgc$  описује прецизно однос између  $w_t$  и  $w_l$ . Пример реда из српског речника простих речи је

*браћу,браћа.N+Hum+MG+Pl:fs4v*

где поље  $SynSem = N+Hum+MG+Pl$  описује да је реч о именици ( $N$ ) која означава људско биће ( $Hum$ ), има природни мушки ( $MG$ ) род и множину као природни број ( $Pl$ ), док код  $fs4v$  одређује да је реч о облику који је акузатив ( $4$ ) сингулара ( $s$ ) граматичког женског рода ( $f$ ) и означава живо биће ( $v$ ). Пример реда из речника композита је

*државним испитом,државни испит.N+Comp:msbq*

и интерпретира се на сличан начин као и ред из речника простих речи. Атрибут  $Comp$  указује да се ради о композиту. Приметимо овде да се присвојни придев „државни” мора слагати у роду, броју и падежу са именицом „испит”, па се отуда уместо 11 могућих граматичких значења овог придевског облика бира само једно које се слаже са именицом [54]. Подручје синтаксичко-семантичких својстава омогућава да се за сложену реч дефинишу додатни атрибути који могу да укажу на њену некомпозиционалност. Тако за лему „црна овца”, поље  $SynSem$  садржи атрибут  $Hum$  који означава особу: овај атрибут се не може одредити аутоматски на основу својстава комонената *црн.A+Col* и *овца.N+Zool* ( $Col$  за боја,  $Zool$  за зоолошки).

Речници простих речи и композита се могу конструисати аутоматски применом теорије коначних аутомата и трансдуктора [25], [55], [56]. Као што је по-

<sup>5</sup>У имплементацији е-речника, за ознаку врсте речи је везан код који дефинише тип флективне парадигме и који омогућава да се на основу  $w_l$  аутоматски генеришу облици  $w_t$ .

<sup>6</sup>Под композитом се овде подразумева контингентни низ простих речи чије се значење, по правилу, не може одредити на основу значења простих речи које је чине.

казано у [57], комбиновање графемских и морфографемских алтернација које одређује флективно понашање једне леме може се изразити и посредством регуларног израза. Из оваквог израза се даље непосредно конструише аутомат који ће генерисати све облике флективне парадигме у формату који дат изразом (3.1).

На пример, у табели 3.2 су упоредно приказани изводи из *DLF* речника са свим флективним облицима за три речи: четинар, аларм и жироскоп. Примећујемо да се синтаксна и семантичка обележја не мењају приликом промене броја и падежа, а мењају се само морфосинтаксичке категорије (флективни код).

Табела 3.2: Упоредни приказ извода из *DLF* речника за три речи

четинар,четинар.N+Bot:ms4q	аларм,аларм.N:ms4q	жироскоп,жироскоп.N+Conc:ms4q
четинар,четинар.N+Bot:ms1q	аларм,аларм.N:ms1q	жироскоп,жироскоп.N+Conc:ms1q
четинара,четинар.N+Bot:mw4q	аларма,аларм.N:mw4q	жироскопа,жироскоп.N+Conc:mw4q
четинара,четинар.N+Bot:mw2q	аларма,аларм.N:mw2q	жироскопа,жироскоп.N+Conc:mw2q
четинара,четинар.N+Bot:mp2q	аларма,аларм.N:mp2q	жироскопа,жироскоп.N+Conc:mp2q
четинара,четинар.N+Bot:ms2q	аларма,аларм.N:ms2q	жироскопа,жироскоп.N+Conc:ms2q
четинару,четинар.N+Bot:ms7q	аларму,аларм.N:ms7q	жироскопу,жироскоп.N+Conc:ms7q
четинару,четинар.N+Bot:ms3q	аларму,аларм.N:ms3q	жироскопу,жироскоп.N+Conc:ms3q
четинаре,четинар.N+Bot:mp4q	аларме,аларм.N:mp4q	жироскопе,жироскоп.N+Conc:mp4q
четинаре,четинар.N+Bot:ms5q	аларме,аларм.N:ms5q	жироскопе,жироскоп.N+Conc:ms5q
четинаром,четинар.N+Bot:ms6q	алармом,аларм.N:ms6q	жироскопом,жироскоп.N+Conc:ms6q
четинари,четинар.N+Bot:mp5q	аларми,аларм.N:mp5q	жироскопи,жироскоп.N+Conc:mp5q
четинари,четинар.N+Bot:mp1q	аларми,аларм.N:mp1q	жироскопи,жироскоп.N+Conc:mp1q
четинарима,четинар.N+Bot:mp7q	алармима,аларм.N:mp7q	жироскопима,жироскоп.N+Conc:mp7q
четинарима,четинар.N+Bot:mp6q	алармима,аларм.N:mp6q	жироскопима,жироскоп.N+Conc:mp6q
четинарима,четинар.N+Bot:mp3q	алармима,аларм.N:mp3q	жироскопима,жироскоп.N+Conc:mp3q

Такође примећујемо да дате три речи мењају облике на исти начин: у истом реду у табели 3.2 све три речи су добиле исти наставак (или ни једна није добила наставак) и исти флективни код. У табели табели 3.3 су приказане одговарајуће комбинације наставака и флективних кодова.

Табела 3.3: Наставци и флективни кодови за речи четинар, аларм и жироскоп

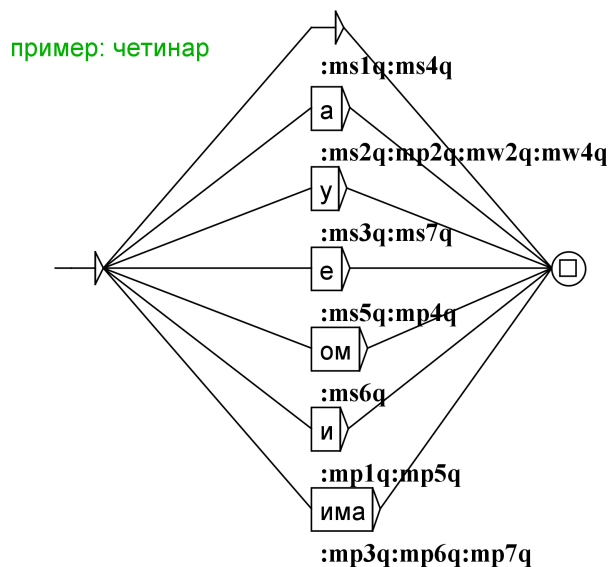
наставак	флективни код	наставак	флективни код
нема	ms4q	е	mp4q
нема	ms1q	е	ms5q
а	mw4q	ом	ms6q
а	mw2q	и	mp5q
а	mp2q	има	mp1q
а	ms2q	има	mp7q
у	ms7q	има	mp6q
у	ms3q	има	mp3q

За ове три именице кажемо да су из исте флективне класе коју у електронском речнику српског језика називамо *N1*. У питању су именице мушког рода које означавају неживу ствар и наставке за падеж и број добијају како је приказано у табели 3.3. Аутор речника има могућност да уместо да наводи редом 16 редова у *DLF* облику, за сваку именицу из класе *N1* зада по један ред у такозваном *DLS* облику:

четинар, N1+Bot  
 аларм, N1  
 жироскоп, N1+Comp

За разлику од *DLF* речника, у *DLS* речнику се задаје само лема, затим се основном синтаксном обележју додаје број, чиме се добија назив флективне класе, наводе се остала синтаксна и семантичка обележја и на крају се не наводе флективни кодови.

*Unitex* нуди могућност аутоматског генерисања *DLF* речника на основу *DLS* речника, што значи да се за претходно наведена три реда у *DLS* облику генеришу *DLF* редови приказани у табели 3.2. За сваку флективну класу њена флективна правила се одређују графом који носи исти назив као и флективна класа. Граф *N1* је дат на слици 3.11. Сваки пролазак кроз овај граф одређује један могући наставак и једну или више варијанти флективног кода, чиме се на крају

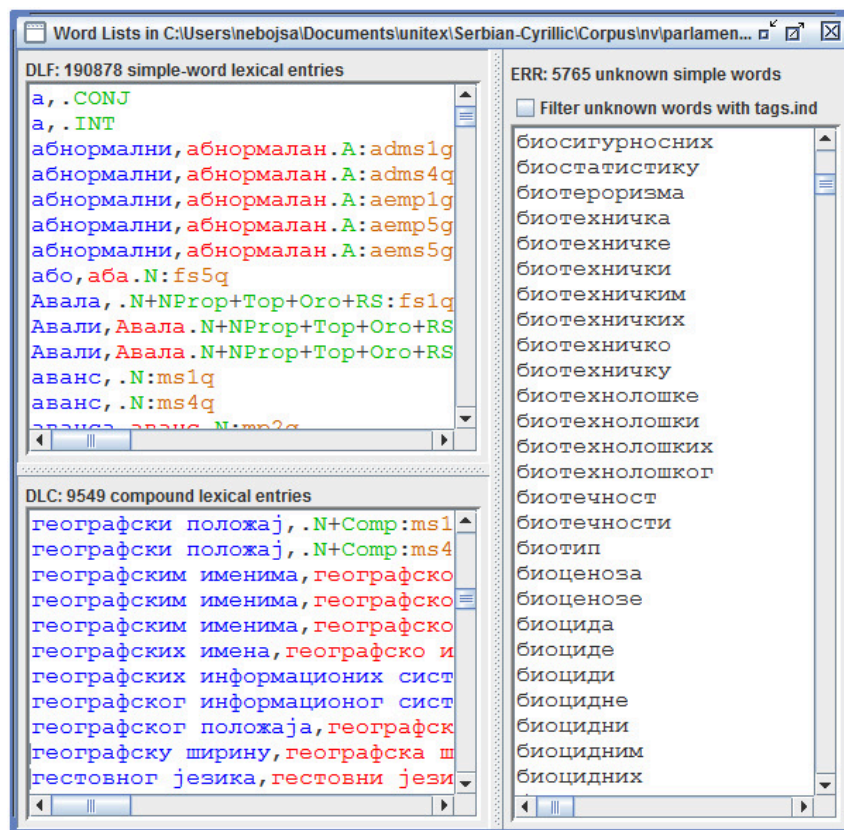


Слика 3.11: Флективни граф *N1*

добија истих 16 комбинација наставака и флективног кода које су приказане у табели 3.3.

У овом примеру се промена речи увек сводила само на додавање наставака. *Unitex* подржава и промене када се флективни наставци додају на лему из које су с краја обрисана нека слова, на коју су на крај или почетак додата нека слова и сл.

*Unitex* омогућава да се сравни, реч по реч, садржај корпуса са садржајем овако добијеног речника. Резултат сравњивања може бити да је ниска алфабетских карактера из корпуса једнака простој речи или композиту ( $w_t$ , тј. да одговара неком реду из речника *DLF* или *DLC*) или да такве ниске нема у речнику. У овом другом случају, такву ниску називамо непозната реч. Овакав прилаз даје високи одзив (енгл. *recall*) у препознавању могућих морфолошких интерпретација појединачних карактерских ниски из корпуса. За разлику од других приступа претраживању корпуса, с обзиром на садржај појединих поља у речницима, могуће је корпус претраживати по свим својствима која су придружена путем електронског речника, укључујући и комбинације тих својстава.



Слика 3.12: Резултат примене електронских речника на корпус прописа

Део резултата примене система електонских речника на корпус прописа је приказан на слици 3.12. У ранијем делу текста, када смо говорили о претходној обради корпуса у оквиру *Unitex*-а (поделељак 3.1.4), рекли смо да *Unitex* формира изведене речнике датог текста у којима се налазе само одреднице које се појављују у тексту. Сада можемо рећи да је то један *DLF* и један *DLC* речник датог текста и они су на слици 3.12 приказани у левом делу прозора. Насупрот речима које су пронађене у примењеним речницима, у десном делу прозора се налази списак непознатих речи. Видимо да је укупно препознато 190.878 различито граматичко значење простих речи и 9.549 композита, док се 5.765 различитих токена који би могли бити речи не налази ни у једном од примењених речника и представља непознате речи.

Имајући на уму феномен хомографије, који се не може отклонити на овом нивоу (ниво токена), по основним врстама речи, корпус садржи 1.518.864 појављивања потенцијалних именица (30.852 различитих облика), 576.307 појављивања потенцијалних глагола (9.906 различитих облика), 634.547 појављи-

вања потенцијалних придева (23.133 различитих облика) и 137.147 појављивања потенцијалних прилога (959 различитих облика)<sup>7</sup>.

Напоменимо да упит корпусу може бити формулисан у облику локалне граматике (коначног аутомата или трансдуктора) која представља сложену комбинацију речничког садржаја. Локална граматика, уколико је то потребно, може и да трансформише полазни текст.

### 3.4 Лексичке особености законских текстова

Језик закона припада административном функционалном стилу и могло би се очекивати да је су у њему присутно све оно што је, према [58, стр. 99], обележје бирократског језика, па се поставља питање да ли постоји корпусно оправдање за такав закључак. Учестаност појединих феномена који одликују бирократски језик се може проверити анализом корпуса законских текстова помоћу система електронских речника.

Погледајмо прво начин употребе саме именице „језик” која се појављује 993 пута и то 217 пута у синтагми „српски језик”, 112 пута као „страни језик”, 43 пута као „службена употреба језика”, 155 пута као „језик националне мањине”, итд. Пример који указује на намеру писца закона да превазиђе оквир наметнут правним формализмима налазимо у Закону о заштити потрошача:

Уговорна одредба обавезује потрошача ако је изражена *једноставним, јасним и разумљивим језиком* и ако би је схватио разуман човек потрошачевог знања и искуства.

Други пример употребе даје следећи извод из Закона о заштити и унапређењу права пацијената:

Ако пацијент не познаје *језик који је у службеној употреби* на територији здравствене установе, мора му се обезбедити *преводаца у складу са прописима о службеној употреби језика и писма*, а ако је *пацијент глувонем*, мора му се обезбедити *тумач*.

---

<sup>7</sup>Уколико се облици две различите врсте речи поклапају, онда се исти облик броји и у оквиру једне и у оквиру друге врсте речи.

Као особина бирократског језика се наводи склоност да се у његовој употреби користе дугачки именички ланци и непотребно прибегава апстрактним и глаголским именицама. Ако се у корпусу закона препознају низови од четири или више узастопних именица, наћи ће се свега 952 појављивања таквих низова у којима се укупно појављује 5052 речи (0,15% од свих речи у корпусу) попут „(испит за) стицање звања чланова посаде бродова трговачке морнарице”, „обављање делатности пружања услуга преноса новца”, „мера забране обављања делатности пружања услуга смештаја странцима”, „поступак утврђивања обавеза плаћања накнада” или „висина цене коштања карте аутобуса.”

Од укупног броја именица, у корпусу се глаголске именице<sup>8</sup> јављају 153.168 пута, што чини 4,66% од свих речи у корпусу. Ако се за конкорданце добијене оваквим упитом одреди фреквенција непосредног левог контекста, видимо да је један део глаголских именица употребљен као компонента композита (као у примерима „социјално осигурање”, „здравствено осигурање”, „клиничко испитивање”, „научно истраживање”, итд) или као део законских формула („дан ступања”, „дан објављивања”, „дан подношења”, „у трајању”, итд). Секвенца облика *глагол вршити + глаголска именица* јавља се ретко, свега 464 пута (или 0,028%). Глаголи који претходе глаголској именици са сличном фреквенцијом су „доносити”(0,029%), „забранити” (0,039%), „издати” (0,006%), и сл.

Појављивања апстрактних именица на „-ост” могу се екстраховати морфолошким филтером <N><<ости?>> који препознаје 41.143 таквих облика (или 1,25% ). Ако даље упоредимо фреквенције појединих придева из којих су изведене ове именице са изведеном именицом, добијамо следећу слику: придев „адекватан” се јавља 72 пута, а „адекватност” 46, „ажуран” 38, а „ажурност” 25, „активан” 515, а „активност” 1957, „алкални” се не јавља, али „алкалност” се појављује 2 пута. Детаљније истраживање би морало проверити у којим случајевима и са којим мотивима, писац закона даје предност апстрактној именици над придевом.

Текстови закона пружају и занимљиве примере прецизног дефинисања. Важење оваквих дефиниција је обично на нивоу једног закона. Један тип дефиниција се уводи изразом „у даљем тексту:” у малим заградама, а чија је намена да

<sup>8</sup>Глаголске именице су у електронском речнику обележене са *N+VN*.



се скрати именичка синтагма која претходи овом изразу. У корпусу има 2255 оваквих дефиниција, а примери су:

Привредни субјект може да продаје дуванске производе и путем посебних витрина (у даљем тексту: хјумидор).

безбедносна дозвола за приступ тајним подацима (у даљем тексту: дозвола)

државни службеници на дипломатским и конзуларним пословима (у даљем тексту: дипломате)

Термин уведен на овај начин је једнак синтагми коју је заменио. Термин не мора бити наведен у граматичком облику који се слаже са претходним контекстом, већ термин наводимо у облику у коме желимо да га дефинишемо, што је најчешће номинатив једнине.

У неким случајевима одређивање на шта се термин односи захтева пажљиву семантичку анализу, а што може укључити и разматрање уобичајеног значења термина који се дефинише. Пођимо од примера: „управљање и коришћење адреса и бројева (у даљем тексту: нумерација)”. Дефинисан термин значи „адреса или број” јер је термин дат у једнини, па ако адресе и бројеве једном речју називамо нумерацијама, онда појединачна нумерација може бити адреса или број. Дакле, да је писало „у даљем тексту: нумерације”, значење таквог термина би било „адресе и бројеви”. Да је којим случајем у заградама стајало „у даљем тексту: управљање и коришћење”, онда би значење дефинисаног термина било „управљање и коришћење адреса и бројева”. Видимо да одређивање синтагме коју дефинисан термин замењује не произилази само из форме, већ се треба руководити и принципом најлогичније интерпретације у односу на уобичајено разумевање.

С обзиром да за интерпретацију овакве дефиниције није довољна анализа лексичких, граматичких и елементарних семантичких својстава текста, већ захтева резонување на нивоу пуног разумевања природно-језичке семантике, ваљана интерпретација овакве дефиниције из информатичког угла се јавља као сложен проблем.

Други тип дефиниције чини исказна реченица у уобичајеној форми дефиниције, која почиње термином чије се значење одређује у смислу предметног закона, како би било недвосмислено јасно на шта се односе одредбе закона у којима се тај термин користи. Примери оваквих дефиниција су:

*опојна дрога* је свака супстанца биолошког, односно синтетичког порекла која се налази на Списку, у складу са Јединственом конвенцијом о опојним дрогама ("Службени лист СФРЈ", број 2/64), односно супстанца која делује примарно на централни нервни систем смањујући осећај бола, изазивајући поспаност или будност, халуцинације, сметње у моторним функцијама, као и друге патолошке или функционалне промене централног нервног система;

или

*синдром зависности* је скуп физиолошких, бихевиоралних и когнитивних феномена који се појављују после понављане употребе неке супстанце или групе супстанци, који карактеришу: јака жеља за узимањем психоактивних контролисаних супстанци, тешкоће у контроли употребе и поред штетних последица, повећање подношљивости, приоритетна усмереност ка психоактивној контролисаној супстанци уз запостављање других активности и обавеза, а понекад и апстиненцијално соматско стање;

или

*домаћа животиња јесте животиња* коју је човек одомаћио и чији опстанак зависи од непосредне бриге човека

или

*поштанска адреса* је адреса на којој прималац редовно прима пошту

У свакој од наведених дефиниција је дефиниендум одређен до нивоа на коме просечно образован човек може да одлучи да ли одређени објекат припада дефинисаној класи или не. Оваква прецизност представља разлику законске дефиниције од традиционалне лексикографске обраде која пружа значајно

скромније описе или их изоставља. На пример, у [59], „опојна дрога” и „домаћа животиња” су наведени као изрази, а „поштанска адреса” и „синдром зависности” се не помињу под одредницама адреса и зависност, а то даље значи да тумач закона не може да се ослони на постојеће речнике у утврђивању значења законске одредбе.

Посебност дефиниција у прописима јесте да се њима одређује правни статус, јер кад се утврди да нека ствар, лице или радња испуњава услове из дефиниције одређеног термина, онда се на ту ствар, лице или радњу аутоматски примењују све прописане одредбе које се односе на тај термин. На пример, Кривични законик прописује кривично дело неовлашћена производња и стављање у промет опојних дрога, тако да ако се испостави да је нешто опојна дрога према дефиницији из Закона о психоактивним супстанцама, то аутоматски може значити да је неко у кривичном делу према Кривичном законнику. Могу постојати супстанце које у стручном смислу имају наркотичка својства, али немају правни статус опојне дроге.

Оквири које треба дати одређеном правном термину који дефинишемо у пропису условљени су логиком целокупног правног система. Са друге стране су стручне дефиниције често обухватније јер укључују све што има смисла посматрати са одређеног аспекта, а посебно када су у питању основни појмови неке области. На пример, поштанска пошиљка у стручном смислу може обухватати и електронску пошту и у том смислу не смета што електронска пошта истовремено представља електронску комуникацију. Међутим, у правном систему је неопходно разграничити поштанске и телекомуникационе услуге, па је у прописима електронска пошта сврстана у електронску комуникацију, јер би у супротном сви телекомуникациони оператори морали да се лиценцирају и као поштански оператори.

У законима се срећу и примери када се два претходно описана типа дефиниције комбинују, попут:

*слободна зона је део територије Републике Србије (у даљем тексту: Србија) који је посебно ограђен и означен, на коме се обављају делатности под условима утврђеним овим законом (у даљем тексту: зона)*

Овај пример је специфичан и по томе што се уводе два термина за исти појам: „слободна зона” и „зона”. Разлог за то је што је у питању термин који се изузетно често појављује у тексту Закона о слободним зонама (224 пута у закону од 39 чланова), па је у таквом случају језички природно користити само „зона”. Са друге стране, „зона” је сувише уопштен термин, тако да се ван Закона о слободним зонама користи термин „слободна зона” када треба реферисати дати појам, као што имамо пример у Царинском закону: „...ако се роба уноси у слободну зону непосредно...”

У закону се експлицитна дефиниција може и изоставити, а да се опет сматра да је одређен појам уведен тим законом. Шта више, у старијим законима није ни био обичај да се експлицирају дефиниције. Тако Закон о чеку, без претходних дефиниција и уводних одредби, почиње са: „Чек садржи:...”

Погледајмо даље какав је однос између традиционалног лексикографског виђења појма „закон” и квантитативних показатеља из корпуса. Са тим циљем ћемо са чињеницама из корпуса упоредити групу 1802 из Систематског речника српскохрватског језика [60] која носи заглавље „закон”. Смисао оваквог поређења је, пре свега, процена употребљивости резултата традиционалне обраде када су у питању информатичке потребе. Један сличан пројекат, *Global WordNet*<sup>9</sup> [61], представља информатизовану слику организовања концепата у једном језику. Српска верзија<sup>10</sup> *WordNet*-а, из информатичког угла, је постала значајан ресурс у проналажењу информација на српском, али и у вишејезичној претрази.

У групи 1802 се као повезани термини, између осталих, налазе речи „закон”, „законски”, „законски пропис” и „законска одредба”. Фреквенције ових одредница у корпусу закона су редом 34.576 (1,0508%), 806 (0,0245%), 5 (0,0002%) и 26 (0,0008%), а у Корпусу савременог српског језика редом 122.697 (0,1003%), 11.127 (0,0091%), 667 (0,0005%) и 556 (0,0005%). Примећујемо да последња два израза имају сразмерно малу фреквенцију у законским текстовима, с обзиром да су у питању правни термини. Разлог томе је што је у оквиру самог законског текста довољно рећи само „пропис” и подразумева се да је то на основу

---

<sup>9</sup><http://globalwordnet.org/>

<sup>10</sup><http://korpus.matf.bg.ac.rs/SrpWN/>

закона донет пропис (или краће законски пропис), док се реч „одредба” углавном користи када упућујемо на одредбе одређеног закона или дела закона, а када говоримо генерално о законским одредбама, довољно је да користимо реч „закон”. Дакле, одредницу „законска одредба” је у неким случајевима могуће заменити одредницом „закон”, као у примеру: „Одлука ... може се побити ако је ... супротна *законским одредбама*”. У другим случајевима то није могуће: „... губи право на приговор којим се позива на непоштовање те *законске одредбе* или тог услова.” Однос између „закон” и „законски” је у оквиру *WordNet*-а описан посебном деривационом релацијом [62]. Одредница из следеће групе, „збирка прописа”, нема ниједно појављивање у корпусу закона и два појављивања у Корпусу савременог српског језика. У наредној групи, поред неколико одредница које се не појављују у корпусу закона, налази се и „кривични закон”. С обзиром да је од 2005. године Кривични закон замењен Кривичним закоником, у корпусу закона се, поред одреднице „кривични закон” са 40 појављивања, јавља још и „кривични законик” (211 пута), а такође и „законик о кривичном поступку” (88 пута) које Систематски речник српскохрватског језика не бележи. Слични подаци се добијају и за друге групе синонимних речи у оквиру групе 1802.

У корпусу закона је присутно имплицитно успостављање релација међу концептима. Погледајмо прво пример именица са значењем транспортног средства. У корпусу наилазимо на низ „транспортно средство”, „возило”, „аутомобил”, „камион”, „аутобус”, „трактор”, итд. Концепт транспортно средство обухвата, поред сувоземних, и пловна и ваздушна транспортна средства. Сувоземна транспортна средства могу бити возила где се даље разликују, с једне стране, моторно возило, прикључно возило, оклопљено возило, специјално возило, службено возило, путничко возило, комби-возило, ватрогасно возило, пољопривредно возило, шинско возило, итд. С друге стране, возила се разлажу на концепте аутомобил, камион, аутобус који и сами могу бити додатно спецификовани. На пример, аутомобил може бити путнички, службени или оклопљени, итд. Из информатичког угла, овакво разликовање омогућава да се успостави јасна семантичка мрежа (у смислу који се захтева у *WordNet*-у). Из угла језика закона, наведена хијерархија допушта да се прецизно одреди

објекат који је предмет закона, али и да се начине одговарајућа уопштавања као у примеру:

Инспектор врши инспекцијски надзор на прописаном месту које омогућава искључење *возила* из саобраћаја и отклањање утврђених недостатака.

где возило може бити било које од наведених.

Сличан хијерархијски ланац се може успоставити и између именица којима се именују особе на које се односе законске одредбе. Особе су у електронском речнику обележене атрибутом *Num*, што омогућава да се над корпусом постави упит облика  $\langle N+Num \rangle$  (именица која означава особу) како би се добио њихов списак. Овакве именице се појављују 230.191 пут, а примери су: агент, министар, корисник, лице, директор, грађанин, инспектор, предузетник, председник, људи, дужник, судија, подносилац, кандидат, корисник, произвођач, оснивач, тужилац, путник, власник, заменик, потрошач, службеник, заступник, осигураник, корисник, дете, итд. Од ових именица, најчешћа је „лице” са 24.282 појављивања, јер се та именица у прописима уобичајено користи као генерички термин за особу, тј. за класу  $\langle N+Num \rangle$ . Именица особа се појављује свега 358 пута и то нарочито у изразу „особа са инвалидитетом”. Када је реч о именовању особе, претходно изложени механизам дефиниција користи се да се прецизира, ако је потребно, значење употребљене одреднице што илуструју следећа два примера:

Под *лицем* из става 3. тачка 1) овог члана не сматра се *лице* које је било овлашћено за представљање и заступање.

или

Под појмом *грађанин*, у смислу овог закона, подразумева се не само *физичко лице* које је домаћи држављанин, већ и свако *физичко лице* страни држављанин, као и свако *домаће или странско правно лице* о чијим правима и обавезама одлучују органи управе из става 1. овог члана.

У последњем примеру видимо случај дефиниције која је ближа другом типу, али заправо представља исказ који по својој семантичкој садржини чини дефиницију, а не по посебној форми. Специфично за ову дефиницију је и да је дата као прецизирање у односу на уобичајено значење термина који се дефинише.

Лексика законских текстова, претражена кључем <N+Zool>, даје занимљиве примере назива животињских врста, као што илуструје фрагмент конкорданци у табели 3.4.

Табела 3.4: Фрагмент конкорданци за претрагу кључем <N+Zool>

погоном или псима, осим дивље свиње, вука, шакала и лисице; 3) употребом озе и свиње, живина (кокошке, ћурке, гуске, патке и друге пернате животи лобања и крзно медведа, вука, риса, дивље мачке и шакала, лобања лисице, чи пригоном, погоном или псима, осим дивље свиње, вука, шакала и лисице; а. SU случају клања домаћих свиња, дивљих свиња, коња и других врста је ловити: S1) јелена европског и дивљу свињу, осим подмлатка ових та јелена и срндаћа, рогови муфлона, дивојарца и дивокозе, зуби (кљове) тара, вирџинијског јелена, муфлона и дивокозу, осим подмлатка ових врста,

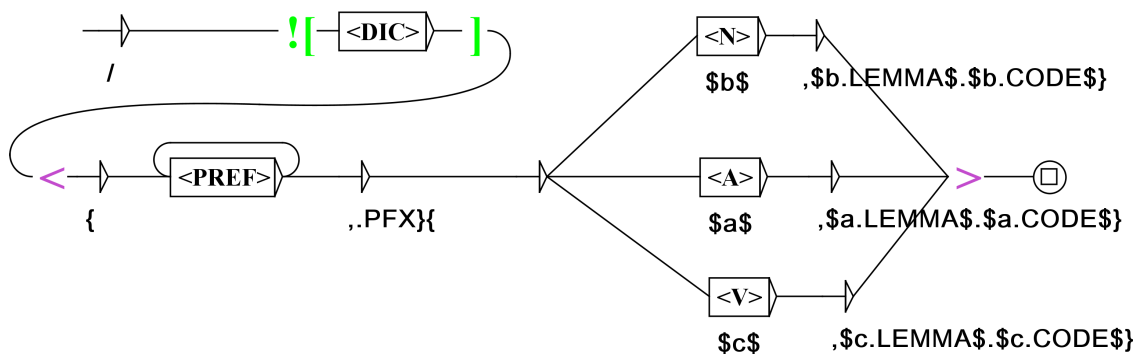
Претходна анализа језика законских текстова показује да је и поред бирократизованости, језик задржао богатство лексичког садржаја.

## 3.5 Непознате речи

Као што је у одељку 3.3 већ истакнуто, у корпусу закона се налази 5.765 речи које нису обухваћене речником и које се у информатичкој литератури називају непозната реч. Непознате речи се појављују 29061 пут што чини 0,88% од укупног броја речи у корпусу.

У односу на класификацију непознатих речи у српском, дату у [63], а с обзиром на раст електронског речника, у корпусу закона се више не појављују неки типови непознатих речи (пре свега, властита имена и скраћенице). Један од извора непознатих речи, регуларна деривација, такође се изузетно ретко среће<sup>11</sup>. Известан број непознатих речи су типографске грешке (нпр. „агортех-

<sup>11</sup>На пример, не појављује се ниједан облик аугментатива, а деминутиви су углавном лексикализовани са новим значењем, нпр. „флашица за бебе”, „звездица хотела”.



Слика 3.13: Речничка граматика

ничких”, „верисповест”, „дисциплинских” и сл.) укључујући и омашком употребљене карактере из других писама (нпр. „трошкова”). Највећи део непознатих речи долази из стручних терминологија и оне могу углавном бити или сложенице које нису у речнику или називи на страним језицима. У овој другој групи се налазе речи као *rent-a-car*, *limo service*, *Digital Object Identifier* које би морале ући у засебан електронски речник или латински називи попут *Vitis vinifera L.*, *Marteilia refringens*, *M. sydneyi*. Осим ових примера, нашу пажњу овде привлаче сложенице у српском [64]. Ослањајући се на концепт морфолошких филтера и речничких граматика, могуће је идентификовати непознате речи које су сложенице и доделити им одговарајуће граматичке информације. Принцип морфолошких филтера је већ био илустрован у препознавању апстрактних именица на „-ост”: они се наводе међу заградама облика << и >>, а сврха им је да открију дефинисани образац у скупу објеката који одговара лексичкој маски. На пример, морфолошки филтер <N><<изациј>> издваја 7145 (0,23%) појављивања 251 различитих облика именица које садрже суфикс „-изациј-” (нпр. имунизација, нормализација, популаризација, информатизација, комерцијализација, валоризација, докапитализација, карактеризација). Снажнији морфолошки механизам пружају речничке граматике. Формат њиховог записивања је сличан запису графа са слике 3.11, а сврха им је да од феномена који је њима описан генеришу редове који су сагласни са редовима речника *DLF* чији је пример дат изразом (3.1).

Размотримо деловање речничких граматика у препознавању сложеница међу непознатим речима. Напоменимо да речник *DLF* садржи као засебне ре-



дове листу префикса, префиксоида и радикаоида, а свима њима је, као врста речи, додељен кључ *PREF*. Користећи се овом чињеницом, формулисали смо речничку граматику приказану на слици 3.13. Значење овог графа је следеће: на његов улаз се доводи једна по једна реч корпуса. Ако се реч са улаза не налази ни у једном од речника (услов  $! [<DIC> ]$ ), онда ће она бити анализирана у морфолошком режиму који допушта да се ниво анализе спусти са нивоа речи на ниво карактера. Део графа у коме је активиран морфолошки режим је ограђен угластим заградама  $< \text{ и } >$ . Граматика даље покушава да изврши декомпозицију улазне речи према обрасцу описаном унутар угластих заграда. Овај образац ће бити успешно препознат уколико улазна (непозната) реч почиње бар једном речју која је у речнику *DLF* има ознаку *PREF*. Уколико то није случај, граматика се напушта, а то значи да непозната реч не припада овом обрасцу. Ако улазна реч прође ову „капију”, анализирани део ће бити запамћен као сегмент који је обележен са *PFX*, а преостали део улазне речи ће бити сравњен са једном од три могућности које су описане у десном делу граматике. Овај део испитује да ли се део речи који још није анализиран може успешно сравнити са неким именичким ( $<N>$ ), придевским ( $<A>$ ) или глаголским ( $<V>$ ) обликом у речнику типа *DLF*. Уколико овакав покушај успе, препознати облик ће бити постављен као вредност одговарајуће променљиве  $\$a\$, \$b\%$  или  $\$c\%$ , а на излазу ће целој улазној речи бити додељени граматички кодови наслеђени преко променљивих (нпр.  $\$a.LEMMA\%$  за додељену лему, а  $\$a.INFLECTED\%$  за граматичке категорије придева). Ознака / на самом почетку графа захтева од *Unitex*-а да од препознатих облика оформи речник у формату *DLF*.

Примена овог графа на корпус законских текстова препознаје укупно 2186 (38%) непознатих речи, а неки примери препознатих облика су: психоактивна, супраструктура, ванлинијски, међуповезивање, поткапитализована, интерресорни, хидрогеолошки, првопонуђен. У табели 3.5 су дати поједини редови из генерисаних конкорданци.

Табела 3.5: Конкорданце добијене применом графа са слике 3.13

руги начин доступан и да је {де, PFX}{компилација, компилација. N:fs1q} ограничена те државе, а једно или више {међу, PFX}{слетања, слетање. N+VN+Ek:np1q} предвиђено односно ако су енклаве или {полу, PFX}{енклаве, енклава. N:fp1q} у комплексу шума у те сврхе; 3) обезбеђивање да {хидро, PFX}{морфолошки, морфолошки. A+PosQ:adms1g}

Сличан граф, који у односу на граф са слике 3.13 уместо ознаке <PREF> има ознаку <A>, препознаје још 775 (13%) непознатих речи као, на пример: јавно-бележнички, сродно-правни, ниско-тиражни, пореско-прекршајни, само-опслуживање, само-вредновање, брзо-ток, шесто-валентни, итд. Применом речничких графова, који симулирају творбене моделе, може се, дакле, значајно редуковати број непознатих речи у тексту што има двоструку примену: прво, текст се може даље анализирати (јер је непознатој речи утврђен граматички статус) и друго, речи, препознате на овај начин, могу се евентуално додати у основну верзију речника типа *DLF*.

## 3.6 Именовани ентитети

Препознавање именованих ентитета спада међу главне задатке у области обраде природног језика и припада подобласти екстракције информација. Именованим ентитетима се сматрају лична имена, топоними, називи организација, називи послова и улога, временске одреднице, новчани износи, мере, проценти и сл.

Појам именованих ентитета је уведен на шестој *Message Understanding Conference* одржаној 1995. године [65], након чега препознавање именованих ентитета заузима значајно место у истраживању и практичној примени обраде природног језика. Свеобухватни преглед области препознавања именованих ентитета кроз приказ истраживачке литературе дат је у [66]. Аутор наведеног рада истиче да и поред тога што се највише истраживања у овој области односи на енглески језик, може се приметити да препознавање именованих ентитета заокупља пажњу шире истраживачке заједнице, и то не само за језике за које постоје добро развијени језички ресурси као што су немачки, француски, хо-

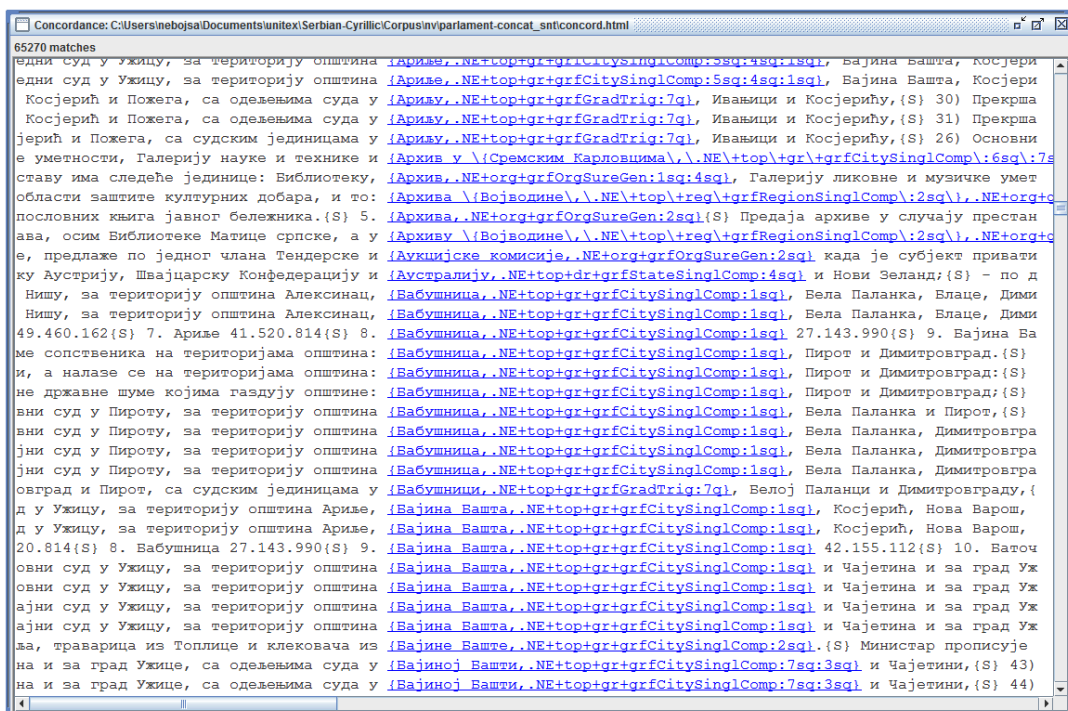


до децембра 2006. године. Вести у просеку нису дуже од две реченице.

На слици 3.14 је у форми конкорданци приказан део именованих ентитета које је систем препознао у корпусу агенцијских вести. Видимо да су именовани ентитети претворени у лексичке етикете са синтаксним кодом *NE* (од енгл. *named entity*). Првих десетак приказаних конкорданци представља топониме, и то насељена места. Затим је „Галеника” препозната као топоним (насеље Галеника), а онда је тај топоним под наводницима препознат као организација (предузеће „Галеника”). Преостали случајеви приказани на слици 3.14 представљају именоване бројеве, где се појављују количине, временске одреднице, проценти и новчани износи.

На слици 3.15 је у форми конкорданци приказан део именованих ентитета које је исти систем препознао у корпусу закона. Међу приказаним конкорданцама се налазе топоними и организације, при чему поједине организације у свом називу садрже топоним, попут Архива Војводине, што је такође уредно препознато.

Упоредан приказ релативних заступљености по пет најприсутнијих типова именованих ентитета за корпус кратких агенцијских вести и за корпус закона



Слика 3.15: Конкорданце именованих ентитета за корпус закона

дат је у табели 3.6. Подебљаним словима су означени типови који се појављују у обе колоне.

Табела 3.6: Најфреквентнији типови именованих ентитета

Редни број	Корпус агенцијских вести		Корпус закона	
	Заступљеност	Тип имен. ент.	Заступљеност	Тип имен. ент.
1	38%	<b>топоним</b>	39%	<b>организација</b>
2	20%	<b>лично име</b>	25%	<b>топоним</b>
3	17%	<b>организација</b>	24%	<b>врем. одредница</b>
4	11%	<b>врем. одредница</b>	13%	именовани број
5	10%	улога	< 1%	<b>лично име</b>

Значење појединих типова ентитета у табели је следеће:

- топоним – назив физичке или административне географске одреднице;
- лично име – пуно име, само име или само презиме;
- улога – функција, чин, звање, занимање и друге улоге;
- организација – назив институције, предузећа, удружења, међународне организације и сл.;
- временска одредница – датум, време, период и релативне временске одреднице (нпр. наредне године);
- именовани број – новчани износ, мера, количина и проценат.

Из табеле 3.6 се може видети да се у агенцијским вестима именовани ентитети користе да одговоре на питања „Где?” (топоними), „Ко?” (лична имена, организације и улоге) и „Када?” (временске одреднице), док се у законима именованим бројевима даје одговор и на питање „Колико?”. Такође се у законима ређе упућује на одређене особе (лична имена и улоге), а уместо тога се више упућује на одређене институције (организације).

## Поглавље 4

### Упућујуће фразе

У текстовима прописа (закона, подзаконских аката и слично) се често упућује на одредбе истог или другог прописа. Типични примери упућивања су ситуације у којима говоримо о поштовању или непоштовању других одредби, као што је: „Јавни позив се објављује у складу са чланом 22. овог закона” или „ако не поступи у складу чланом 33. овог закона”.

Поред тога, упућивањем можемо одредбу прописа искористити као посредну дефиницију појмова који се у тој одредби појављују, као што је случај у следећем ставу из Закона о рударству и геолошким истраживањима:

Општим актом привредног субјекта одређују се радна места у оквиру којих се врше послови из става 1. овог члана, а чије је обављање неопходно за обезбеђивање услова из члана 131. овог закона, као и објекти, односно просторије из става 2. овог члана.

У наведеном примеру се фразе „послови из става 1. овог члана”, „услови из члана 131. овог закона”, „објекти из става 2. овог члана” и „просторије из става 2. овог члана” користе као претходно дефинисани појмови, а да их одредбе на које се упућује експлицитно не дефинишу, већ се у њима само на неки начин помињу. На пример, став 2. истог члана (на који је упућивано) гласи:

Забрањено је организовање и спровођење штрајка у јамским просторијама или у другим објектима и просторијама у којима постоји потенцијална опасност по живот и здравље људи.

Свако надовезивање на неки део претходног текста прописа, осим у оквиру исте реченице, односно става, повлачи потребу за јасним упућивањем на одредбу на коју се надовезујемо. Као пример наводимо прва два става члана 218. Закона о парничном поступку:

У парници између странке и умешача који јој се придружио, умешач не може да оспорава утврђено чињенично стање, као и правне квалификације садржане у образложењу правноснажне пресуде (интервенцијско дејство пресуде).

Изузетно од става 1. овог члана, странка која је била умешач има право да истакне приговор да је странка из раније парнице којој се придружио као умешач погрешно водила претходну парницу или да је суд пропустио да јој доставља позиве, поднеске или одлуке.

У мање формалном контексту бисмо другу реченицу могли започети само са „Изузетно, странка која...”, али је у правно-нормативном тексту потребно експлицитно упутити на одредбу на коју се надовезујемо, па макар то била претходна одредба.

Због учесталости коришћења упућивања у прописима, као и важности упућивања у разумевању значења правних норми, кроз праксу израде прописа су се успоставили стандарди у начину изражавања упућивања, што је на крају прецизирано и одређеним формалним актима [22], [23].

Улога упућивања у прописима, као и потреба да се форма упућивања стандардизује није специфичност српске законодавне праксе, већ се у различитим облицима среће у законодавствима свих земља. На пример, Европска унија (ЕУ) издаје „Међуинституционални приручник за стил” [69] у коме се налазе правила за израду званичних докумената ЕУ, укључујући и детаљан опис форме упућивања у одељцима 3.1 и 3.2 тог приручника.

## 4.1 Упућивање и упућујућа фраза

Термин упућујућа фраза користимо да означимо синтаксну конструкцију којом се изражава упућивање. Према томе, термин упућивање се може одно-

сити и на саму упућујућу фразу и на семантичку садржину упућивања. У раду ћемо користити оба термина у зависности од тога да ли је потребно да експлицирамо да мислимо на синтаксну конструкцију, насупрот семантичкој садржини упућивања.

Основна класификациона јединица прописа на коју се врши упућивање је члан, а упућивање се може вршити и на делове члана као што су ставови, тачке, подтачке и алинеје, као и на пропис у целини. У одлукама, наредбама и упутствима основна класификациона јединица је тачка, али се специфичном структуром тих аката нећемо бавити у овом раду.

У оквиру једне упућујуће фразе се може упутити на више одредби. На пример, уколико бисмо желели да упутимо на одредбе из Јединствених методолошких правила за израду прописа [22] које говоре о претходно наведеним класификационим јединицама, рекли бисмо да су то одредбе из *чл. 24-29. и 36. и члана 37. ст. 4-5. Јединствених методолошких правила за израду прописа (Сл. гласник РС бр. 21/10).*

Као пример сложеног упућивања навешћемо и члан 160. став 2. Закона о ветеринарству (Сл. гласник РС бр. 91/05, 30/10, 93/12) који гласи:

Новчаном казном из става 1. овог члана казниће се за прекршај физичко лице ако учини радње из *члана 156. став 1. тач. 4) - 7) и тачка 11), члана 157. став 1. тач. 2), 3а), 5), 6), 7), 9) и тач. 12) - 16) и члана 158. став 1. тач. 1), 2), 3), 5), 7), 8), 9), 10), 12), 13), 14), 15), 16), 17), 19) и тач. 29) - 35) овог закона.*

Циљ овог поглавља је прецизирање синтаксних правила за упућујуће фразе, до нивоа детаљности који омогућава њихову аутоматску обраду. Поред формално установљених правила у [22], [23] постоје и уобичајена нормотехничка правила која ћемо покушати да реконструиремо из постојеће праксе. Све у свему, приликом формулисања синтаксних правила за упућујуће фразе руководићемо се следећим начелима:

1. *уважавање формално установљених правила у [22], [23] – ова правила представљају основ, али сама по себи нису довољно детаљна;*



2. *уважавање преовлађујуће праксе у савременим законима* – с обзиром да се уобичајена нормотехничка правила доследније примењују у законима него у подзаконским актима због природе процеса израде и доношења закона, примере из закона можемо сматрати релевантнијим;
3. *уважавање генералних језичких правила* – упућивање се појављује као део реченице на српском језику и сходно томе се примењују генерална језичка правила; на пример, када иза нумеричке ознаке стоји тачка, примењују се језичка правила за писање редних бројева из чега произилази када тачку треба изоставити;
4. *начело логичности* – из различитих разлога у законима нису потпуно доследно примењена истоветна нормотехничка правила, те је у том случају поред уважавања преовлађујуће праксе потребно оцењивати и шта је логично да буде обухваћено правилом, а шта је логично сматрати одступањем од правила;
5. *разумљивост и што мања сложеност правила* – правила не треба да буду сувише сложена и треба да буду исказана на разумљив начин;
6. *начело једнозначности* – и у случају најсложенијих форми упућујућих фраза синтаксна правила треба да воде ка једнозначном тумачењу.

Посебно осетљиво место у примени претходних начела је успостављање равнотеже између језичке слободе и потребе за већим степеном формалности у односу на уобичајене језичке конструкције, што је посебно изражено код последња два начела која воде ка томе да правила буду рестриктивнија од онога што би језичка слобода дозвољавала.

У наредним одељцима овог поглавља ћемо приказати синтаксна правила за упућујуће фразе која су уобличена у складу са претходним начелима, а затим ћемо у анализи разлика у нормативној пракси додатно образложити нека опредељења у формулисању правила. На крају поглавља ћемо описати и могућности рачунарске имплементације исказаних правила.

## 4.2 Цитирање одредби прописа у оквиру овог рада

Када у оквиру овог рада цитирамо одредбе прописа, суочавамо се са чињеницом да постоји разлика између начина навођења извора у научним и стручним радовима и начин упућивања на одредбе прописа:

- у научном или стручном раду се извор наводи до нивоа књиге или чланка (евентуално се наводи и страна), сви извори се наводе у листи литературе на крају рада, а на месту цитирања се наводи кратко обележје из листе;
- код упућивања на прописе у оквиру правног текста обично се прецизира део прописа на који се упућује до најнижег могућег нивоа који је логички оправдан, а на месту где се врши упућивање налази се цео опис упућивања у упућујућој фрази.

По питању начина цитирања одредби прописа, у оквиру овог рада ћемо се руководити следећим:

1. у случајевима када је пропис тематски повезан са овим радом, навешћемо га у списку литературе и реферисати попут [22], [23];
2. у случајевима када пропис само користимо као извор прикладног примера, користимо стил упућујућих фраза, с тим да ћемо у неким случајевима изабрати једноставнију или мало слободнију форму.

## 4.3 Елементарне форме упућивања

У циљу описа свих облика упућујућих фраза, прво ћемо описати одређене елементарне форме упућивања, да бисмо касније описали правила конструкција чијом применом се добијају преостали облици упућујућих фраза.

### 4.3.1 Упућивање на акт

Упућивање на акт садржи назив акта и ознаку гласила у којима је акт објављен укључујући касније измене и допуне тог акта, као на пример: *Закон о пореском поступку и пореској администрацији (Сл. гласник РС бр. 80/02, 84/02*

- исправка, 23/03 - исправка, 70/03, 55/04, 61/05, 85/05 - др. закон, 62/06 - др. закон, 61/07, 20/09, 72/09 - др. закон, 53/10, 101/11, 2/12 - исправка, 93/12).

Напомена иза броја Службеног гласника наводи се када промена није извршена законом о изменама и/или допунама предметног закона, већ на неки други начин (објављивање исправке, измена одредби предметног закона из другог закона, примена одлуке Уставног суда и сл.)

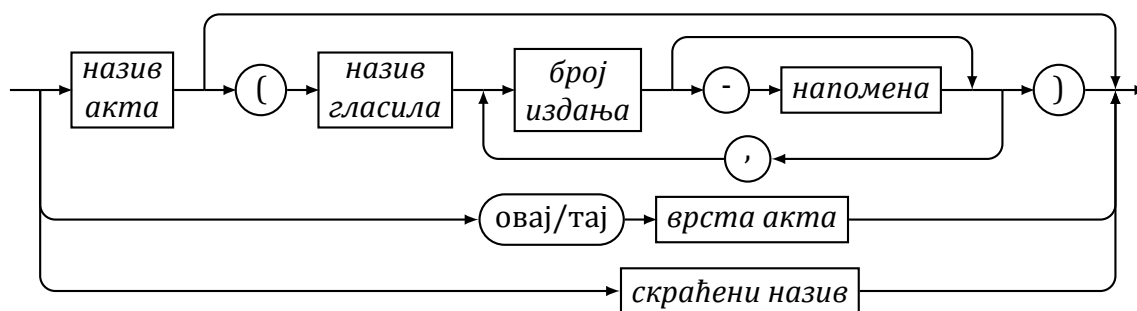
Уколико се на исти акт упућује више пута у истом тексту, онда се само први пут наводи где је акт објављен, а даље само назив акта или скраћени назив, уколико је претходно уведен.

Упућивање може да се врши и на сам акт у коме се налази упућујућа фраза, коришћењем показне заменице „овај” као на пример: „у границама које одређује овај закон” (из Закона о прекршајима).

Када је акт већ наведен у претходном делу реченице, на њега се може упутити и показном заменицом „тај”, као на пример у прелазној одредби Царинског закона:

Даном почетка примене овог закона, престаје да важи Царински закон ("Службени гласник РС", бр. 73/03, 61/05, 85/05 - др. закон и 62/06 - др. закон), осим чл. 252. до 329. тог закона.

Према томе, фраза која упућује на један акт може да има структуру описану дијаграмом 4.1.



Дијаграм 4.1: Упућивање на акт

Кућице на дијаграму представљају елементе од којих се слаже форма коју дијаграм описује, што је у овом случају упућивање на акт. Овалне кућице представљају конкретан текст (слова, бројеве или знаке интерпункције) који је у

њима уписан, док је у правоугаоним кућицама уписан назив или опис елемента који кућица представља. Сваки пролазак кроз дијаграм који почиње од улазне тачке са леве стране и завршава се у излазној тачки са десне стране одређује једно слагање форме од датих елемената редом којим се пролази кроз кућице.

### 4.3.2 Упућивање на делове акта

Делови акта на које се може вршити упућивање су члан, став, тачка, подтачка и алинеја. Називамо их класификационим јединицама. Следећи пример из Закона о финансирању локалне самоуправе илуструје све наведене типове класификационих јединица осим алинеје:

Приходи од уступљених пореза

Члан 35.

Јединици локалне самоуправе припадају следећи порези остварени на њеној територији:

1) порез на доходак грађана, и то на приходе од:

(1) пољопривреде и шумарства,

(2) самосталне делатности,

(3) непокретности,

(4) давања у закуп покретних ствари,

(5) осигурања лица,

(6) 80% од пореза на зараде који се плаћа према пребивалишту запосленог,

(7) остали приходи у складу са законом;

2) порез на наслеђе и поклон;

3) порез на пренос апсолутних права.

Изузетно од става 1. тачка 1) подтачка (6) овог члана, граду Београду припада 70% од пореза на зараде који се плаћа према пребивалишту запосленог.

Наведен члан има два става, при чему став 1. има три тачке, од којих тачка 1) има седам подтачака. Приметимо да се кроз тачке и подтачке наставља претходно започета реченица, због чега подтачке почињу именицом у генитиву, а на крају последње тачке се налази знак интерпункције за крај целе реченице.

Приметимо да у том смислу подтачка (7) није језички исправно формулисана, јер кад се подтачка (7) надовеже на уводни део тачке 1), добија се: „и то на приходе од остали приходи у складу са законом“. Језички исправно би било да се уводни део тачке 1) завршава са „и то:“, а да се у првих пет подтачака дода „на приходе од“. Међутим, као што видимо и из овог примера, у законодавној пракси се дешава да код дуже листе тачака, подтачака или алинеја, ставке при крају листе буду језички неусклађене са одговарајућим уводним делом, јер би језичко усклађивање захтевало преформулацију свих ставки.

Ово можемо сматрати пропустом или прихватљивом језичком недоследношћу која суштински не уноси нејасноће, а целу формулацију чини прегледнијом. Није реално очекивати да ћемо у сложеној структури ставова са тачкама и подтачкама увек успети да направимо формулацију која би остала јасна и језички исправна и када бисмо склонили ознаке за тачке и подтачке и заиста направили једну дугачку реченицу. Структура тачака и подтачака уноси додатно значење у став и на то се можемо ослонити, мада свакако треба настојати да тачке, подтачке и алинеје буду језички усклађене са одговарајућим уводним делом набрајања, колико год је то могуће.

Став се по правилу састоји од једне реченице, а само изузетно од више реченица, када је због разумљивости потребно раздвојити текст става у две или више реченица, а те реченице су по својој садржини сувише повезане да би се одвојиле у посебне ставове.

Као пример става са две реченице навешћемо члан 27. став 6. Закона о електронским комуникацијама који гласи:

Средства која представљају разлику прихода и расхода утврђених годишњим финансијским извештајем Агенције уплаћују се на одговарајући рачун прописан за уплату јавних прихода буџета Републике Србије и користе се преко Министарства за унапређење и развој области елек-

тронских комуникација и информационог друштва. Део разлике средстава, сразмеран приходима оствареним од оператора електронских комуникационих мрежа и услуга на територији аутономне покрајине, уплаћује се на рачун буџета аутономне покрајине и користе се преко органа аутономне покрајине надлежног за послове електронских комуникација за унапређење и развој области електронских комуникација и информационог друштва на територији аутономне покрајине.

На почетку сваког члана се наводи нумерација члана, тако што се у посебном пасусу, непосредно пре првог става у члану, наводи реч "Члан" иза кога следи број члана и тачка.

Ставови се приликом навођења не нумеришу, већ се сваки нови пасус сматра новим ставом, па је сваком ставу придружен број према месту у редоследу ставова у оквиру члана. У разним издањима збирки правних прописа, издавачи се често опредељују да додају нумерацију код ставова, што је корисно за читаоца, посебно код чланова са више ставова. Међутим, у званичном тексту прописа ставови се не нумеришу.

Тачка се нумерише бројем иза кога стоји затворена заграда, док се подтачка нумерише бројем у заградама. Код алинеја се приликом навођења не користи нумерација, већ водоравна црта, а број придружен алинеји се, као и код ставова, одређује на основу места у редоследу алинеја у оквиру подтачака.

Као пример тачака, подтачака и алинеја навешћемо члан 73. став 1. тачку 5) Закона о буџетском систему из 2002. године где се описују послови Управе за трезор:

5) буџетско рачуноводство и извештавање, које обухвата:

(1) рачуноводствене послове за:

- обраду плаћања и евидентирање примања;
- вођење дневника, главне књиге и одабраних помоћних књига за сва примања и издатке по корисницима буџетских средстава;
- међународне донације и друге видове помоћи;

(2) финансијско извештавање;

(3) рачуноводствену методологију, укључујући:

- одржавање система класификације;
- прописивање правила буџетског рачуноводства;
- прописивање захтева у погледу интерног и екстерног извештавања;

### **4.3.3 Суфикси у нумерацијама код измена и допуна прописа**

Код измена и допуна прописа могу се увести словни суфикси у нумеричке ознаке нових чланова, тачака и подтачака, да се не би поклопиле нумерације нових и постојећих чланова, тачака, односно подтачака. У случају уметања нових ставова и алинеја у пропису о изменама и допунама је потребно експлицитно навести да наредни ставови, односно алинеје мењају нумерацију.

На пример, Пословник Народне скупштине (Сл. гласник РС бр. 52/10) је измењен Одлуком о измени и допуни Пословника Народне скупштине (Сл. гласник РС бр. 13/11) чиме је, између осталог, иза члана 237, додат члан 237а. На пословник који је резултат примене измена и допуна упућујемо са „Пословник Народне скупштине (Сл. гласник РС бр. 52/10, 13/11)”. Резултат примене измена и допуна није званично објављен, па се упућивање односи на замишљен резултат примене измена и допуна.

Када неко запише измењен текст, према сопственом тумачењу примене измена и допуна, то се назива незванично пречишћен текст. Према томе, код измењених прописа, за рачунарску обраду резултата примене измена и допуна су нам потребни незванично пречишћени текстови.

Орган који је донео пропис може објавити и званично пречишћен текст, али не постоји пракса да се то ради за све прописе. Такође, у званично пречишћеном тексту се врши пренумерација класификационих јединица. За Пословник Народне скупштине је објављен званичан пречишћен текст у Сл. гласнику РС бр. 20/12, тако да на крају имамо да:

- члан 237а постоји у Пословнику Народне скупштине (Сл. гласник РС бр. 52/10, 13/11), али

- члан 237а не постоји у Пословнику Народне скупштине (Сл. гласник РС бр. 20/12 - пречишћен текст).

У званично пречишћеном тексту је члан 237а пренумерисан у 238, а сви наредни чланови су добили нумерацију померену за један.

Дакле, резултат примене измена и допуна се по питању упућивања суштински разликује од одговарајућег званично пречишћеног текста, јер, на пример, следеће две фразе упућују на различите чланове:

- члан 240. Пословника Народне скупштине (Сл. гласник РС бр. 52/10, 13/11);
- члан 240. Пословника Народне скупштине (Сл. гласник РС бр. 20/12 - пречишћен текст).

Законске одредбе којима се врши измена и допуна имају специфичну форму, што ћемо илустровати одредбама Закона о изменама Закона о државној ревизорској институцији из 2007. године:

#### Члан 1.

У члану 16. Закона о Државној ревизорској институцији ("Службени гласник РС", број 101/2005), у ставу 1. реч: "седам", замењује се речју: "пет".

Став 2. се брише.

Досадашњи ст. 3. и 4. постају ст. 2. и 3.

#### Члан 2.

У члану 19. став 1. речи: "већином гласова на седници којој присуствује већина од укупног броја народних посланика", замењују се речима: "већином гласова свих народних посланика".

#### Члан 3.

Овај закон ступа на снагу даном објављивања у "Службеном гласнику Републике Србије".



Одредбама о изменама и допунама се могу мењати, брисати и додавати делови текста из закона који се мења, а може се и променити нумерација класификационих јединица. На крају закона се наводе прелазне и завршне одредбе (члан 3. у претходном примеру).

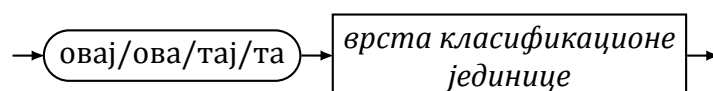
Могућности аутоматског пречишћавања текста закона на српском језику и својства језика „измена и допуна” анализирани су у раду „Ка програму за аутоматско мењање и допуњавање законских текстова” [70]. Неки од страних радова на ову тему су „Automatic Generation of Amendment Legislation” [71] и „Automated Handling of Amending Documents and Resulting Consolidations” [72].

#### 4.3.4 Повратно упућивање

Повратно упућивање упућује на класификациону јединицу у којој се налази сама упућујућа фраза, када користимо показну заменицу „овај” испред речи која означава врсту класификационе јединице, као на пример „овај члан”, „овај став” или „ова тачка”. Повратна упућивања су се више пута појављивала у досадашњим примерима, а навешћемо и пример из члана 28. Закона о судијама: „Дужност обавештавања из *овог члана* тече од дана пријема предмета у суд”.

Као посебан случај повратног упућивања сматраћемо и упућивање на класификациону јединицу која је већ поменута у оквиру исте реченице, када се користи показна заменица „тај” (дијаграм 4.2). Овај облик повратног упућивања се, на пример, појављује у одредби из члана 24. Закона о здравственом осигурању:

Права из обавезног здравственог осигурања утврђена овим законом обезбеђују се и члановима породице осигураника из члана 17. став 1. овог закона, осим из става 1. тач. 24) и 25) тог члана.



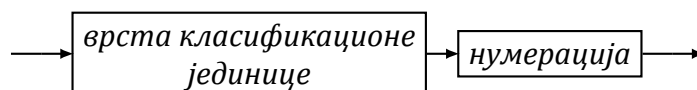
Дијаграм 4.2: Повратно упућивање

### 4.3.5 Просто једноструко упућивање

Под простим једноструким упућивањем ћемо подразумевати упућујућу фразу у којој се наводи врста класификационе јединице и нумерација (дијаграм 4.3), као, на пример, „члан 33.“, „став 5.“, „тачка 7а)“, „подтачка (9)“ или „алинеја трећа“. Свака врста класификационе јединице има другачија правила за нумерацију приликом упућивања:

- члан, тачка и подтачка се нумеришу истом ознаком која је наведена у тексту члана, тачке, односно подтачке;
- став се нумерише слично као и члан, бројем иза кога следи тачка, осим што за став немамо словне суфиксе, јер се број одређује према месту става у редоследу;
- алинеја се нумерише редним бројем који је исписан словима.

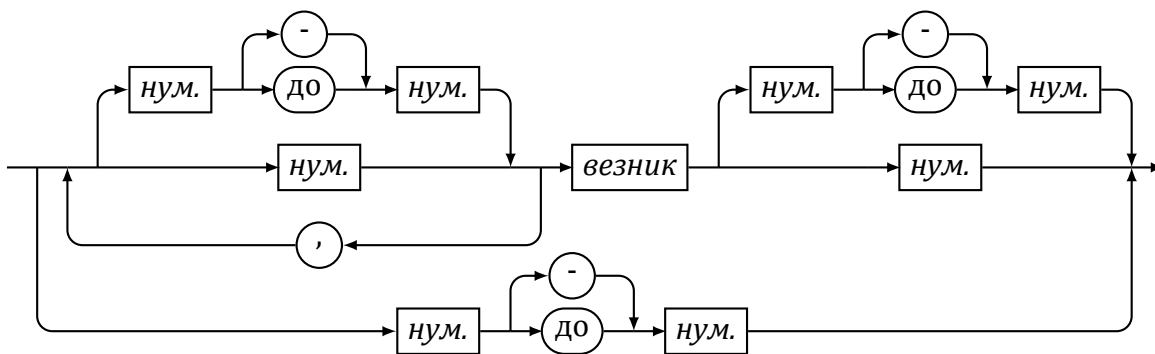
На дијаграму 4.3 ради једноставности нисмо приказали слагање врсте класификационе јединице и одговарајуће нумерације.



Дијаграм 4.3: Просто једноструко упућивање

### 4.3.6 Просто вишеструко упућивање

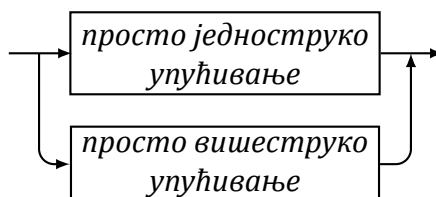
Код простог вишеструког упућивања, за разлику од једноструког, наводи се више нумерација, и то набрајањем, навођењем опсега или комбинацијом та два начина (дијаграм 4.4). При томе се врста класификационе јединице скраћује на првом слогу, као на пример: „ст. 1, 3-5, 9. и 11.“. У примеру видимо и примену правила за изостављање тачке у нумерацији испред запете и цртице, што је различито од случаја када се у нумерацији користи заграда, као на пример: „тач. 1), 3)-5), 9) и 11)“. Уместо везника „и“ на крају набрајања у неким случајевима се користи „или“ или „односно“.



Дијаграм 4.4: Вишеструка нумерација – скраћеница „нум.“ значи „нумерација”

### 4.3.7 Просто упућивање у општем случају

Просто упућивање је или просто једноструко упућивање или просто вишеструко упућивање (дијаграм 4.5).



Дијаграм 4.5: Просто упућивање

### 4.3.8 Ниво простог и повратног упућивања

Под нивоом простог, односно повратног упућивања ћемо подразумевати врсту класификационе јединице на коју се упућује. И код простог и код повратног упућивања, врста класификационе јединице се наводи на једном месту у упућивању: код простог на почетку, а код повратног као друга реч. Тако, на пример, ако просто упућивање почиње са „тачка” или „тач.,” онда је то просто упућивање нивоа тачке, док повратно упућивање „овог члана” има ниво члана.

## 4.4 Затворене и отворене упућујуће фразе

Упућујућу фразу „члан 156. став 1. тач. 4) - 7) Закона о ветеринарству” сматрамо затвореном, као и фразу „став 1. овог члана” јер је фразом у целини одређен контекст упућивања, односно упућивање је до краја прецизирано. У упућујућој фрази „члан 156. став 1. тач. 4) - 7)” упућивање почиње од нивоа члана, али је даље остало отворено у ком закону је тај члан, па ћемо овакву упућујућу фразу сматрати отвореном.

У прописима се користе затворене упућујуће фразе, а отворене једино у саставу затворених. У нормативној пракси се могу пронаћи изузеци, али и тада мора да постоји подразумеван контекст који семантички затвара упућивање, јер би у супротном упућивање остало неодређено. На пример, у казним одредбама Закона о јавним набавкама (Сл. гласник РС бр. 124/12), у члану 170. став 1. стоји:

Новчаном казном од 100.000 до 1.000.000 динара казниће се за прекршај понуђач, односно подносилац пријаве ако:

- 1) не чува поверљиве податке о наручиоцу (члан 15);
- 2) поступи супротно одредби члана 25. овог закона;
- 3) не обавести наручиоца о промени података или ако достави нетачне податке о испуњености услова за учешће у поступку или даје нетачне податке у погледу стручних референци (члан 77);
- 4) као подизвођача ангажује лице које није наведено у понуди и уговору о јавној набавци супротно одредбама овог закона (члан 80);
- 5) на основу одлуке Републичке комисије не надокнади наручиоцу трошкове поступка заштите права (члан 156. став 4).

У цитираном ставу се на крају појединих тачака у заградама налази упућивање на одредбу истог закона са којом је одређена тачка повезана. У заградама се наводи отворена упућујућа фраза којој недостаје одређење акта у оквиру кога се врши упућивање, јер се подразумева да се упућивање врши у оквиру

истог закона. Ова форма се у законодавној пракси често среће код казних одредби.

Отворена упућујућа фраза прецизира упућивање унутар претпостављене класификационе јединице. Претпостављена класификациона јединица најчешће произилази из контекста веће упућујуће фразе у чијем саставу је посматрана отворена упућујућа фраза, а може бити одређена и на други начин, као што је случај у претходно наведеном примеру из казних одредби Закона о јавним набавкама, где су у заградама наведене отворене упућујуће фразе, а претпостављамо да се упућивање врши унутар истог закона.

Дакле, отворено упућивање се у сваком случају мора семантички затворити, без обзира да ли се то чини кроз синтаксно затварање (стављање у састав затворене упућујуће фразе) или на други начин. Ако тај други начин има и препознатљиву синтаксну форму, можемо га сматрати и посебним начином затварања упућујуће фразе. Тако можемо сматрати да стављањем отворене упућујуће фразе у заграде добијамо посебну форму затворене упућујуће фразе, где се претпоставља контекст текућег акта, односно текуће класификационе јединице одговарајућег нивоа.

## **4.5 Почетни и крајњи ниво упућујуће фразе**

Упућујућа фраза „члан 156. став 1. тач. 4) - 7)” почиње од нивоа члана и упућује до нивоа тачке, па ћемо рећи да упућујућа фраза има почетни ниво члана и крајњи ниво тачке.

Почетни и крајњи ниво су семантичка својства која не морају да су повезана са тим шта се појављује на почетку, а шта на крају фразе. На пример, упућивање „члан 2. овог закона” има почетни ниво акта, а крајњи ниво члана. Дакле, почетни ниво је највиши ниво који је упућивањем одређен, а крајњи ниво је онај испод кога упућивање није даље прецизирано.

За затворене упућујуће фразе можемо рећи да увек имају почетни ниво акта, јер затворена фраза свакако одређује и акт у оквиру кога упућује, па према томе почетни ниво не може бити нижи (ужи) од акта. На пример, у затвореној упућујућој фрази „став 1. овог члана” је одређен закон у контексту кога се врши

упућивање, јер „овај члан” једино може да се налази у „овом закону”.

Упућујућа фраза увек има јединствено одређен почетни ниво: ако је затворена, то је ниво акта, а ако је отворена то је ниво који је први нижи од претпостављене класификационе јединице унутар које се врши упућивање.

Са друге стране, упућујућа фраза не мора да има јединствено одређен крајњи ниво упућивања. На пример „члан 156. став 1. тач. 4) - 7) и члан 157. став 2.” у једном делу упућује до нивоа тачке, а у другом делу до нивоа става, па нема јединствено одређен крајњи ниво упућивања.

Упућујућа фраза попут „члан 3. и став 2.” није исправна. Очигледно није затворена, а ако је третирамо као отворену, долазимо до тога да ни једна претпостављена класификациона јединица унутар које би оваква фраза упућивала не одговара: ако би упућивала унутар акта, онда не знамо у ком је члану „став 2”, а ако би претпостављена класификациона јединица била члан или нека јединица нижег (ужег) нивоа, онда „члан 3.” не би могао да упућује унутар члана, односно ниже класификационе јединице.

## 4.6 Слагање упућујућих фраза

Правила конструисања сложенијих упућујућих фраза полазећи од једноставнијих се могу свести на два основа типа слагања: хоризонтално слагање (набрајање) и вертикално слагање (додатно одређивање контекста отворене упућујуће фразе).

Упућујуће фразе истог почетног нивоа се могу набрајати и тако се може формирати нова упућујућа фраза. Као пример ћемо навести Члан 90. став 1. тачку 5) Закона о управљању отпадом (Сл. гласник РС бр. 36/09, 88/10), који гласи:

5) не обезбеди спровођење радног плана постројења, као оператер постројења за третман отпада или оператер на депонији не води прописане евиденције или не одреди квалификовано лице одговорно за стручни рад у постројењу, односно на депонији (члан 29. тач. 1), 8) и 9) и члан 30. став 1. тач. 1), 8) и 9);

Упућујућа фраза са краја претходног примера је почетног нивоа члана и

формирана је набрајањем две фразе истог почетног нивоа: „члан 29. тач. 1), 8) и 9)“ и „члан 30. став 1. тач. 1), 8) и 9)“.

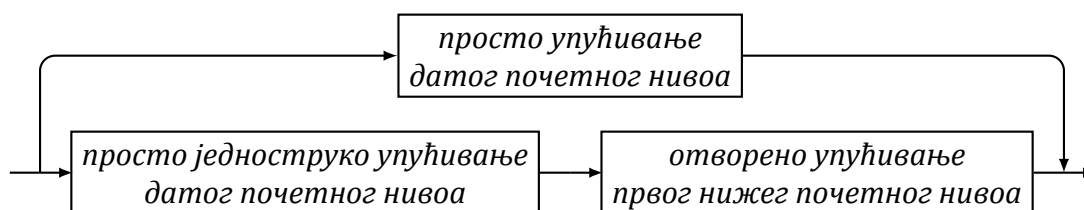
Такође, отвореној упућујућој фрази можемо додатно одредити контекст вишег нивоа и тако добити упућујућу фразу тог вишег почетног нивоа. На пример, ако испред упућујуће фразе „став 1. тач. 1), 8) и 9)“ додамо просто једноструко упућивање „члан 30.“ добијамо отворену упућујућу фразу почетног нивоа члана „члан 30. став 1. тач. 1), 8) и 9)“, а ако истој упућујућој фрази на крај додамо „овог члана“ добијамо затворену упућујућу фразу „став 1. тач. 1), 8) и 9) овог члана“.

#### 4.6.1 Конструкција отворених упућујућих фраз

Све отворене упућујуће фразе се могу конструисати од простих упућивања вертикалним и хоризонталним слагањем.

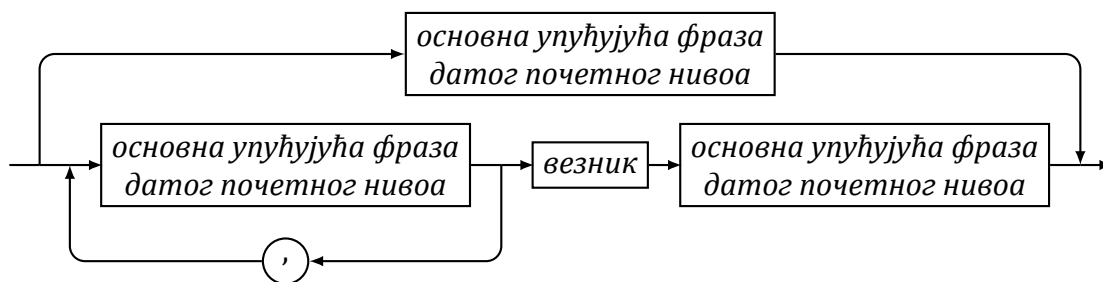
Отворену упућујућу фразу која не представља набрајање више фразе истог почетног нивоа називамо основном упућујућом фразом. Дакле, у претходном примеру смо имали отворену упућујућу фразу која је конструисана набрајањем две основне упућујуће фразе: „члан 29. тач. 1), 8) и 9)“ и „члан 30. став 1. тач. 1), 8) и 9)“.

Према томе, упућујућа фраза може бити или просто упућивање (једноструко или вишеструко) или резултат вертикалног слагања простог једноструког упућивање иза кога следи отворена упућујућа фраза првог нижег почетног нивоа (дијаграм 4.6).



Дијаграм 4.6: Основана упућујућа фраза датог почетног нивоа

Отворена упућујућа фраза може бити или основна упућујућа фраза или резултат набрајања (хоризонталног слагања) основних упућујућих фразе истог почетног нивоа (дијаграм 4.7).



Дијаграм 4.7: Отворена упућујућа фраза датог почетног нивоа

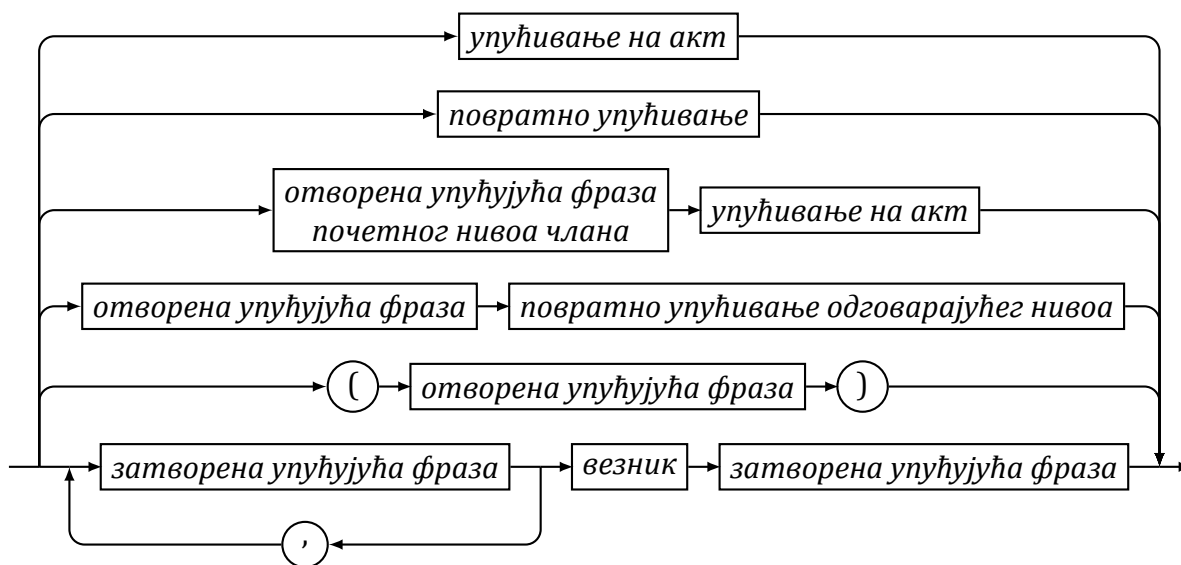
#### 4.6.2 Конструкција затворених упућујућих фраза

Затворена упућујућа фраза се може појавити као:

- елементарна затворена упућујућа фраза (упућивање на акт или повратно упућивање);
- резултат вертикалног слагања, односно затварање отворене упућујуће фразе (затварање додатним упућивањем на акт, повратним упућивањем или заградама) и
- резултат хоризонталног слагања, односно набрајања других затворених упућујућих фраза.

Сви наведени случајеви су приказани дијаграму 4.8.





Дијаграм 4.8: Затворена упућујућа фраза

## 4.7 Разлике у нормативној пракси

Пропис се не може оспорити само зато што није израђен у складу са методолошким или језичким правилима. Уколико је јасно шта одредба неког прописа исказује и ако то што исказује није спорно, онда у правном смислу одредба није спорна, без обзира на одступања од уобичајених или формално установљених језичких правила. Од тренутка када је закон или други пропис донет и објављен у Службеном гласнику, он је на снази такав какав је докле год се не измени, исправи или престане да важи на основу одлуке надлежног органа у складу са Уставом и законима.

Стога се у нормативној пракси појављују разлике у начину означавања класификационих јединица и изражавања упућивања. То може бити последица промена праксе током времена, различитих уверења и навика оних који су учествовали у изради прописа, омашки или непознавања правила.

### 4.7.1 Историјске разлике у нормативној пракси

Закони који су тренутно на снази доношени су током дужег временског периода. На пример, актуелни Закон о меници и Закон о чеку су оригинално обја-

вљени 1946. године у Службеном листу Федеративне Народне Републике Југославије, а последња измена оба закона је објављена 1996. године у Службеном листу Савезне Републике Југославије. У та два закона се, на пример, на више места користи скраћеница „чл.“ и када се упућује само на један члан. У многим законима из тог времена тачке су означаване словима, као на пример у Закону о „Партизанској споменици 1941“.

Због дисконтинуитета правног система који је направљен Законом о неважности правних прописа донетих пре 6. априла 1941. године и за време непријатељске окупације (Сл. лист ФНРЈ бр. 86/46, 105/46) немамо примере важећих прописа из ранијих периода, али нас то не спречава да направимо кратак осврт и на старија правна акта.

Табела 4.1: Начин означавања чланова и тачака у српским (и југословенским) уставима од 1835. до 1931. године

назив устава	означавање члана	означавање тачке
Устав Књажества Србије из 1835. године	5.) <i>текст члана</i>	б. <i>текст тачке</i>
Устав Књажества Србије – Султански хатишериф из 1838. године	5. <i>текст члана</i>	б.) <i>текст тачке</i>
Устав Књажества Србије из 1869. године.	Члан 5. <i>текст члана</i>	2., <i>текст тачке</i>
Устав Краљевине Србије из 1888. године	Члан 5. <i>текст члана</i>	2) <i>текст тачке</i>
Устав Краљевине Србије из 1901. године	Члан 5. <i>текст члана</i>	б) <i>текст тачке</i>
Устав за Краљевину Србију из 1903. године	Члан 5. <i>текст члана</i>	2) <i>текст тачке</i>
Устав Краљевине Срба, Хрвата и Словенаца из 1921. године	<b>Чл. 5.</b> — <i>текст члана</i>	2) <i>текст тачке</i>
Устав Краљевине Југославије из 1931. године	Члан 5. <i>текст члана</i>	2. <i>текст тачке</i>

У табели 4.1 је дат приказ начина означавања члана и тачке у српским уставима од 1835. до 1931. године.

У законима Краљевине Југославије је била честа пракса да се основна класификациона јединица (која одговара члану) означава знаком за параграф, на пример „§ 5.” У том случају је и код упућивања коришћен знак за параграф попут „§ 44” (код упућивања се не наводи тачка иза броја), при чему су се користили наставци за падеже попут „поменуте у §-у 3”, а за множину је коришћен двоструки знак за параграф попут „§§ 32 и 33”. Код упућивања на тачку је коришћена специфична форма „под бр. 5 §-а 6” или „§ 23 бр. 17”. У последња два примера се подразумева да се тачка на коју се упућује налази у првом ставу, а према потреби се прецизира и став „§ 17 ст. 1 бр. 1 и 2”. Примери у којима се користи знак за параграф узети су из Закона о јавним бележницима (нотарима) (Сл. новине Краљ. Југославије бр. 220/30).

Такође је у законима Краљевине Југославије коришћено и означавање чланова на начин сличан данашњем, па се чак дешавало да се у закону о измени закона и у закону који се мења користе различити начини обележавања основне класификационе јединице. Навешћемо пример из Закона о измени и допуни закона о земљишним књигама (Сл. новине Краљ. Југославије бр. 281/31):

#### Члан 1.

Закон о земљишним књигама од 18. маја 1930 год. мења се и допуњује како следи:

1) Став 2 §-а 38 мења се и гласи: ...

У наведеном примеру видимо и да на закон није упућивано према гласилу у којем је објављен, већ према датуму објављивања.

### **4.7.2 Разлике у савременој нормативној пракси**

Разлике у савременој нормативној пракси анализираћемо у односу на овде описану синтаксу упућујућих фраза, а кроз карактеристичне типове одступања.

Према овде описаној синтакси упућујућих фраза, алинеја се може појавити само у оквиру подтачке, али је у нормативној пракси чест случај да се алинеја уместо тачке појављује непосредно унутар става, односно да се уместо нумеричких ознака тачака употребе водоравне црте које означавају алинеје. Један од примера је Пословник Народне скупштине (Сл. гласник РС бр. 20/12 - пречишћен текст) где се у већини случајева уместо тачака користе алинеје, попут члана 179. став 3:

Образложење садржи:

- правни основ за доношење скупштинског буџета;
- образложење предлога потребних средстава за рад Народне скупштине.

У методологијама утврђеним од стране Народне скупштине и Владе [22], [23] јасно је наведено да се члан дели на ставове, ставови на тачке, тачке на подтачке, а подтачке на алинеје, при чему став не може имати само једну тачку нити тачка само једну подтачку, а такође се тачка и подтачка обавезно означавају бројевима. Све у свему, овде описана синтакса упућујућих фраза је усклађена са званичним методологијама по томе да не обухвата могућност постављања алинеје непосредно унутар става.

И у савременим законима, мада ретко, могу се пронаћи случајеви где упућујућа фраза почиње најнижом класификационом јединицом, на пример „тачка 3) става 1. овог члана” уместо „став 1. тачка 3) овог члана”. Наведен пример потиче из члана 20. став 3. Закона о личној карти при чему у оригиналу стоји „3.” уместо „3)”:

Лице које је у поседу личне карте умрлог лица, односно лице из тачке 3. става 1. овог члана дужно је да врати личну карту надлежном органу ради поништења у року од 15 дана од дана престанка важења личне карте.

Овај стил упућивања, где се прво наводи најнижа класификациона јединица је редак у савременој законодавној пракси и може се сматрати као одступање

од преовлађујуће праксе, тако да није обухваћен овде описаном синтаксом упућујућих фраза.

Један од честих типова одступања од званичних методологија па тиме и од овде описане синтаксе упућујућих фраза јесу неправилне нумеричке ознаке. Имамо случајеве, као што је и претходни пример, да иза нумеричке ознаке тачке не стоји заграда већ знак интерпункције „тачка” или да се алинеје означавају арапским бројевима. У Закону о изменама Закона о општем управном поступку (Сл. гласник РС бр. 30/10) имамо и случај да се ставови означавају бројем у загради, као на пример у члану 3. тог закона:

У члану 25. став (1) речи: ”Савезној Републици Југославији” замењују се речима: ”Републици Србији”.

У ст. (2) и (3) речи: ”савезни орган надлежан за иностране послове” замењују се речима: ”министарство надлежно за спољне послове”, у одговарајућем падежу.

У Закону о девизном пословању имамо случај да су у члану 2. тачке означене са бројем у загради, док су подтачке означене бројем и заградом, па су сходно томе и у упућивањима на тачке члана 2. коришћена нумеричка обележја у заградама.

Неодговарајући избор форме назива класификационе јединице за једину и множину („став” или „ст.” и слично) се такође може срести у пракси, као на пример у члану 24. Закона о оптичким дисковима, где став 1. тачка 2) гласи: „не поштује обавезе обележавања, чувања и вођења евиденције прописане овим законом (члан 12. ст. 1. тачка 5).” Такође, у члану 198. став 1. тачка 1) Закона о полицији се појављује упућивање: „чл. 59. до 65. и чл. 69.” (друго „чл.” би требало исправити у „члан”).

Претходно описан тип пропуста се често појављује при изради закона, али је врло редак у коначном тексту закона, захваљујући правнотехничкој редакцији у Републичком секретаријату за законодавство у случају када Народној скупштини закон предлаже Влада, што је најчешћи случај.

Следећи тип одступања од овде описане синтаксе упућујуће фразе је изостављање навођења става када се упућује на тачку, као на пример „члан 2. тачка

7) овог закона” уместо „члан 2. став 1. тачка 7) овог закона”. Наведен пример је из члана 3. Закона о здрављу биља где члан 2. има само један став са набројаним тачкама. У оваквом случају је упућивање потпуно одређено, али га ипак не сматрамо добром праксом јер би се у евентуалним допунама у којима би се додао нов став са тачкама на крај члана 2. нарушило упућивање које више не би било недвосмислено одређено. Није логично да се додавањем става на крај члана може нарушити упућивање на постојеће ставове у члану.

На крају ћемо навести више примера одступања од овде описане синтаксе упућујуће фразе који су везане за набрајање.

У набрајању се дешава да се изостави везник „и” на крају, као што је случај у примеру „Одредбе члана 1, члана 2. став 1, члана 7. став 1, члана 8. став 1, члана 9, у делу који...” из члана 23. Закона о изменама и допунама Закона о инвестиционим фондовима (Сл. гласник РС бр. 51/09). У оваквом случају је неспорно потребан везник „и”.

У казним одредбама Закона о буџетском систему (Сл. гласник РС бр. 54/09, 73/10, 101/10, 101/11, 93/12, 62/13) у члану 103. став 1. тачка 2) стоји:

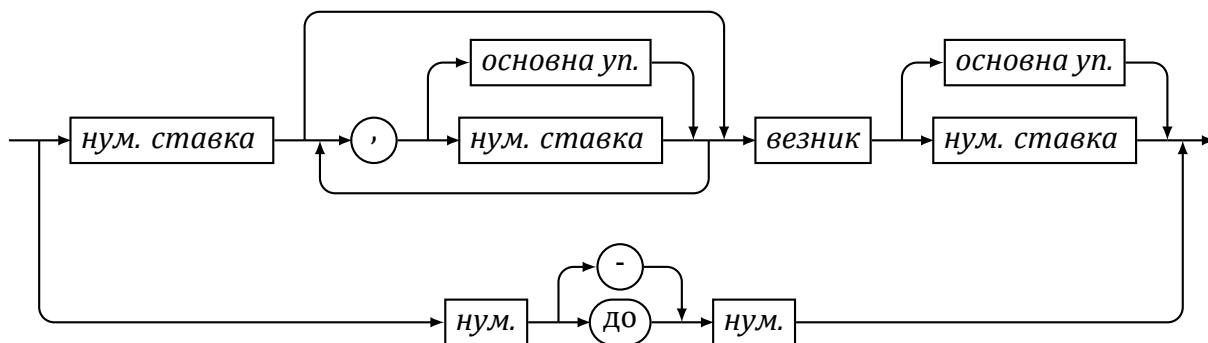
2) ако не поштује одредбе чл. 30, 37, 38, 40, 41. и члана 42. став 6. овог закона у поступку припреме и усвајања буџета.

Упућивање садржано у цитираној одредби није у складу са овде описаном синтаксом јер иза „чл.” треба да следи набрајање нумеричких ознака, па је везник „и” у склопу тог набрајања, што значи да фраза „члана 42. став 6.” стоји на месту где се очекује нумеричка ознака. Са друге стране, језички гледано, ово није неодржива форма, тако да овај пример указује да су правила која смо поставили нешто рестриктивнија од онога што би језичка слобода дозвољавала. Према тим правилима иста тачка би требало да гласи:

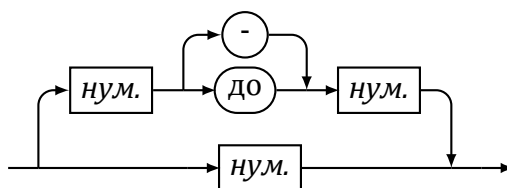
2) ако не поштује одредбе чл. 30, 37, 38, 40. и 41. и члана 42. став 6. овог закона у поступку припреме и усвајања буџета.

Тиме смо са првим „и” завршили набрајање нумеричких ознака, а онда друго „и” повезује две подфразе: „чл. 30, 37, 38, 40. и 41.” и „члана 42. став 6.”

Правила би било могуће променити тако да обухвате неизмењен пример из Закона о буџетском систему. У том случају би требало дијаграм 4.4 заменити дијаграмима 4.9 и 4.10.



Дијаграм 4.9: Вишеструку нумерација (алтернативно) – скраћеница „нум. ставка” значи „нумеричка ставка”, „нум.” значи „нумерација”, а „основна уп.” значи „основна упућујућа фраза одговарајућег почетног нивоа”



Дијаграм 4.10: Нумеричка ставка

У суштини смо дијаграме променили тако да смо у простом вишеструком упућивању одређеног нивоа свуда осим на самом почетку дозволили да се појави основна упућујућа фраза одговарајућег почетног нивоа. Овим се знатно усложњавају правила и постају мање разумљива, јер се уводи додатно кружно реферисање скоро са почетка на крај. Просто вишеструко упућивање више не би било просто, изгубили бисмо поступну изградњу фраза од простијих ка сложенијим, а не постоји логично образложење зашто је то потребно, осим да се дозволи један детаљ у језичкој слободи. Из измењеног примера који смо уподобили са претходно описаном синтаксом упућујућих фраза видимо да тај детаљ не оптерећује израду прописа. Када смо све то узели у обзир, а у складу са у уводном делу изнетим начелима, определили смо се да могућност навођења

основне упућујуће фразе на месту нумеричке ознаке код набрајања нумеричких ознака не укључимо у овде описану синтаксу упућујућих фраза.

Поред примера који су неисправни у смислу овде описане синтаксе упућујућих фраза, срећемо и случајеве када је непотребно изабран сложенији облик упућујуће фразе, али још увек у оквиру овде описане синтаксе.

Узмимо пример члана 279. став 1. Закона о здравственој заштити, који гласи:

Даном ступања на снагу овог закона престаје да важи Закон о здравственој заштити ("Службени гласник РС", бр. 17/92, 26/92, 50/92, 52/93, 53/93, 67/93, 48/94, 25/96 и 18/02), осим чл. 77а до 77е и чл. 78. до 85.

Упућујућа фраза „чл. 77а до 77е и чл. 78. до 85.“ је могла да се напише и као „чл. 77а до 77е и 78. до 85.“, а да представља семантички идентично упућивање. Обе варијанте су у складу са овде описаном синтаксом упућујућих фраза. У првој варијанти имамо два проста упућивања која су као основне упућујуће фразе истог почетног нивоа повезане везником „и“ (дијаграм 4.7), а у другој варијанти имамо једно просто вишеструко упућивање где су набројани сви чланови на које се упућује (дијаграм 4.4). У Закону о добровољним пензијским фондовима и пензијским плановима у члану 14. став 4. имамо пример који одговара другој варијанти:

Народна банка Србије издаје сагласност када процени испуњеност услова из члана 12. став 1. тач. 1) до 3) и 5) до 7) овог закона.

## 4.8 Препознавање упућујућих фраза у корпусу

Једна од првих примена рачунара у лингвистици била је аутоматско генерисање конкорданци, што је и даље често коришћено средство у лингвистичким истраживањима. На слици 4.1 приказане су конкорданце за упућујуће фразе које је генерисао *Unitex* на корпусу правних текстова.

Да би *Unitex* могао да препозна упућујуће фразе и прикаже их у конкорданцама, било је неопходно да се формално специфицира синтакса упућујућих



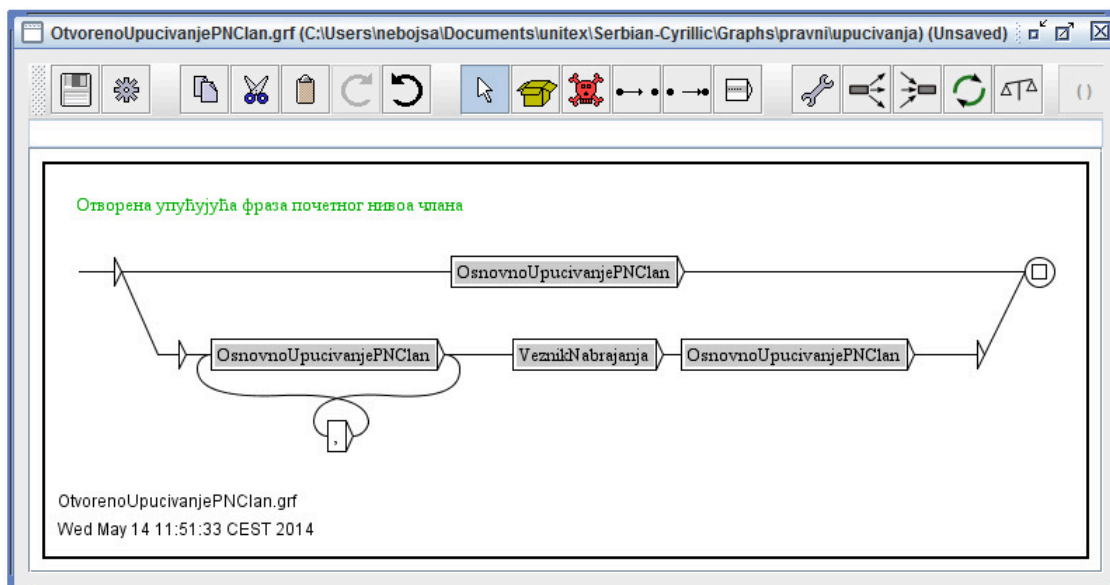
фраза. *Unitex* омогућава да се синтакса специфицира помоћу графова (синтаксни графови су били поменути и у одељку 3.5), а затим граф може да се користи као упит по коме ће се претражити текст и формирати конкорданце. На слици 4.2 видимо како изгледа један синтаксни граф у корисничком интерфејсу *Unitex*-а, а у прилогу А (одељак А.2) приказани су графови на основу којих су генерисане конкорданце са слике 4.1.

У појединим графовима из прилога А можемо приметити текст који стоји испод одређене кућице у графу и обично у себи садржи витичасту заграду. У питању је излаз трансдуктора (текст који се укључује у резултат обраде), о чему ће бити речи у поглављу посвећеном синтаксној анализи (поглавље 6), а у оквиру овог поглавља излаз трансдуктора није релевантан и приликом генерисања конкорданци са слике 4.1 га нисмо користили.

Конкорданце су генерисане у односу на граф *Uprisanje* који обједињује за-



Слика 4.1: Конкорданце за упућујуће фразе на корпусу правних текстова



Слика 4.2: Пример синтаксног графа у корисничком интерфејсу *Unitex*-а

творене и отворене упућујуће фразе (граф 4 из прилога А). Графове ћемо реферисати према називу који стоји у заградама на крају наслова испод слике графа јер се под тим називом графови реферишу у оквиру других графова. Тај назив уједно представља назив датотеке (не рачунајући наставак *.grf*) у којој је граф сачуван. Због чињенице да се назив графа подудара са називом датотеке, у називу смо користили само слова из енглеског алфабета.<sup>1</sup>

Према томе, граф *Upucivanje* укључује графове *ZatvorenoUpucivanje* и *OtvorenoUpucivanje*. Кућицу која укључује граф *B* са тачке гледишта овог поглавља можемо просто занемарити, као да је у питању кућица са празном речју.

У оквиру овог рада нећемо детаљно објашњавати структуру и семантику синтаксних графова, о чему се може прочитати у [29]. Међутим, графичка презентација синтаксе коју дају графови је довољно интуитивна да у многим случајевима и читаоци који нису посебно упућени у синтаксне графове могу наслутити шта поједини графови представљају.

Приметићемо и да графови *ZatvorenoUpucivanje* и *OtvorenoUpucivanje* даље

<sup>1</sup>И поред тога што већина модерних оперативних система подржава ћириличне карактере у називима датотека, у пракси се још увек релативно често срећу проблеми са словима изван ASCII скупа карактера у именима датотека приликом преноса датотека или њиховог коришћења у разним апликацијама.

укључују друге графове, и тако даље, па граф *Upriscivanje* непосредно и посредно укључује све графове из одељка A.2.

Можемо приметити сличност графова из одељка A.2 и дијаграма из овог поглавља, јер су ти дијаграми представљали основ за израду графова. Такође примећујемо да су, у односу на дијаграме, графови значајно разрађени да би прецизно покрили све случајеве.

Конкорданце смо формирали тако да обухватају и затворене и отворене упућујуће фразе да бисмо имали прилике да анализирамо и случајеве када се у тексту стицајем разних околности појави отворена упућујућа фраза која није део затворене. Типичан такав случај су неисправно формиране упућујуће фразе, па се неки делови неисправних фраза препознају као мање упућујуће фразе.

На слици 4.1 при дну видимо да је више пута препозната фраза „члана 153.” и из десног контекста конкорданце видимо да је то заправо део неисправно формиране затворене упућујуће фразе. У првом од таквих случајева примећујемо фразу „члана 153. тач. 1), 2), 4), 5) и 6)” која је требало да гласи „члана 153. став 1. тач. 1), 2), 4), 5) и 6)”. У питању је честа омашка да када имамо тачке у првом ставу, а посебно ако је то једини став члана, пропустимо да наведемо став у упућивању. Видимо да нешто изнад, код упућивања унутар члана 152, није направљена слична омашка.

Други карактеристичан случај где се препознају отворене упућујуће фразе су саме ознаке чланова које су у истој форми као просто једноструко упућивање на члан, што није обухваћено на слици 4.1, али имамо прилике да видимо на почетку сваког члана. Ово сматрамо нежељеним препознавањем јер део текста који не представља упућивање има форму упућујуће фразе. Да смо препознали само затворене упућујуће фразе, не би се десило ово нежељено препознавање. Од оваквог нежељеног препознавања се можемо заштитити и ако претходно препознамо логичку структуру прописа и међу препознатим ознакама чланова даље не покушавамо да препознамо упућујуће фразе, а о томе ће бити више речи у наредним поглављима.

## 4.9 Препознавање неисправних упућујућих фраза

Код препознавања неисправних упућујућих фраза немамо унапред дефинисану границу између онога што треба препознати као неисправну упућујућу фразу и онога што не треба препознати нити као исправну нити као неисправну упућујућу фразу. На пример, у фрази „тачке 3.” имамо очигледну омашку да иза броја не треба да стоји тачка, већ заграда: „тачке 3)”. Међутим, за примере „3 тачке” или „три тачке” је мање вероватно да је то требало да буде упућивање.

Дакле, да бисмо лоцирали потенцијалне грешке у упућујућим фразама, потребно је да прво одредимо шта све сматрамо да личи на упућујућу фразу, па у оквиру тога да идентификујемо шта нису исправне упућујуће фразе.

У одељку А.3 приказани су синтаксни графови који описују фразе сличне упућујућим.

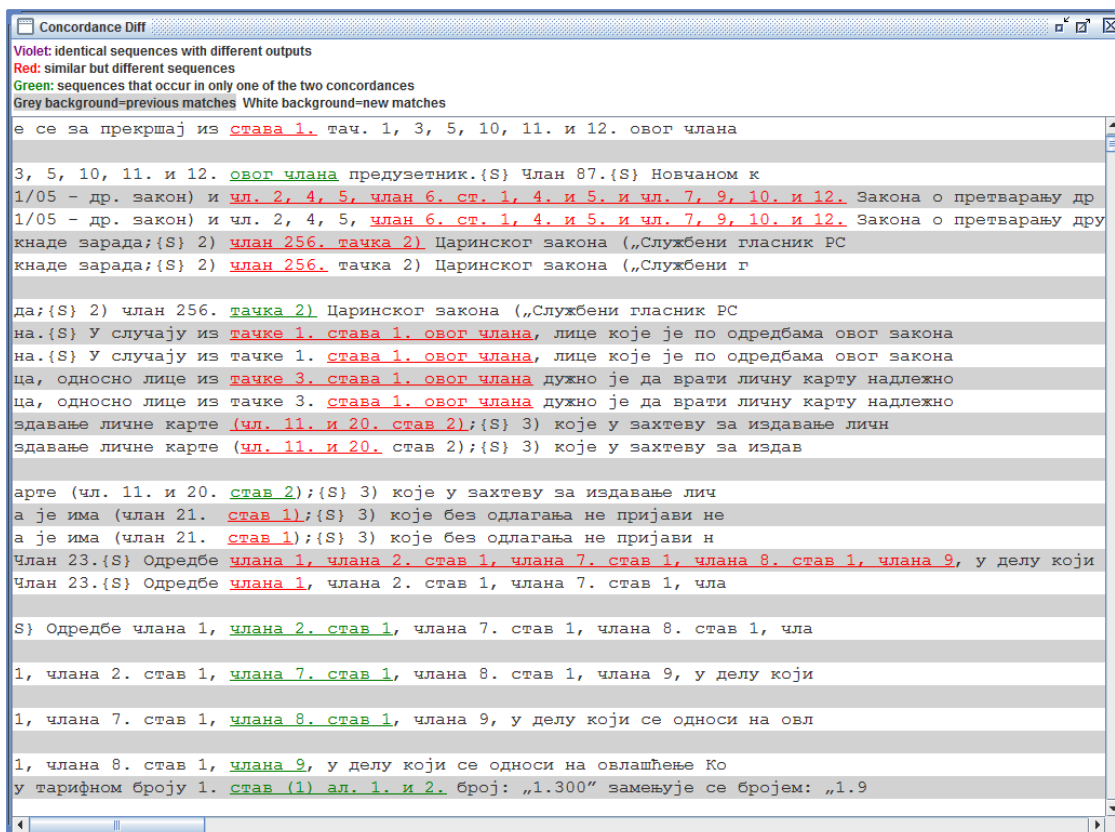
Граф *NumeracijaL* (последњи у одељку А.3) обухвата све варијанте нумерација које се користе у упућивањима, укључујући и неисправне комбинације бројева и интерпункције.

Граф *ProstoJednostrukoUpucivanjeL* комбинује све називе типова класификационих јединица, у пуној и скраћеној форми, са свим варијантама нумерације, чиме су укључене и неисправне комбинације.

Код графа *OtvorenoUpucivanjeL* не водимо рачуна о правилном слагању класификационих јединица по питању нивоа упућивања, јер обухватамо и случајеве када нивои упућивања нису правилно усаглашени (нпр. набрајамо упућивања различитих почетних нивоа).

Код формирања графова за фразе сличне упућујућим просто занемарујемо одређена синтаксна правила јер желимо да препознамо и фразе у којима се омашком та синтаксна правила нису поштовала. Тиме смо уједно и поједноставили синтаксне графове. Синтаксна правила занемарујемо до нивоа до кога остаје довољно карактеристичних детаља у „разлабављеној” синтакси да вероватноћа да ћемо лоше препознати делове текста за које није ни била намера да представљају упућивања није велика.

*Unitex* омогућава приказ разлика двају конкорданци, тако да можемо да упо-



Слика 4.3: Поређење конкорданци у *Unitex*-у

редимо конкорданце за граф *UrucivanjeL* и *UrucivanjeL*, што је приказано на слици 4.3.

Линије са белом позадином представљају конкорданце за граф *Urucivanje*, док линије са сивом позадином представљају конкорданце за граф *UrucivanjeL*.

Приметимо сиву линију у којој је препознато „чл. 2, 4, 5, члан 6. ст. 1, 4. и 5. и чл. 7, 9, 10. и 12.“. Испод ње је бела линија у којој је препознат само део почев од „члан“. Грешка је у томе што набрајање иза првог „чл.“ није завршено са „и“: „чл. 2, 4 и 5, члан...“

У претходном примеру је у оквиру једне фразе сличне упућујућој препозната једна упућујућа фраза и то је приказано у две узастопне линије. Међутим, када дужа упућујућа фраза није исправна, може се десити да се у оквиру ње препозна више мањих. Такав случај имамо код „тачке 1. става 1. овог члана“, где је грешком тачка наведена пре става, па су у оквиру ове фразе сличне упућивању препознате две упућујуће фразе: „тачке 1.“ и „става 1. овог члана“. Пошто *Unitex* упоређује конкорданце тако што упарује конкорданце из једног и из



*UpscivanjeL* и истовремено једно или више препознавања делова текста графом *Upscivanje*.

Приказан део преклопљених конкорданци обухвата лоцирана појављивања упућујућих фраза у Закону о изменама и допунама Закона о Државној ревизорској институцији. Примећујемо да су лоциране и фразе које представљају обележја чланова јер имају форму која одговара упућивању на члан (у датом примеру их препознајемо по великом почетном слову у речи „Члан”). Такође можемо приметити да је у дотичном закону коришћена тачка иза броја члана са суфиксом, у ком случају је тачку требало изоставити, затим видимо да недостаје везник „и” у једном набрајању, због чега су на том месту препознате четири мање упућујуће фразе и на крају видимо да су на једном месту коришћене упућујуће фразе различитог почетног нивоа повезане везником „и”, што није исправна конструкција.

# Поглавље 5

## Логичка структура

### 5.1 Основни елементи логичке структуре прописа

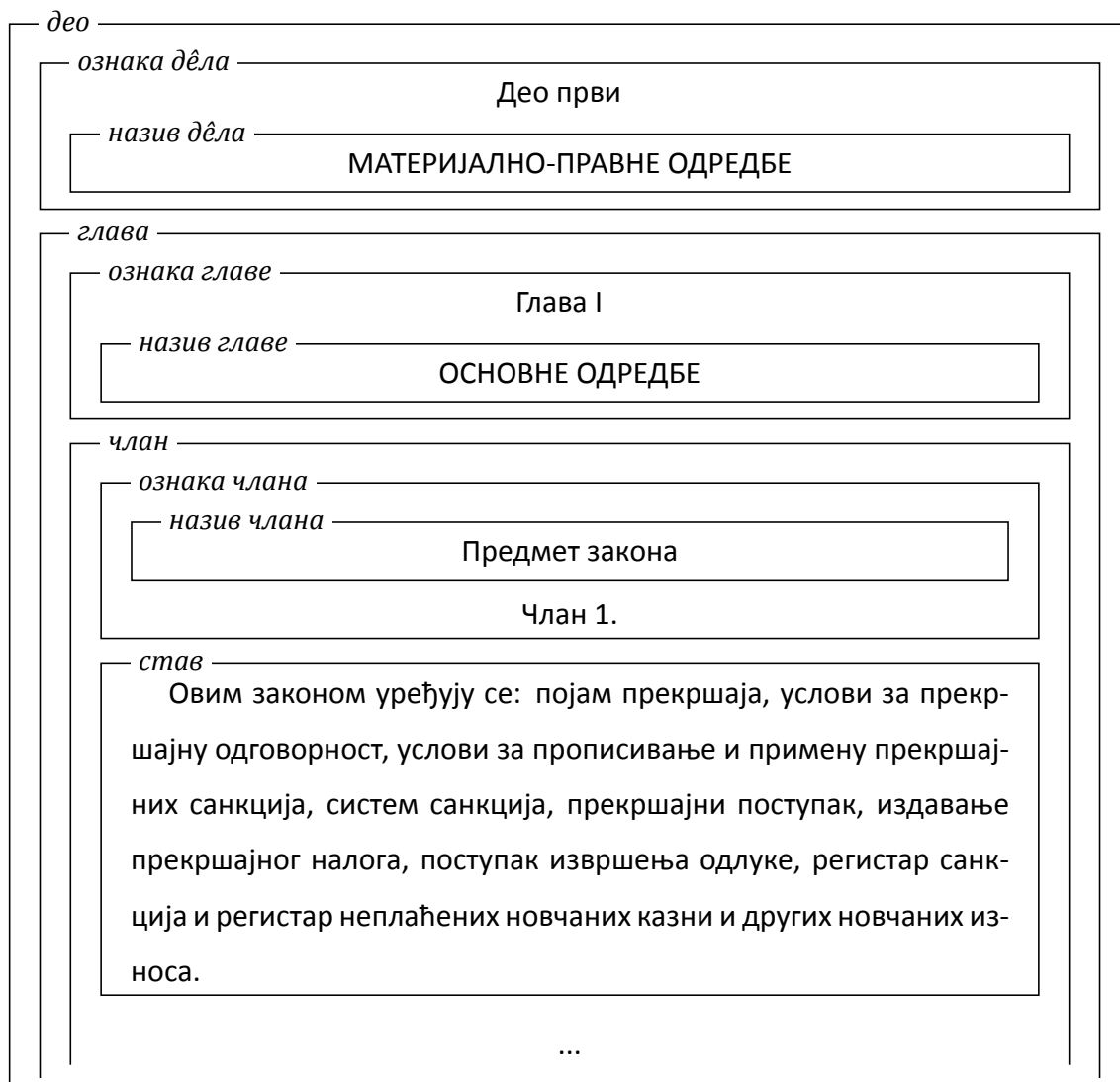
У претходном поглављу у пододељку 4.3.2 описане су класификационе јединице на које се може вршити упућивање (члан, став, тачка, подтачка и алинеја). Подела на класификационе јединице чини основну логичку структуру акта. У прописима је по правилу основна класификациона јединица члан, а изузетно је у одлукама, наредбама и упутствима основна класификациона јединица тачка. Специфичностима логичке структуре аката у којима је основна класификациона јединица тачка нећемо се бавити у овом раду.

Пропис може имати и шире класификационе јединице од члана, и то: део, главу, одељак и пододељак. Увођење ширих класификационих јединица зависи од броја чланова прописа и потребе да се истакне груписање чланова у шире класификационе јединице. Опште правило је да прописи који немају више од 20 чланова не треба да имају шире класификационе јединице [22], [23].

Циљ овог поглавља је да се опишу елементи логичке структуре закона на сличан начин на који су у претходном поглављу описане упућујуће фразе.

На слици 5.1 приказан је део Закона о прекршајима на коме је илустрована подела на логичке целине. Навођење назива делова, глава и чланова није обавезно, а у неким случајевима се назив може писати у истом реду са нумеричком ознаком тако да је ознака главе могла да се напише и у форми „I. ОСНОВНЕ ОДРЕДБЕ”. Када се узме у обзир да такође није обавезно да закон има шире кла-





Слика 5.1: Елементи логичке структуре прописа

сификационе јединице, једино са слике 5.1 што је обавезан део сваког закона је ознака „Члан 1.“ и садржај члана који следи.

## 5.2 Поглед на логичку структуру прописа прилагођен локалним граматикама

Методe обраде природно-језичког текста засноване на локалним граматикама, које користимо у овом истраживању, подразумевају да се синтаксним правилима описују релативно мали делови текста до нивоа реченице или евен-

туално неколико повезаних реченица (од чега и потиче назив „локалне грамати- тике”).

Алати који су базирани на локалним граматикама, што је у случају овог ис- траживања већ више пута поменути систем *Unitex*, нису прилагођени случају када цео текст или значајан део текста треба да се препозна као једна целина у односу на граматику. У питању је другачији приступ од граматика које се ти- пично користе за опис формалних језика, а посебно програмских језика, где је уобичајено да се цео текст препознаје као целина која се даље дели на подце- лине [56].

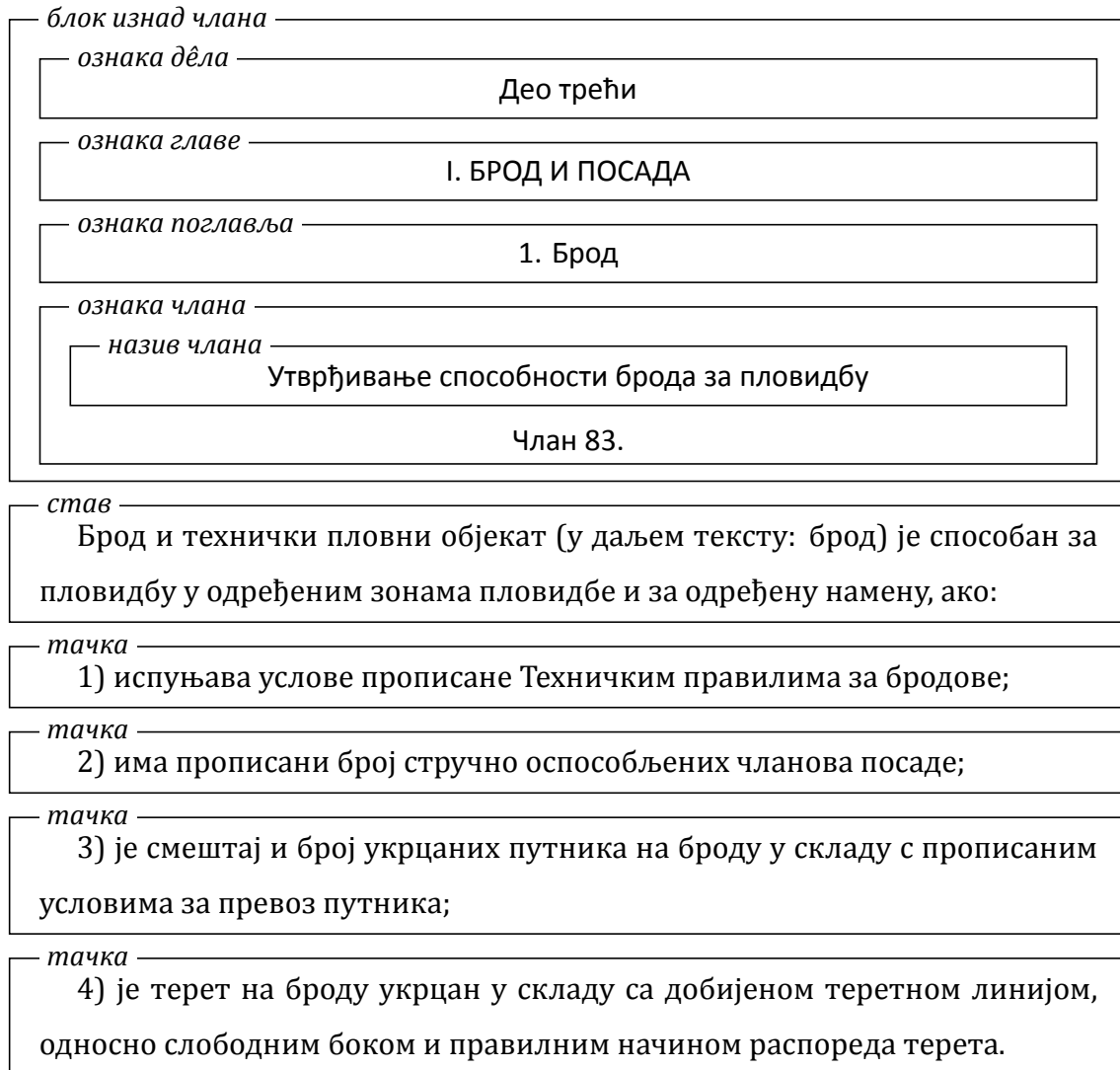
Другим речима, код локалних граматика јединица препознавања није цео текст, већ мање језичке конструкције које се препознају унутар текста. С тачке гледишта обраде природног језика ово је логичан приступ јер граматичка пра- вила природног језика углавном одређују односе унутар појединих реченица.

При томе треба имати у виду да уситњавање јединица препознавања по- већава могућност да дата синтаксна правила препознају и случајеве које ни- смо желели. На пример, ако бисмо јединице препознавања уситнили толико да назив изнад члана постане засебна јединица препознавања, таква форма извучена из контекста би се препознала на почетку сваке реченице, јер не са- држи ништа карактеристично што је разликује од неколико почетних речи ре- ченице.

Овом проблематиком се нисмо тако детаљно бавили код упућујућих фраза јер за тим није постојала потреба: упућујућа фраза по својој природи не из- лази из оквира једне реченице, а у себи садржи довољно карактеристичних детаља који смањују могућност да нека форма буде нежељено препозната као упућујућа фраза. Међутим, сада морамо да обратимо пажњу да јединице препо- знавања изаберемо тако да имају довољно карактеристичних елемената који обезбеђују низак степен нежељеног препознавања.

Дакле, у примеру са слике 5.1, целине текста које представљају део, главу и члан су сувише крупне за јединице препознавања, па ћемо стога дати нешто другачији поглед на структуру прописа, где се текст прописа види као секвенца мањих целина које ће се касније имплементирати као засебне јединице препо- знавања. Више детаља о даљој имплементацији синтаксне анализе дато је у

поглављу 6. Нови поглед на структуру прописа је илустрован на примеру дела текста из Закона о пловидби и лукама на унутрашњим водама, што је приказано на слици 5.2.



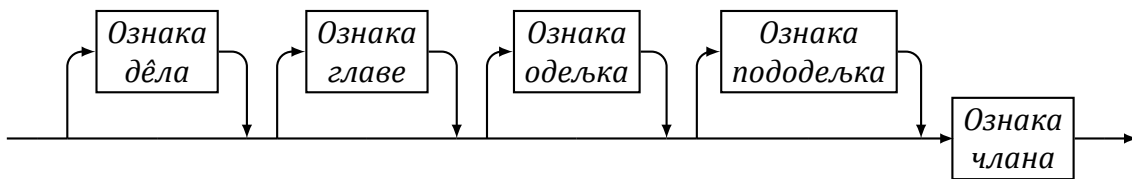
Слика 5.2: Прилагођен поглед на логичку структуру

С обзиром да се ознаке ширих класификационих јединица увек налазе непосредно изнад ознаке члана, синтаксно их посматрамо као једну групу коју смо назвали „блок изнад члана”, са идејом да то буде једна јединица препознавања. Одлучили смо се за овакво груписање зато што ознаке појединих класификационих јединица саме за себе немају довољно карактеристичних елемената који би обезбедили низак степен нежељеног препознавања, а при томе на овај начин уоквирена целина није сувише велика.

На основу наведеног можемо очекивати, а што ће се у даљем истраживању и потврдити, да је блок изнад члана добар избор за јединицу препознавања. Успешним препознавањем блокова изнад члана је уједно одређена подела на чланове који се налазе између два блока и иза последњег, чиме је решен један ниво логичке структуре прописа.

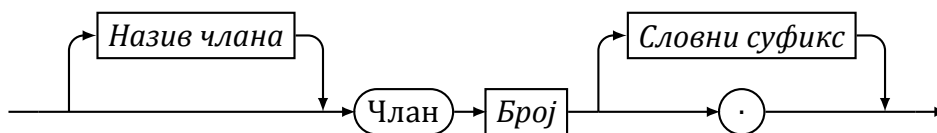
### 5.3 Структура блока изнад члана

У овом одељку ћемо прецизније описати структуру блока изнад члана, која је приказана на дијаграму 5.1. Као што видимо, изнад сваког члана се налази ознака члана, а ознаке ширих класификационих јединица могу и не морају да се појаве, при чему се поштује логичан редослед њиховог навођења. Такође је у оквиру ознаке члана назив члана опциони, што видимо на дијаграму 5.2.



Дијаграм 5.1: Структура блока изнад члана

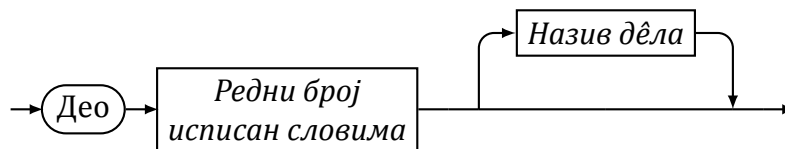
Једино што је увек присутно у блоку изнад члана јесте да се он завршава навођењем речи „Члан” која почиње великим словом и иза које следи број у арапском запису, а иза броја тачка или словни суфикс (уколико се наводи словни суфикс иза њега не иде тачка). Ова форма, када се појављује у засебном реду, је карактеристична за ознаку члана и не може се очекивати да ће се појавити у другом значењу.



Дијаграм 5.2: Ознака члана

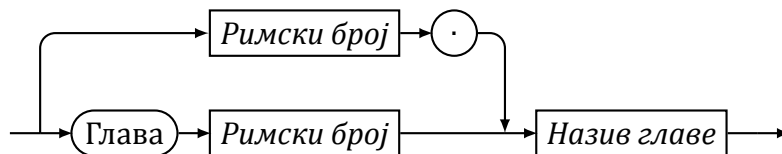
Назив члана, уколико се наводи, налази се у засебном реду изнад броја и почиње великим словом. На дијаграмима није истицана подела на редове.

За разлику од чланова, делови прописа се нумеришу редним бројевима који су исписани словима (дијаграм 5.3). Слично као за чланове, реч „Део” иза које следи нумеричка ознака дѐла наводи се у посебном реду, с тим да се евентуални назив дѐла наводи у следећем реду и пише се великим словима.



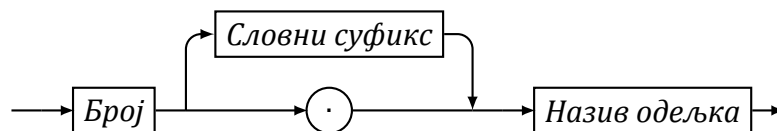
Дијаграм 5.3: Ознака дѐла

Главе се нумеришу римским бројевима (дијаграм 5.4), при чему су могућа два стила означавања глава: да се иза римског броја наведе тачка и затим назив у истом реду или да се испред римског броја дода реч „Глава”, а у том случају се назив наводи у следећем реду. Назив главе се пише великим словима и обавезан је.



Дијаграм 5.4: Ознака главе

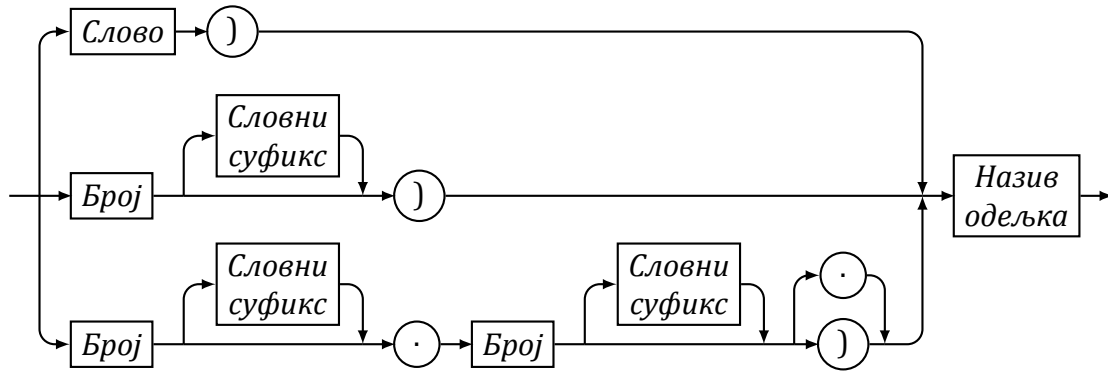
Главе се даље могу делити на одељке. Одељак се нумерише арапским бројем иза кога следи тачка или словни суфикс, а затим назив одељка у истом реду (дијаграм 5.5). Назив одељка почиње великим словом.



Дијаграм 5.5: Ознака одељка

Када постоји потреба да се одељак даље дели у пододељке, требало би да се као редна ознака у пододељцима користи слово иза кога следи заграда, међутим у пракси се уместо тога често користи број иза кога следи тачка или заграда, а томе у неким случајевима претходи број одељка раздвојен тачком. У

дијаграму 5.6 смо описали сва три случаја. У неким законима, попут Закона о привредним субјектима, се појављује додатан ниво ширих класификационих јединица испод нивоа пододељка, али је то већ изузетак за који ћемо сматрати да је одступање од правила.



Дијаграм 5.6: Ознака пододељка

Уобичајена пракса је да се не изоставља подела на главе, када год пропис има шире класификационе јединице. То практично значи да када се приликом израде прописа појави потреба за ширим класификационим јединицама, прво се уводи подела на главе, а затим се, према потреби, главе деле на одељке и пододељке и евентуално се групишу у делове. У актима [22], [23] изричито се наводи да пропис не може бити подељен само на делове, што значи да ако има поделу на делове, мора имати и поделу на главе. Са друге стране се изричито не истиче да највиша шира класификациона јединица треба да буде глава или део, али у корпусу правних текстова који је коришћен у истраживању (поглавље 3) није пронађен пример у коме је највиша класификациона јединица одељак, што је и логично.

Наиме, изузетак је пронађен у тексту Закона о трансфузиолошкој делатности који је преузет са сајта Народне скупштине и у коме се користе одељци као највише класификационе јединице; међутим, у тексту истог закона који је објављен у Службеном гласнику РС бр. 72/09, нумерација испред назива највиших класификационих јединица је промењена у римске бројеве, чиме су по форми то постале главе, али је истовремено нижи ниво класификационих јединица остао са словним редним ознакама, односно у форми пододељака. Можемо

претпоставити да се ова промена десила у фази правно-техничке редакције где се сматрало да унутрашња подела поменутог прописа треба да почне од нивоа главе.

Такође треба имати у виду да се нумерација чланова не везује за шире класификационе јединице (члан 1. је увек први члан у оквиру целог прописа, а не у оквиру неке шире класификационе јединице) и да се шире класификационе јединице не наводе приликом упућивања на одредбе прописа. Шире класификационе јединице формално немају нормативни значај, мада, самим тим што помажу разумевању прописа, у некој мери утичу на то како ће се пропис разумети, односно тумачити у реалним ситуацијама.

Наравно, ако је једина измена у неком пропису то што су одељци постали главе са истим називима и истим редним бројевима (само записаним римским уместо арапским бројевима), то нема никакав утицај на значење и разумевање тог прописа. Дакле, разлика у тексту Закона о трансфузиолошкој делатности између сајта Народне скупштине и Службеног гласника нема никакав утицај на примену тог закона, али представља занимљив пример за питање избора нивоа ширих класификационих јединица.

## **5.4 Унутрашња структура члана**

У 5.2 смо констатовали да је препознавањем блокова изнад члана уједно одређена подела на чланове, тако да је следећи корак препознавање унутрашње структуре члана.

Подсетимо се примера са слике 5.2 где је илустрован поглед на логичку структуру прописа. Примећујемо да у унутрашњој структури члана сваки пасус представља став или тачку, а у општем случају став, тачку, подтачку или алинеју.

Скрећемо пажњу да на слици 5.2 реч став означава тип пасуса, док став као класификациона јединица обухвата и тачке унутар става. Сличну разлику имамо и када речи тачка и подтачка означавају тип пасуса у односу на то када означавају класификациону јединицу. Реч алинеја као тип пасуса и као класификациона јединица увек означава исти део текста, јер алинеја у себи нема

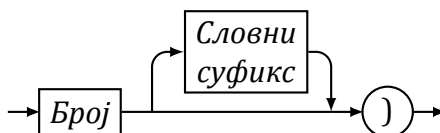
даљу поделу која би увела додатне пасусе.

Карактеристични детаљи који разликују пасусе који садрже став, тачку, подтачку и алинеју налазе се на почетку и на крају пасуса и приказани су у табели 5.1.

Табела 5.1: Карактеристични детаљи за типове пасуса у члану

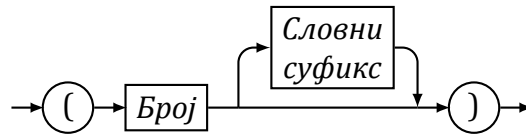
Тип	Почетак	Завршетак
став	било шта што није доле наведено	двотачка (када следе тачке) или тачка (када је крај става)
тачка	арапски број, опциони словни суфикс и заграда	двотачка (када следе подтачке), запета или тачка-запета (када следи наставак), „и” или „или” (на крају претпоследње) или тачка (на крају последње)
подтачка	у заградама арапски број са опционим словним суфиксом	двотачка (када следе алинеје), запета или тачка-запета (када следи наставак), „и” или „или” (на крају претпоследње) или тачка (на крају последње)
алинеја	цртица	запета или тачка-запета (када следи наставак), „и” или „или” (на крају претпоследње) или тачка (на крају последње)

Дакле, на почетку тачке и подтачке се налази одговарајућа нумерација (дијаграми 5.7 и 5.8), док алинеја почиње цртицом.



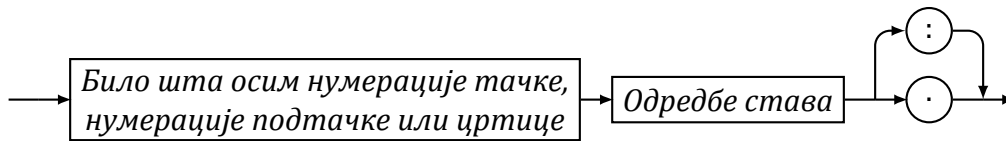
Дијаграм 5.7: Нумерација тачке



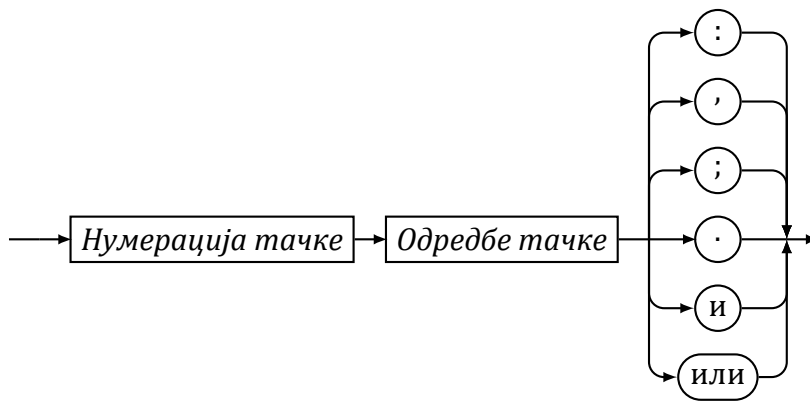


Дијаграм 5.8: Нумерација подтачке

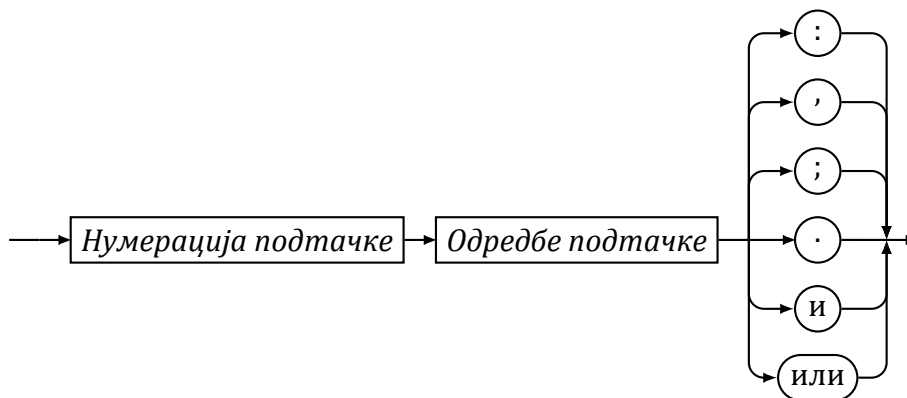
Структуре пасуса за став, тачку, подтачку, односно алинеју приказане су на дијаграмима 5.9, 5.10, 5.11 и 5.12.



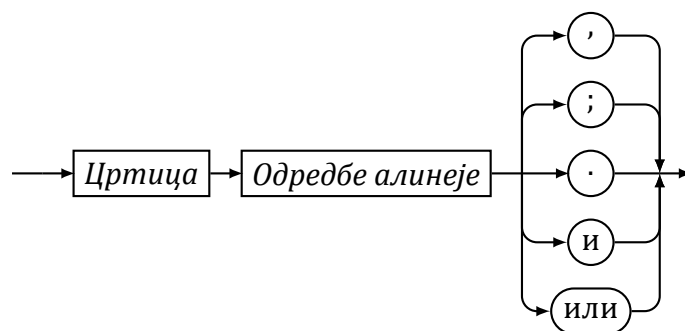
Дијаграм 5.9: Став



Дијаграм 5.10: Тачка



Дијаграм 5.11: Подтачка



Дијаграм 5.12: Алинеја

Примећујемо да немамо пуно карактеристичних детаља за препознавање елемената унутар члана, што се посебно односи на ставове. Да бисмо ипак смањили могућност нежељеног препознавања ставова, тачака, подтачака и алинеја, прво ћемо препознавати блокове изнад члана (чија структура оставља мању могућност нежељеног препознавања), па ћемо их на одређен начин издвојити и након тога преостале пасусе препознавати као ставове, тачке, подтачке и алинеје. Овим смо ушли у тематику метода синтаксне анализе, чему је посвећено наредно поглавље.

## 5.5 Подела на пасусе

У претходно израженим синтаксним правилима за елементе логичне структуре прописа више пута се показала значајна подела на пасусе, односно редове, мада у самим дијаграмима та чињеница није изражавана.

Када се текст обрађује *Unitex*-ом, у фази претходне обраде се врши сегментација реченица и сваки крај реченице се означава са  $\{S\}$ .

С обзиром на специфичну структуру текста прописа, уобичајена правила за сегментацију реченица не дају задовољавајуће резултате. У примеру са слике 5.2 можемо видети да се уобичајено означавање краја реченице тачком појављује једино на самом крају (тачка код нумеричке ознаке члана је део нумеричке ознаке, у случају када нумеричка ознака има словни суфикс тачка се не појављује). То значи да би члан са блоком изнад члана, ставом и тачкама, применом уобичајених правила, у многим случајевима био сегментиран као једна

реченица.

Како *Unitex* врши сегментацију реченица у раној фази обраде, када још не знамо логичку структуру текста закона, правила за сегментацију не можемо прилагођавати деловима логичке структуре. Када узмемо у обзир и чињеницу да се у законима ретко дешава да један пасус садржи више реченица, долазимо до тога да постављање ознаке  $\{S\}$  на крај сваког пасуса даје прихватљивију поделу на реченице него што то чине уобичајена правила. При томе крајем пасуса подразумевамо сваки прелазак у нови ред који није последица преламања текста ради уклапања у ширину ступца.

С обзиром да је, са друге стране, у анализи текста прописа у одређеним случајевима значајна подела на пасусе, као што је то случај код препознавања назива и садржаја класификационих јединица, коначно смо се определили да ознака  $\{S\}$  стоји на крају пасуса.

## 5.6 Синтаксни графови

Синтаксни графови израђени на основу дијаграма из овог поглавља приказани су у прилогу А (одељак А.4). За разлику од упућујућих фраза, ови графови нису толико погодни за анализу путем конкорданци над корпусом прописа, јер обухватају читаве пасусе у једној јединици препознавања, па им је основна сврха да се користе у синтаксној анализи прописа, што је тема наредног поглавља.

## Поглавље 6

### Синтаксна анализа

Задатак синтаксне анализе текста прописа је да се у датом тексту ефективно одреди структура која произилази из синтаксних правила, што ћемо илустровати примером једне упућујуће фразе.

На слици 6.1 је приказан могућ резултат синтаксне анализе упућујуће фразе „члан 13. ст. 2, 4. и 7. овог закона”. Синтаксно дрво са слике 6.1 директно прати редослед примене графова из прилога А: сваки чвор дрвета који је представљен правоугаоном кућицом садржи назив графа којим се препознаје грана дрвета испод тог чвора, док чворови представљени овалним кућицама садрже делове текста.

Сврха синтаксне анализе је да припреми текст за даљу обраду и из те перспективе се може рећи да синтаксно дрво на слици 6.1 сувише одражава детаље синтаксних правила. Из перспективе значења упућујуће фразе, што је предмет даље обраде, потребно је да истакнемо елементе који су семантички релевантни, па синтаксно дрво можемо свести како је приказано на слици 6.2.

Друга верзија синтаксног дрвета је у једном детаљу чак изражајнија: број је издвојен од знака интерпункције иза броја, јер знак интерпункције није семантички релевантан.

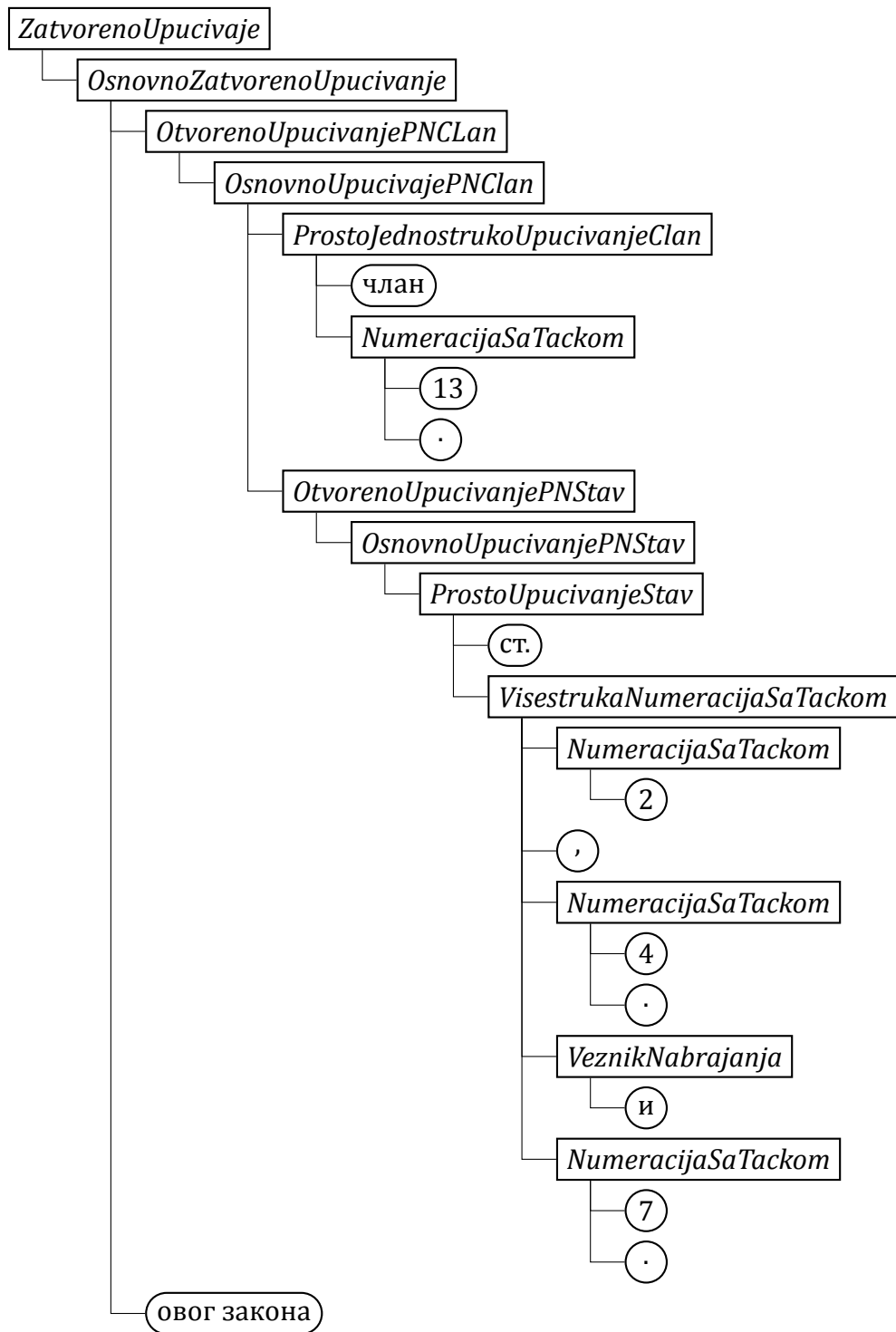
Дрво са слике 6.2 можемо представити у *XML*<sup>1</sup> формату као што је приказано на слици 6.3.

У оквиру овог истраживања ће циљ синтаксне анализе текста прописа бити

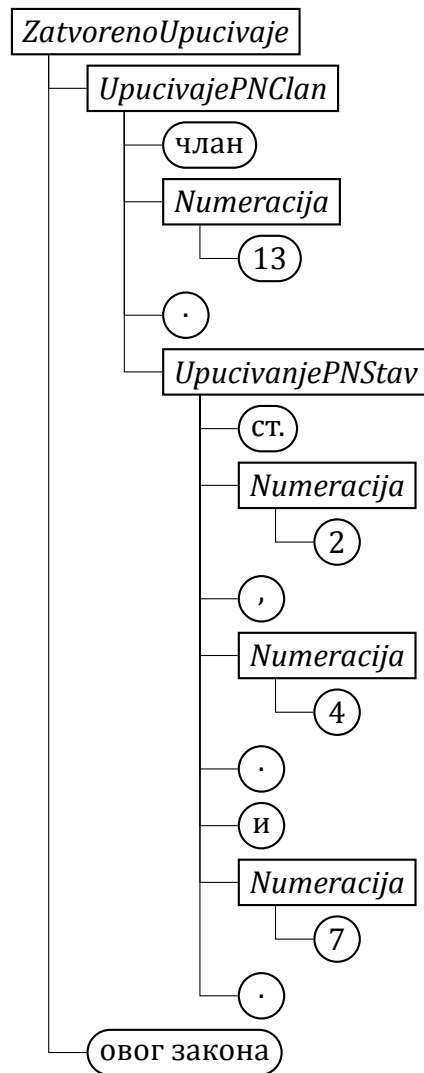
---

<sup>1</sup>*Extensible Markup Language*, <http://www.w3.org/XML/>

да произведе текст са уметнутим XML обележјима попут примера са слике 6.3.



Слика 6.1: Синтаксно дрво за „члан 13. ст. 2, 4. и 7. овог закона”



Слика 6.2: Сведено синтаксно дрво за „члан 13. ст. 2, 4. и 7. овог закона”

```

<ZatvorenoUpucivanje>
  <UpucivanjePNClan>члан <Numeracija>13</Numeracija>.
    <UpucivanjePNStav>ст. <Numeracija>2</Numeracija>,
      <Numeracija>4</Numeracija>. и <Numeracija>7</Numeracija>.
    овог члана
  </UpucivanjePNClan>
</ZatvorenoUpucivanje>
  
```

Слика 6.3: Синтаксно дрво у XLM формату

## 6.1 Лексичке етикете у синтаксној анализи

Систем *Unitex* у оквиру претходне обраде токенизује текст, што је описано у пододељку 3.1.4. Токени могу бити секвенце узастопних слова, појединачни карактери који нису слова и посебне врсте токена у шта спадају лексичке етикете. Лексичке етикете дају могућност да неки део текста експлицитно обележимо као токен и доделимо му синтаксна, семантичка и флективна својства попут оних који се наводе у електронском речнику. Такав токен називамо лексичком етикетом, а у следећој реченици дајемо пример лексичке етикете (у витичастим заградама) којом смо обележили једну реч:

С првим {пролећем,пролеће.N+Ek:ns6q} настављамо радове.

Лексичке етикете имају структуру сличну редовима у електронском речнику у *LADL*-формату (одељак 3.3) јер изражавају исту врсту својстава, при чему се у случају коришћења електронског речника та својства додељују проналажењем одговарајућих ставки у речнику, док се код лексичких етикета одговарајућа својства експлицитно наводе у тексту. У конкретном примеру лексичка етикета означава да реч „пролећем” представља облик именице „пролеће” (а не глагола „пролетати”), синтаксни кодови говоре да је у питању именица (N) у екавском изговору (Ek), а од флективних својстава је наведен средњи граматички род (n), једнина (s), инструментал (6) и граматички неживо (q). Овој речи су додељени исти кодови који се налазе у електронском речнику, с тим да би електронски речник препознао и прво лице једнине глагола „пролетати”.

У графовима из прилога А на појединим местима примећујемо трансдукторске излазе (текст који стоји испод појединих кућица у графовима), које смо у примерима из претходних поглавља занемаривали. Када у анализу текста укључимо опцију да се трансдукторски излази умећу на месту где се појављују и произведемо конкорданце за упућујуће фразе, добијамо резултат попут оног на слици 6.4 (у примеру са слике се користи Закон о привредним субјектима).

Прва од линија конкорданци које се виде на слици 6.4 односи се на упућујућу фразу „чланом 50. став 1. тачка 1) овог закона”. На слици 6.5 је приказана





```

{  {чланом {50,.tagNumeracija}.
    {став {1,.tagNumeracija}.
      {тачка {1,.tagNumeracija}},.tagUpucivanjePNTacka}
    ,.tagUpucivanjePNStav},
  .tagUpucivanjePNClan}
  овог закона
,.tagZatvorenoUpucivanje}

```

Слика 6.5: Препозната упућујућа фраза са уметнутим трансдукторским излазима

препознаје као један токен, док сличну могућност нема за XML елементе.

Овакво коришћење лексичких етикета је специфично по томе што синтаксне конструкције претварамо у посебне лексичке јединице. Када *Unitex* у наредном пролазу токенизује текст, овакве лексичке етикете препознаје као један токен, то јест не види више њихову угњеждену структуру. На тај начин ћемо у сваком пролазу да „запакујемо” делове које смо тим пролазом препознали и тек на крају ћемо све то да „препакујемо” у *XML*.

Оваква примена више трансдуктора, где резултат примене једног постаје улаз следећег, називамо каскадом трансдуктора. За каскаде трансдуктора користимо алат *CasSys* [73] који је интегрисан у *Unitex*.

Сваки трансдуктор у каскади је представљен графом који садржи (непосредно или кроз графове које укључује) одређене трансдукторске излазе. За синтаксну анализу правних текстова у овом истраживању смо користили каскаду од четири трансдуктора, као што је приказано у табели 6.1. Спецификација трансдукторских каскада у формату који користи *CasSys* је дата у прилогу Б.

Табела 6.1: Каскада трансдуктора за синтаксну анализу прописа

	Граф	део прилога А
1.	BlokIznadClana	пододељак А.4.1
2.	Upucivanje	одељак А.3
3.	UzeKlasJedinice	пододељак А.4.2
4.	PocetakAkta	пододељак А.4.3

Као што смо истакли на крају одељка 5.4, у првом пролазу ћемо обрађивати блокове изнад члана да бисмо смањили могућност нежељеног препознавања ставова, тачака, подтачака и алинеја. Из сличног разлога упућивања обрађујемо након блокова изнад члана, јер ознака члана има форму као упућивање на члан (о чему је било речи на крају одељка 4.8). Упућивања обрађујемо пре ужих класификационих јединица зато што се налазе унутар њих. Након прва три пролаза сви чланови са својим садржајем и ознакама изнад члана треба да су препознати и спаковани у одговарајуће лексичке етикете. На крају остаје да се обради део који се налази на почетку који обухвата назив прописа.

Алат *CasSys* коначан резултат обраде производи у два формата: у основном и у *XML* формату.

Лексичке етикете су у основном формату записане онако како су приказане на сликама 6.4 и 6.5 с том разликом да је код угњеждених лексичких етикета испред витичастих заграда и других карактера који имају посебно значење додата обрнута коса црта. Ове обрнуте косе црте заправо сакривају угњеждене лексичке етикете, јер *Unitex* у ствари не подржава угњежденост лексичких етикета и све заједно види као једну велику лексичку етикету, што смо већ помињали раније у овом одељку.

Када *CasSys* формира резултат обраде у *XML* формату, тада лексичку етикету са слике 6.5 записује као што је приказано на слици 6.6.

Дакле, преостало је да резултат каскадне обраде претворимо у циљни *XML* формат који је приказан на слици 6.3 било из основног формата или из *XML* формата са слике 6.6. То можемо учинити на два начина:

1. да додамо још пролаза у каскаде, чиме ће се „распакивати” етикете и уметати *XML* тагови или
2. да користимо резултат у форми *XML*-а и да одговарајућим алатом трансформисемо *XML* формат лексичких етикета у циљни *XML* формат.

*Unitex* са *CasSys*-ом пружа све неопходне могућности да би се реализовао први начин, али то у овом раду нећемо детаљније објашњавати јер се нисмо определили за ту опцију.

Други начин такође не искључује могућност да и за тај задатак употребимо *Unitex*. И поред тога што *Unitex* свакако није најбољи алат за трансформацију *XML*-а, његова предност може бити у томе што га већ користимо и не морамо да уводимо додатну технологију у обраду.

Ми смо се руководили принципом избора најбољег алата за дати задатак и изабрали *XSLT*<sup>2</sup> технологију за трансформисање *XML*-овског резултата *CasSys*-а у циљни *XML* формат. Опис трансформације у *XSLT* нотацији дат је у прилогу В.

```

<csc>
  <form>
    <csc>
      <form>чланом
        <csc><form>50</form><code>tagNumeracija</code></csc>.
      <csc>
        <form>став
          <csc><form>1</form><code>tagNumeracija</code></csc>.
          <csc>
            <form>тачка
              <csc><form>1</form><code>tagNumeracija</code></csc>)
            </form>
            <code>tagUpucivanjePNTacka</code>
          </csc>
        </form>
        <code>tagUpucivanjePNStav</code>
      </csc>
    </form>
    <code>tagUpucivanjePNClan</code>
  </csc> ОВОГ ЗАКОНА
</form>
<code>tagZatvorenoUpucivanje</code>
</csc>

```

Слика 6.6: Лексичка етикета у *XLM* формату са угњежденим етикетама

Записивањем синтаксног дрвета у планираном *XML* формату окончали смо синтаксну анализу.

<sup>2</sup>*Extensible Stylesheet Language Transformations*, <http://www.w3.org/style/XSL/>

## Поглавље 7

### Контролни документ

Синтаксна анализа је основ за разне врсте обрада правних текстова, а у оквиру овог рада ћемо се посебно бавити формирањем контролног документа. Сврха контролног документа је да помогне у откривању техничких пропуста при изради прописа, тако што је у контролном документу:

- прегледно приказано шта је и на који начин препознато и
- указано на неспорне пропусте.

Контролни документ пре свега треба да буде уредно форматиран у односу на препознату логичку структуру. Уколико неки делови текста остану непрепознати, они ће у контролном документу бити посебно истакнути. Већ на овом нивоу контролни документ остварује своју сврху тиме што ће део текста који је на неочекиван начин форматиран указивати на пропусте који су довели до тога да логичка структура није препозната како је очекивано.

Контролни документ такође треба на прегледан начин да прикаже детаље мреже упућивања. Под мрежом упућивања подразумевамо сва потенцијална одредишта упућивања, све упућујуће фразе и њихове међусобне релације. При истицању неправилности у мрежи упућивања, поред упућивања на непостојеће одредиште, треба истаћи и пропусте у редоследу нумерације.

Генерисање контролног документа представља практичан алат који може помоћи при изради и редакцији закона, као и за анализу раније израђених закона. Са друге стране, генерисање контролног документа представља импле-

ментацију теоријских поставки из претходних поглавља и даје могућност да се оне експериментално провере на конкретним примерима прописа.

## 7.1 Основна техничка и технолошка опредељења

Контролни документ генеришемо у оквирима следећих почетних техничких захтева:

1. контролни документ користи *XML* или *XHTML*<sup>1</sup> формат тако да може да се прикаже у веб прегледачима;
2. за дефинисање изгледа документа се користи *CSS*<sup>2</sup> спецификација која је уграђена у документ тако да документ чини само једна датотека;
3. *JavaScript* нећемо користити уколико не буде неопходно (а показало се да није било неопходно);
4. коначни документ ћемо формирати мењајући *XML* датотеку која је резултат синтаксне анализе, са идејом да измене буду што мање.

За генерисање контролног документа на основу резултата синтаксне анализе користили смо програмски језик *Java* и стандардну подршку *Java* платформе за обраду *XML* докумената.

Основни елементи изгледа контролног документа постигнути су само специфицирањем изгледа путем *CSS*-а јер није било потребе за изменама у структури *XML*-а.

Измене у структури *XML* елемената се свде искључиво на уметање додатних *XML* елемената, при чему се за неке могућности, као што су хиперлинкови, користе поједини *XHTML* елементи. Сви модерни веб прегледачи подржавају овакву могућност уградње *XHTML*-а у неки *XML* формат.

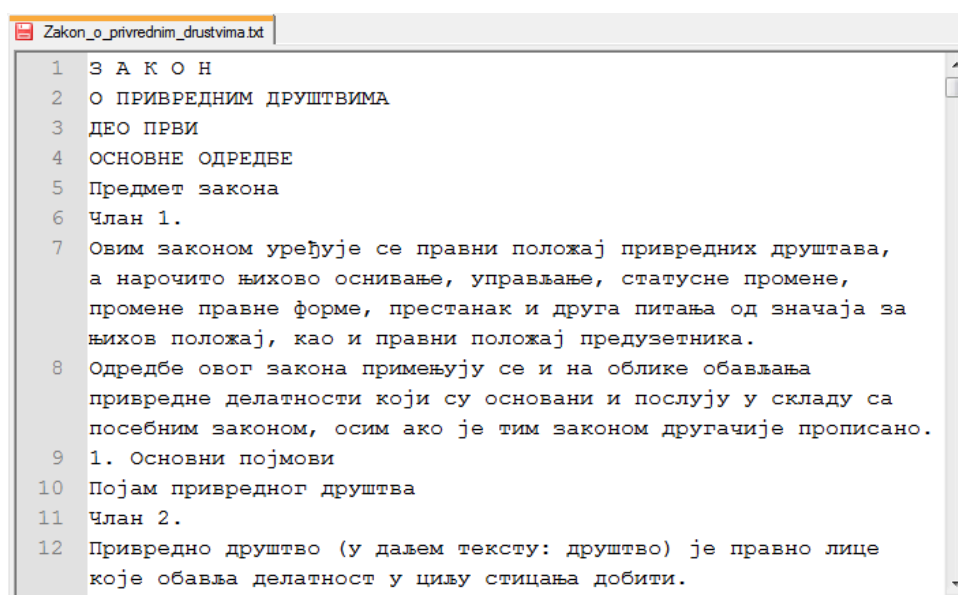
За генерисану датотеку ћемо у називу користити наставак *.html* из практичних разлога јер датотеку треба отворати из апликација које подржавају *XHTML*, мада би на основу садржаја формално одговарала екстензија *.xml*.

<sup>1</sup>*Extensible HyperText Markup Language* <http://www.w3.org/TR/xhtml11/>

<sup>2</sup>*Cascading Style Sheets* <http://www.w3.org/style/CSS/>

## 7.2 Изглед и функционалност

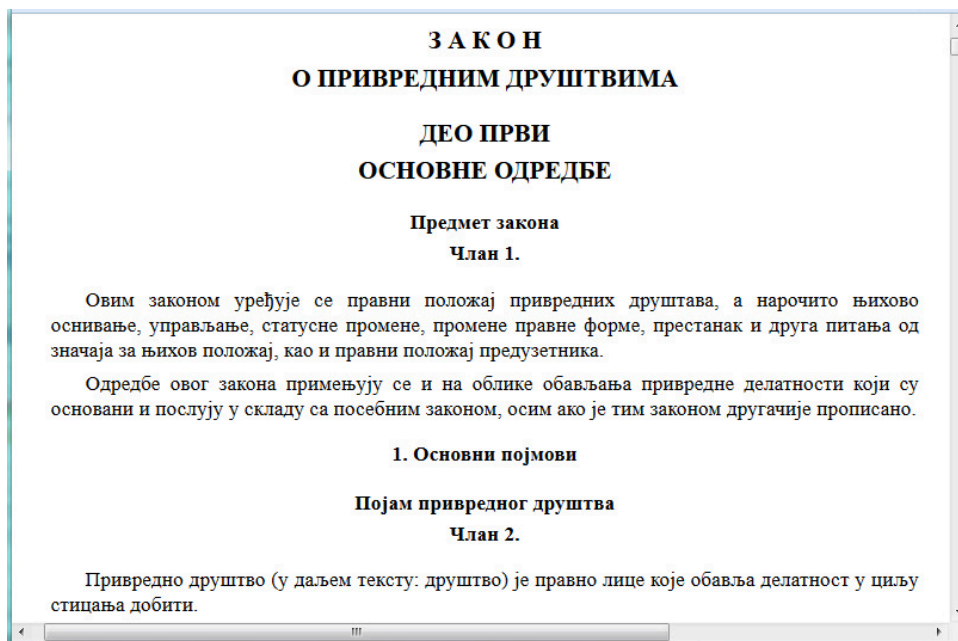
Изглед и функционалност контролног документа описаћемо на примеру Закона о привредним друштвима. На слици 7.1 је приказано како у уређивачу текстуалних датотека изгледа текст закона пре уласка у обраду, док је на слици 7.2 приказан изглед одговарајућег дела контролног документа у веб прегледачу.



Слика 7.1: Полазни текст који је предмет обраде

Видимо како је на основу препознате логичке структуре одређено поравнање редова, величина и подебљаност слова. Промена величине и подебљаности слова у називима ширих класификационих јединицама се не захтева званичним методолошким правилима, али је то овде додато због прегледности.

У оквиру текста закона се појављује и неколико пасуса који нису успешно класификовани, па су обојени црвеном бојом, као на примерима са слике 7.3. На слици су приказана два случаја из различитих делова текста. Прво имамо пасус који се завршава са тачком-зарез уместо тачком и због тога није препознат као став. У другом случају примећујемо ознаку пододељка иза чега следи ознака која по свој прилици треба да се односи на ширу класификациону јединицу испод нивоа пододељка („подпододељак”), а та врста ширих класификационих јединица није предвиђена. Приметимо да је и ознака пододељка остала не-



Слика 7.2: Изглед контролног документа

препозната јер се та ознака препознаје искључиво у склопу блока изнад члана (види одељак 5.3 овог рада), па због проблематичне ознаке која следи, све заједно није могло бити препознато као део блока изнад члана.

Новчаном казном од 20.000 до 150.000 динара казниће се за прекршај физичко лице ако обавља делатност а то не чини кроз прописане облике обављања привредне делатности, укључујући и привредно друштво и предузетника;{S}

7.1. Скупштина{S} 7.1.1. Састав и делокруг{S}

Слика 7.3: Примери пасуса који нису успешно класификовани

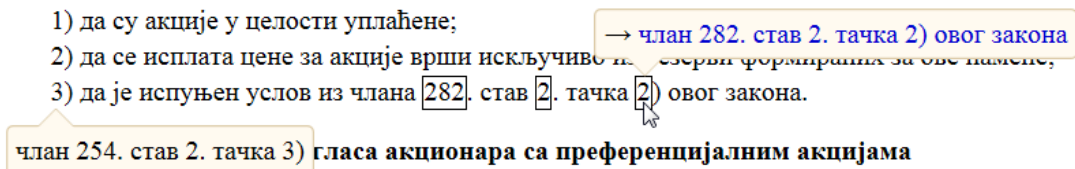
Када у оквиру веб прегледача курсор миша позиционирамо изнад неког пасуса који може бити одредиште упућивања, испод тог пасуса се појављује облачић који говори где се налазимо. Пример је дат на слици 7.4, где се види да се

Актом друштва из става 4. овог члана се:  
 1) као пословна тајна може одредити само податак који испуњава услове из става 3. овог члана и  
 2) не могу као пословна тајна одредити сви подаци који се односе на пословање друштва.  
 (члан 72. став 5. тачка 2) овог члана може бити производни, технички, технолошки, финансијски подаци, резултат истраживања, као и документ, формула, цртеж, објекат, метод, поступак, обавештење или упутство интерног карактера и слично.

Слика 7.4: Облачић који говори где се налазимо

налазимо у члану 72. став 5. тачка 2).

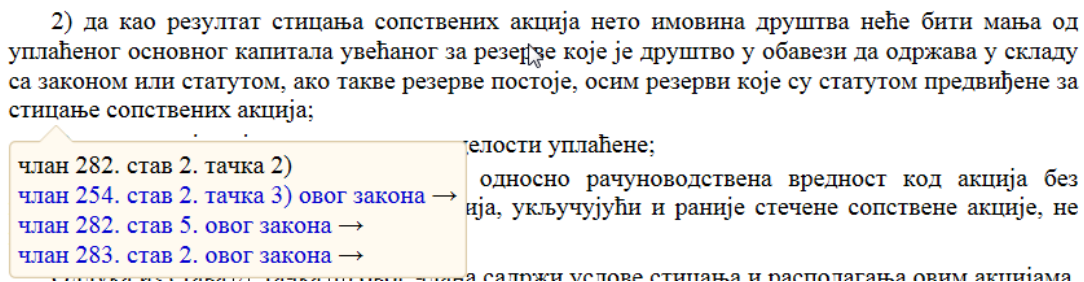
На истој слици примећујемо да су бројеви у упућивањима уоквирени. На слици 7.5 видимо шта се дешава када курсор позиционирамо изнад тако уоквиреног броја.



Слика 7.5: Облачић за упућивање

У облачићу се појављује хиперлинк на који можемо да кликнемо да бисмо прешли на члан 282. став 2. тачку 2) која је приказана на слици 7.6. Видимо да су у облачићу који показује где се налазимо наведена места са којих се упућује на место на коме се налазимо. То су такође хиперлинкови.

Приметимо да су на слици 7.5 уоквирена сва три броја у упућујућој фрази. Ако бисмо се позиционирали изнад средњег броја приказало би се упућивање на одговарајући став и слично би се изнад броја 282 приказало упућивање на члан под тим бројем.



Слика 7.6: Облачић показује места са којих се упућује на дату тачку

Код упућивања на непостојеће одредиште уоквирени број у упућивању добија црвену позадину, као што видимо на примеру са слике 7.7.

Приметимо да је црвеном позадином (у црно-белој штампи тамнија позадина) означен само број тачке, што значи да члан 259. став 2. постоји (као и став 1). Позиционирањем курсора изнад броја који означава став добићемо облачић са хиперлинком на то одредиште (као на слици 7.5). Ако бисмо на тај начин отишли на члан 259. став 2, видели бисмо да тај став не садржи тачке,



преос не постоји: аљем року од 30 дана од дана пријема тог захтева  
 испл члан 259. став 2. тачка 1) овог закона ти.  
 Тржана вредност из ст. 4. и 5. утврђује се у складу са одредбама члана 259. став 1. и став  
 2. тачка 1) овог закона.  
 из става 3. овог члана долази до повећања основног капитала, односно до  
 члан 257. став 6. капитала, а које у оба случаја не прелази 1% основног капитала, друштво није

Слика 7.7: Упућивање на непостојеће одредиште

али да тачке садржи наредни став. Може се претпоставити да је приликом израде закона у неком тренутку уметнут став 2. чиме је претходни став 2. постао став 3, а да при томе није ажурирано упућивање чије је одредиште унутар помереног става.

На текућу класификациону јединицу увек треба да се упућује повратним упућивањем (овај члан, овај став и сл.) тако да када се на текућу класификациону јединицу упућује на други начин, можемо претпоставити да је то резултат пропуста на који треба указати у контролном документу, као у примеру са слике 7.8. Када се пређе мишем преко броја 5, у облачићу се појављује одговарајућа порука.

акција призна и исплати исту вредност акцији упућује на овај став нице да ли су ти акционари поднели тужбу из става 1. овог члана.  
 Ако друштво не поступи у складу са ставом 5. овог члана у року за извршење пресуде из става 4. овог члана, сваки несагласни акционар може тужбом надлежном суду тражити исплату разлике до пуне вредности акција, односно вредност акција која је утврђена том пресудом.

Слика 7.8: Бројем се упућује на текућу класификациону јединицу

Пример са слике би могао да буде резултат брисања неког од претходних ставова у току израде закона (или спајања нека два става), при чему није ажурирано упућивање.

Поред наведеног, имплементирана је и могућност указивања на грешке у редоследу нумерација класификационих јединица, а у тексту Закона о привредним друштвима који смо овде узели као пример такве грешке нису идентификоване. Међутим, да бисмо илустровали и ову могућност, у тексту смо унели две измене: члан 311. је постао члан 311а, а тачка 2) из тог члана је постала тачка 3).

На почетку контролног документа налази се преглед свих идентификована

них грешака, уколико их има, што је приказано на слици 7.9.

### **Идентификовани су технички пропуссти у тексту!**

**На шест места се упућује на непостојећу одредбу:**

1. члан 257. став 6. → члан 259. став 2. тачка 1) овог закона
2. члан 297. став 2. → члан 297. став 2. тачка 1) овог закона
3. члан 297. став 2. → члан 297. став 2. тачка 9) овог закона
4. члан 322. став 4. → члан 322. став 2. тачка 2) овог закона
5. члан 534. став 3. → члан 534. став 1. тачка 3) овог закона
6. члан 588. став 1. тачка 6) → члан 89. став 10. овог закона

**На три места се бројем упућује на текућу класификациону јединицу:**

1. члан 297. став 2.
2. члан 415. став 4.
3. члан 476. став 5.

**На два места нумеричка ознака класификационе јединице није по реду:**

1. члан 311а.
2. члан 311а. став 1. тачка 3)

**На 15 места део текста није успешно класификован:**

1. иза члана 198. став 6.
2. иза члана 200. став 1. тачка 26)
3. иза члана 215. став 1.
4. иза члана 216. став 1.

Слика 7.9: Списак идентификованих грешака

## Поглавље 8

# Израђени софтверски алати

Током истраживања је израђен одређен број софтверских алата од којих су неки усмерени ка имплементацији резултата истраживања, а неки више ка практичном коришћењу у истраживачком раду.

За имплементацију софтверских алата је претежно коришћен програмски језик *Java*<sup>1</sup>. Поред могућности које пружа програмски језик и стандардна библиотека, односно *Java* платформа, укључивана су и поједина независна решења, што ћемо прецизније навести када будемо говорили о појединим израђеним алатима.

У већини алата се користи интеграција са системом *Unitex*. Интеграција се односи на коришћење постављеног подешавања из система *Unitex*, ишчитавање и исписивање датотека у форматима које користи *Unitex*, извршавање команди из дела *Unitex*-а који је имплементиран у програмском језику *C++* и коришћење, независно од корисничког интерфејса, појединих класа из дела *Unitex*-а имплементираних у програмском језику *Java*.

Приликом израде алата смо се руководили начелом шире употребљивости, то јест могућности каснијег коришћења и изван овог истраживања, према чему алате можемо поделити у три групе:

1. без шире употребљивости — алати који су по својој природи везани за специфичан случај коришћења, као што је преузимање материјала са конкретног веб-сајта, где је тежиште израде алата на детаљима коју су специфични за

---

<sup>1</sup><http://www.oracle.com/technetwork/java/>

дати веб-сајт;

2. потенцијална шира употребљивост — алати који су израђени тако да су употребљиви и изван истраживања, али није развијен готов производ путем кога би тај алат био ефективно доступан ширем кругу корисника (овој групи припада већина израђених алата);
3. ефективна шира употребљивост — алати који су ефективно доступни ширем кругу корисника путем одговарајућег готовог производа, као што су унапређења система *Unitex* која су доступна кроз нове верзије *Unitex*-а.

Развој софтверског производа је посебана врста активности која обухвата решавање многих пратећих питања, као што су: документација, тестирање на разним платформама, инсталациона процедура и инсталациони пакет, веб-сајт производа и успостављање канала обостране комуникације са корисницима. При томе је потребно систематски уважавати потребе циљне групе корисника током целокупног животног века производа. За некомерцијалне производе отвореног кода треба успоставити механизме сарадње унутар заједнице која учествује у изради производа, а за комерцијалне производе треба имати тржишну стратегију и пословни план. Све што смо навели можемо решавати и на неком елементарном нивоу, али никако не можемо занемарити уколико имамо намеру да производ живи у пракси.

Када се све узме у обзир, није било потребно да се развојем софтверских производа бавимо у оквиру овог истраживања, али то може бити предмет будућих активности.

## 8.1 Претходна обрада корпуса

Када се раније у овом раду говорило о претходној обради корпуса (одељак 3.1), истакнути су кораци претходне обраде за које је израђен посебан софтверски алат:

1. преузимање са веб-сајта Народне скупштине Републике Србије датотека у *DOC* формату са текстовима објављених закона, као и преузимање додат-

них података о закону (пре свега наслова који су приказани у табели на веб-страни);

2. конверзија из *DOC* формата у чист текст;
3. спајање текстова закона у једну датотеку у формату чистог текста уз анонотацију на библиографском нивоу (означавање шта припада ком закону);

Алат који имплементира наведене кораке претходне обраде има два дела: преузимање материјала са веб-сајта (покрива први корак) и обрада преузетог материјала (покрива други и трећи корак).

### 8.1.1 Преузимање лексичког материјала са веб-сајта

За преузимање лексичког материјала са веб-сајта Народне скупштине Републике Србије израђен је програм у програмском језику *Java* и при томе је коришћен *Jsoup HTML Parser*<sup>2</sup>. *Jsoup* подржава преузимање садржаја путем *HTTP* протокола и анализу преузетих *HTML* страна. На слици 8.1 је приказан пример методе у програмском језику *Java* која преузима све *DOC* датотеке са текстовима закона које су излистане за задатој веб страни. Изглед једне такве веб-странице је дат на слици 8.2.

Дати пример је поједностављена верзија методе која се стварно користи. У примеру је искључен део кода који из *HTML*-а издваја додатне податке о закону и формира одговарајућу *XML* датотеку. У линији 5 (слика 8.1) се успоставља веза са веб-сервером и преузима *HTML* страна. Затим се проналазе редови из одговарајуће *HTML* табеле, на основу задатог упита (линија 8). На слици 8.2 је дат изглед дела табеле и *HTML* запис једног реда у табели, то јест изглед одговарајућег елемента *TR*. Упит из линије 8 у слободној интерпретацији значи: пронађи све елементе *DIV* код којих атрибут *class* има вредност *delegate\_list* и затим у оквиру тих елемената пронађи све елементе *TR* који у себи садрже елемент *TD* код кога атрибут *class* има вредност *lawtitle*. Резултат упита је листа пронађених елемената *TR*, попут оног са слике 8.2. Након тога се у сваком пронађеном елементу *TR* одређује последњи елемент *TD* и у њему последњи еле-

---

<sup>2</sup><http://jsoup.org>

```

1 public static void grab(String pageUrl, String folder)
2     throws IOException {
3     final String userAgent = "Mozilla/5.0...";
4
5     Document doc = Jsoup.connect(pageUrl).userAgent(userAgent).
6         timeout(5000).maxBodySize(10 * 1024 * 1024).get();
7
8     Elements trs = doc.select("div.delegate_list tr:has(td.lawtitle)");
9
10    for (Element tr : trs) {
11        Elements tds = tr.select("td");
12        String link = tds.last().select("a").last().attr("href");
13        URL fileUrl = new URL(new URL(pageUrl), link);
14        downloadFile(fileUrl, folder);
15    }
16 }

```

Слика 8.1: Метода за преузимање докумената са одређене веб-странице

мент *A*, што је линк који упућује на *DOC* датотеку, из кога још треба узети атрибут *href* (линија 12). Након што је одређен линк, преузима се датотека у задат локални директоријум (линија 14).

Назив	Датум	Предлагач	Преузми
ПРЕДЛОГ ЗАКОНА О ПОТВРЂИВАЊУ СПОРАЗУМА ИЗМЕЂУ ВЛАДЕ РЕПУБЛИКЕ СРБИЈЕ И ВЛАДЕ РЕПУБЛИКЕ МОЛДАВИЈЕ О САРАДЊИ У ОБЛАСТИ ОДБРАНЕ	18. јун 2014.	Влада Републике Србије	PDF, DOC
ПРЕДЛОГ ЗАКОНА О ПОТВРЂИВАЊУ СПОРАЗУМА ИЗМЕЂУ ВЛАДЕ РЕПУБЛИКЕ СРБИЈЕ И ВЛАДЕ РУСКЕ ФЕДЕРАЦИЈЕ О ВОЈНОЈ САРАДЊИ	18. јун 2014.	Влада Републике Србије	PDF, DOC

```

<tr>
  <td align="left" class="lawtitle">
    <strong>
      <a href="/upload/.../2034-14.pdf">Предлог закона о
      потврђивању Споразума између Владе Републике Србије и
      Владе Републике Молдавије о сарадњи у области одбране</a>
    </strong>
  </td>
  <td nowrap>18. јун 2014.</td>
  <td>Влада Републике Србије</td>
  <td nowrap>
    <a href="/upload/.../2034-14.pdf">PDF</a>,
    <a href="/upload/.../2034-14.doc">DOC</a>
  </td>
</tr>

```

Слика 8.2: Изглед табеле на веб-страни и опис једног реда табеле у *HTML*-у

Преузета датотека је у неким случајевима *ZIP* архива, па је након преузимања потребно и распаковати архиву, за шта се користи подршка за *ZIP* формат из стандардне библиотеке *Java* програмског језика.

### 8.1.2 Обрада преузетог лексичког материјала

Обрада преузетог лексичког материјала обухвата селекцију докумената у складу са условима који су дати на почетку поглавља 3, затим издвајање чистог текста из сваког од селектованих докумената, додавање лексичке етикете за ознаку почетка новог закона и спајање у један документ.

За потребе ове обраде је израђен програм у програмском језику *Java*. За манипулацију датотекама у *DOC* формату је коришћен *Microsoft Word* који је програмски контролисан из програмског језика *Java* путем *Automation*<sup>3</sup> интерфејса. Технологија *Automation* (раније позната као *OLE Automation*) заснована је на систему *Microsoft Component Object Model (COM)*<sup>4</sup> и користи се да би једна апликација омогућила приступ из других апликација према својим функционалностима и ресурсима. Да бисмо из програмског језика *Java* приступили систему *COM*, а тим путем и *Microsoft Word*-у кроз *Automation* интерфејс, користимо решење *JACOB*<sup>5</sup> као мост између *Java*-е и система *COM*. За читање *XML* датотека у којима се налазе додатни подаци о закону, користи се подршка за *XML* из стандардне библиотеке програмског језика *Java* и интерфејс *Document Object Model (DOM)*<sup>6</sup> за приступ структури и садржају *XML* докумената.

Чисти текст из *Microsoft Word* документа можемо добити на барем три начина, од кога су прва два доступна и путем графичког корисничког интерфејса *Word*-а и кроз *Automation* интерфејс, а трећи начин је могућ само кроз *Automation* интерфејс:

1. коришћењем *Save As* опције и избором чистог текста (*Plain Text*) као типа датотеке;

---

<sup>3</sup><http://msdn.microsoft.com/en-us/library/dt80be78.aspx>

<sup>4</sup><http://msdn.microsoft.com/en-us/library/ms694363.aspx>

<sup>5</sup><http://sourceforge.net/projects/jacob-project/>

<sup>6</sup><http://www.w3.org/DOM/>

```

1 public class Doc2Txt {
2
3     private ActiveXComponent xComponent;
4     private Clipboard clipboard;
5     public static final int wdDoNotSaveChanges = 0;
6
7     public Doc2Txt() {
8         ComThread.InitSTA();
9         xComponent = new ActiveXComponent("word.Application");
10        Toolkit toolkit = Toolkit.getDefaultToolkit();
11        clipboard = toolkit.getSystemClipboard();
12    }
13
14    public void convert(String docFile, String txtFile, LawMeta meta)
15        throws Exception {
16        Dispatch documents = xComponent.getProperty("Documents")
17            .toDispatch();
18        document = wdCom.openDocument(documents, docFile);
19        Dispatch selection = xComponent.getProperty("Selection")
20            .toDispatch();
21        Dispatch.call(selection, "wholeStory");
22        Dispatch.call(selection, "Copy");
23        OutputStream = new FileOutputStream(txtFile);
24        Writer = new OutputStreamWriter(outStream,
25            Charset.forName("UTF8"));
26        String text;
27        try {
28            text = (String) clipboard.getData(DataFlavor.stringFlavor);
29        } catch (IllegalStateException e) {
30            Thread.sleep(500);
31            text = (String) clipboard.getData(DataFlavor.stringFlavor);
32        }
33        try {
34            clipboard.setContents(new StringSelection(""), null);
35        } catch (IllegalStateException e) {
36            Thread.sleep(500);
37            clipboard.setContents(new StringSelection(""), null);
38        }
39        writeTag(writer, meta);
40        writer.write(text.replace("\n", "\r\n"));
41        writer.close();
42        Dispatch.call(document, "Close", new Variant(wdDoNotSaveChanges));
43    }
44    //...

```

Слика 8.3: Конверзија DOC датотека у чисти текст



2. коришћењем *copy and paste* механизма, где означимо цео текст у *Word* документу, изведемо операцију *copy* и у другој апликацији изведемо операцију *paste* у форми чистог текста;
3. приступом унутрашњој структури *Word* документа путем *Automation* интерфејса ради издвајања текста из сваког од пасуса.

У почетној фази истраживања, док конверзија није била аутоматизована, коришћен је први начин. Извор могућих проблема код коришћења опције снимања у облику чистог текста је чињеница да *Word* покушава да неке елементе форматирања текста прикаже у форми чистог текста. Нешто од тога се може искључити, као што је преламање пасуса на више редова, нешто не производи суштинске проблеме, као што је уметање размака на почетак реда ради увлачења, али има детаља који могу да праве проблеме, а тешко их је избећи избором опција при снимању, као што је укључивање текста из заглавља и подножја стране, попут нумерације страна.

Када се прешло на коришћење *Automation* интерфејса, у почетку смо користили приступ унутрашњој структури документа, читање листе свих пасуса, преузимање чистог текста из једног по једног пасуса у променљиве програмског језика *Java* и даље уписивање у текстуалну датотеку из *Java*-е. Обрада је релативно дуго трајала, мада је проблем са перформансама могао да се поднесе с обзиром да обрада није морала често да се понавља. Други уочен проблем јесте да се код поља која аутоматски добијају вредност (као што је аутоматска нумерација која се појављује у неким *Word* документима) не појављује тренутна вредност поља, која се види кроз кориснички интерфејс. Проблем са вредностима поља је вероватно могуће решити, али смо у међувремену прешли на други начин конверзије.

На крају смо дошли до решења да се кроз *Automation* интерфејс изврши операција селектовања целог текста, а затим операција *copy* и да се након тога на страни *Java* програма непосредно имплементира еквивалент операције *paste*. Испоставило се да је и по перформансама и по чистоћи конвертованог текста ово најбоље од свих разматраних решења. Пример *Java* кода који на овај начин обавља конверзију дат је на слици 8.3. И у овом случају је пример поједно-

стављен у односу на онај који се стварно користи, пре свега по питању детаља управљања изузецима.

У примеру са слике 8.3 приказан је део кода из дефиниције класе *Doc2Txt*. У конструктору класе се иницијализује приступ систему *COM* и покреће *Microsoft Word* апликација (линије 8 и 9), а одмах након тога се иницијализује приступ системској локацији на којој се смешта резултат операције *copy* (енгл. *System Clipboard*).

Конверзија појединих датотека се обавља позивом методе *convert*. Трећи параметар методе је објекат који садржи додатне податке о закону, претходно преузете из одговарајуће *XML* датотеке. У оквиру методе се прво отвара датотека у оквиру *Word*-а (линије 16-18), затим се селекује цео текст у отвореном документу (линије 19-21) и извршава операција *copy*, чиме је завршен део посла на страни *Word*-а. Са друге стране се прво отвара датотека у коју ће се уписати конвертован текст (линије 23-25), након чега се преузима текст који је „послат” операцијом *copy* (линија 28). Операција преузимања текста може да не успе, па је у том случају треба поновити након неког времена (линије 30 и 31). Следеће што радимо је практично чишћење текста који је остао у *System Clipboard* (линије 33-38).

Линију са лексичком етикетом која означава почетак закона уписујемо позивом методе *writeTag* (линија 42), након чега уписујемо и преузет текст у текстуалну датотеку (линија 41) и на крају затварамо текстуалну датотеку и *Word* документ.

Након свега овога је потребно још спојити све текстуалне датотеке у једну и тиме је обрада преузетог лексичког материјала завршена.

## 8.2 Додатне статистичке анализе корпуса

Систем *Unitex* даје могућност приказивања разних статистичких података о тексту, а постављањем одговарајућих упита се могу добити додатне статистике о заступљености појединих речи, врста речи и израза у тексту. Те могућности смо користили за добијање статистичких података који су приказани у одељку 3.2.

Међутим, ми у корпусу прописа имамо већи број закона, што *Unitex* обрађује као јединствен текст. Анотације у облику лексичких етикета које смо увели дају нам могућност да разликујемо шта припада ком закону, али је то метод који смо сами увели и који није посебно подржан кроз разне могућности *Unitex*-а. Конкретно, *Unitex* не уме да прикаже статистичке индикаторе груписане на нивоу закона, а остављене анотације нам могу помоћи да сами направимо алат који би нам омогућио да се у рачунању бројчаних показатеља спустимо на нивоу закона. Уз помоћ таквог алата смо одредили поједине статистичке показатеље из одељка 3.2, као и податке приказане у прилогу Г.

Да бисмо имплементирали алат за додатне статистичке анализе корпуса, прво морамо да прочитамо токенизовану верзију текста која је произведена током претходне обраде у оквиру *Unitex*-а (пододељак 3.1.4). Токенизован текст је приказан помоћу три датотеке у поддиректоријуму који се креира за сваки претходно обрађен текст и који носи назив *<naziv teksta>.snt*. Те три датотеке су:

1. *tokens.txt* — садржи листу различитих токена из текста: по један токен у сваком реду и у првом реду уписано колико их има (ако се неки токен више пута појављује у тексту, он се само једном појављује у *tokens.txt*);
2. *text.cod* — одређује како су токени из *tokens.txt* поређани у тексту: садржи секвенцу целих бројева у бинарном запису, где сваки број упућује на одговарајући ред из *tokens.txt*, а бројеви се нижу оним редом којим се у тексту нижу одговарајући токени;
3. *enter.pos* — позиције у *text.cod* где се налазе токени који су од карактера за крај реда претворени у размак: секвенца целих бројева у бинарном запису.

Ако посматрамо само *tokens.txt* и *text.cod*, из те две датотеке можемо реконструисати текст, али тако да се на местима прелазака у нови ред појављују размаци. Иза тога стоји идеја да при анализи текста није од значаја где је крај реда. Међутим, у датотеци *enter.pos* је ипак сачувана информација о крају реда да би се омогућило пуно реконструисање текста. Да бисмо отклонили непотребну сложеност која је уведена двојаким третманом краја реда, у интерној структури података која представља токенизован текст просто ћемо увести разлику

између краја реда и размака: поред токена из *tokens.txt* увешћемо додатан токен за крај реда и затим њега користити уместо токена за размак на местима где треба (на основу *enter.pos*). При томе нас и даље ништа не спречава да у обрадама на исти начин третирамо размак и нови ред уколико је то потребно.

У интерној структури података се за сваки ред из *tokens.txt* креира по један објекат класе *Token* (слика 8.4), а затим се према редоследу из *text.cod* формира низ објеката типа *Token*. Тај низ чини интерну репрезентацију токенизованог текста.

Пошто у програмском језику *Java* елементи низа заправо упућују на објекте за које кажемо да су садржани у низу, ми смо практично упућивање преко редног броја које се користи у *text.cod* превели на стандардно упућивање на објекат програмског језика *Java*. То значи да на неке објекте типа *Token* упућујемо са више места из низа, па смо се зато определили да објекат класе *Token* не може да се мења након креирања (сва поља су означена са *final*).

```
1 public class Token {
2     public final int code;
3     public final String image;
4     public final TokenType type;
5     public final Object value;
6
7     public Token(int code, String image,
8                 TokenType type, Object value) {
9         this.code = code;
10        this.image = image;
11        this.type = type;
12        this.value = value;
13    }
14    //...
```

Слика 8.4: Класа *Token*

Поља у класи *Token* имају следећа значења:

1. *code* — редни број токена из *tokens.txt*, то јест број којим се из *text.cod* упућује на тај токен;
2. *image* — текст који токен представља, то јест садржај реда из *tokens.txt*;
3. *type* — тип токена који може бити: секвенца слова (реч), цифра, размак, нови

ред, токен  $\{S\}$ , токен  $\{STOP\}$ <sup>7</sup>, лексичка етикета за анотацију закона или друга лексичка етикета;

4. *value* — могућност уписа додатне вредности у зависности од типа токена, као што је назив закона у случају лексичке етикете за анотацију закона.

Након што учитамо токенизован текст у овакву структуру података, даља статистичка рачунања се обављају одговарајућим елементарним алгоритмима.

### 8.3 Статистике препознатих именованих ентитета

У одељку 3.6 приказане су упоредне статистике учесталости појединих типова именованих ентитета за корпус закона и корпус агенцијских вести. За рачунање те врсте статистика израђен је програм у програмском језику *Java*.

Као што је констатовано у одељку 3.6 и илустровано на слици 3.15, након примене Система за препознавање именованих ентитета, сваки именовани ентитет је обележен лексичком етикетом са синтаксним кодом *NE*, као на пример:

```
{10 година,.NE+time+duration+abs+grfDuration2}
```

Након синтаксног кода *NE* следи семантички код који одређује тип именованог ентитета, а затим следе семантички кодови за подтипове (највише три нивоа). На крају могу бити присутни и семантички кодови који почињу са *grf* који представљају траг о томе да је примењен одређен граф. Пуном ознаком типа именованог ентитета сматраћемо надовезане семантичке кодове за тип и подтипове, раздвојене тачком, што би за претходни пример било: *time.duration.abs*.

Систем за препознавање именованих ентитета се заснива на каскадној примени трансдуктора о којој је било речи у одељку 6.2. Том приликом је описано да резултат каскадне примене трансдуктора може бити и у *XML* формату, у ком случају лексичка етикета из претходног примера изгледа:

---

<sup>7</sup>Токен  $\{STOP\}$  не користимо у овом истраживању а уврштен је јер се као могућност наводи у документацији система *Unitex*[29].

```
<csc>
  <form>10 година</form>
  <code>NE</code>
  <code>time</code>
  <code>duration</code>
  <code>abs</code>
  <code>grfDuration2</code>
</csc>
```

Користећи подршку за *XML* из стандардне библиотеке програмског језика *Java*, програм проналази све лексичке етикете са синтаксним кодом *NE*, формира пуне ознаке типова и рачуна учесталости према пуним ознакама и према основним типовима.

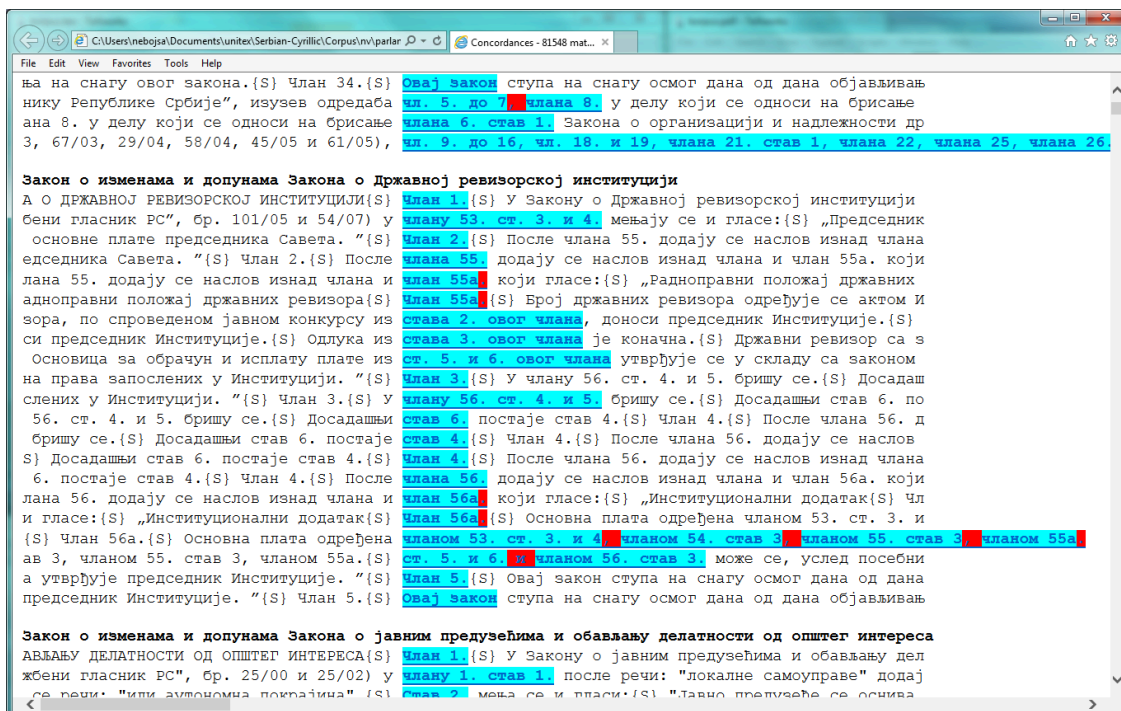
## 8.4 Преклопљене и вишедокументске конкорданце

За потребе овог истраживања смо направили посебан алат за генерисање конкорданци који решава два захтева:

1. преклопљено приказивање конкорданци два графа, где први увек препознаје исте или шире делове текста од другог и
2. конкорданце над корпусом који чини већи број докумената, што је у овом случају корпус прописа.

Примена функционалности преклопљених конкорданци је описана у одељку 4.9. Алат користи модуле система *Unitex* за производњу конкорданци, а затим одређује преклопљене конкорданце и формира резултат у облику *HTML* датотеке која се може отворити у веб прегледачу, као што је и приказано на слици 8.5. У датом примеру су конкорданце произведене у редоследу појављивања у тексту.

Алат који смо развили за потребе овог истраживања такође препознаје библиографске анотације које смо увели у корпусу прописа, тако да у приказу конкорданци може да истакне шта припада ком законском тексту, као што се види на слици 8.5. Наравно, алат може да се користи и за било који други скуп



Слика 8.5: Преклопљене конкорданце

документата који је припремљен кроз поступак претходне обраде на сличан начин као што је то учињено за корпус прописа.

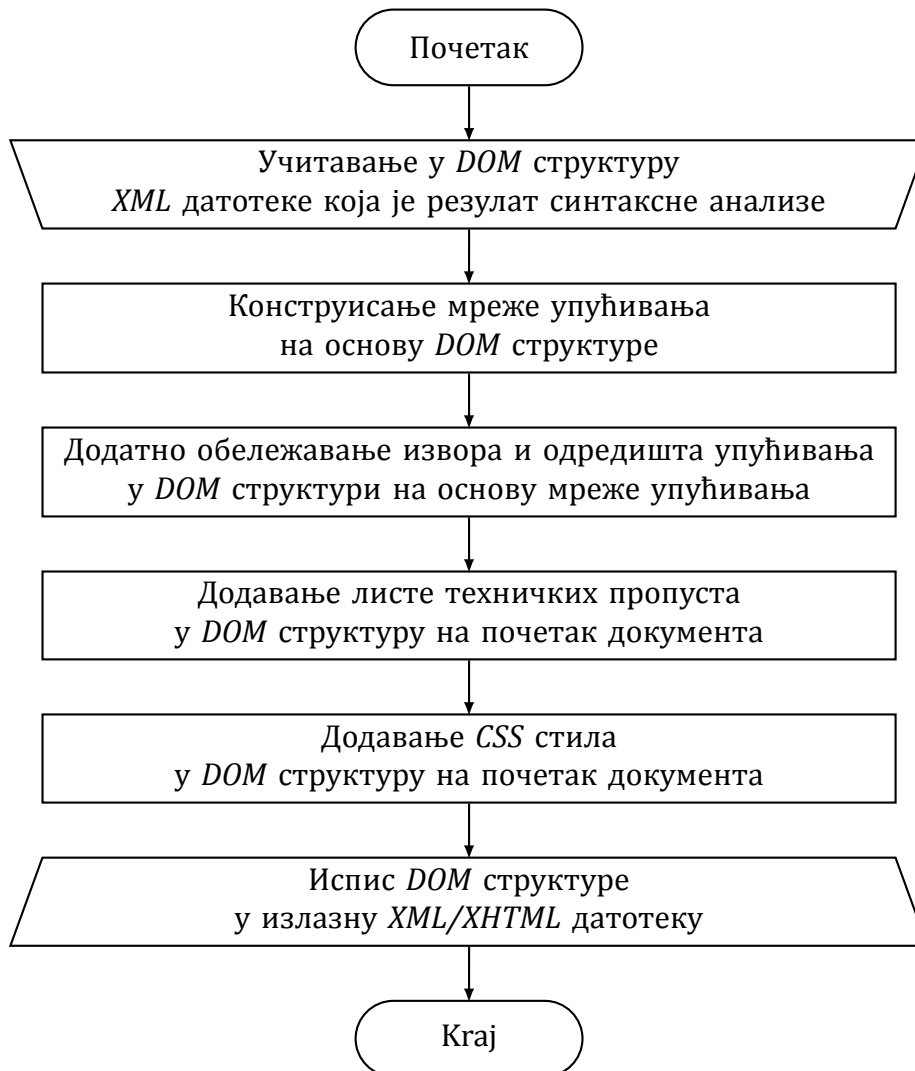
Алат је имплементиран у програмском језику *Java* и користи *Unitex* команде *locate* и *concord* које се и иначе користе за генерисање конкорданци. Алат затим анализира конкорданце које су генерисане *Unitex* командама и претходно обрађен корпус да би на основу свега формирао коначан резултат у облику *HTML* датотеке.

## 8.5 Генерисање контролног документа

У поглављу 7 су описани изглед и функционалност контролног документа, као и основне техничко-технолошке карактеристике контролног документа. Том приликом је истакнуто да се за генерисање контролног документа на основу резултата синтаксне анализе користи програм у програмском језику *Java*, као и да се детаљи изгледа контролног документа специфицирају путем *CSS*-а, што ћемо детаљније описати у овом поглављу.

Главни кораци *Java* програма за генерисање контролног документа прика-

зани су на дијаграму 8.1.



Дијаграм 8.1: Главни кораци *Java* програма

За манипулацију садржајем *XML* документа користи се подршка за *XML* из стандардне библиотеке програмског језика *Java*. Том приликом се *XML* документ у меморији рачунара представља у *Document Object Model (DOM)* структури<sup>8</sup>.

<sup>8</sup><http://www.w3.org/DOM/>

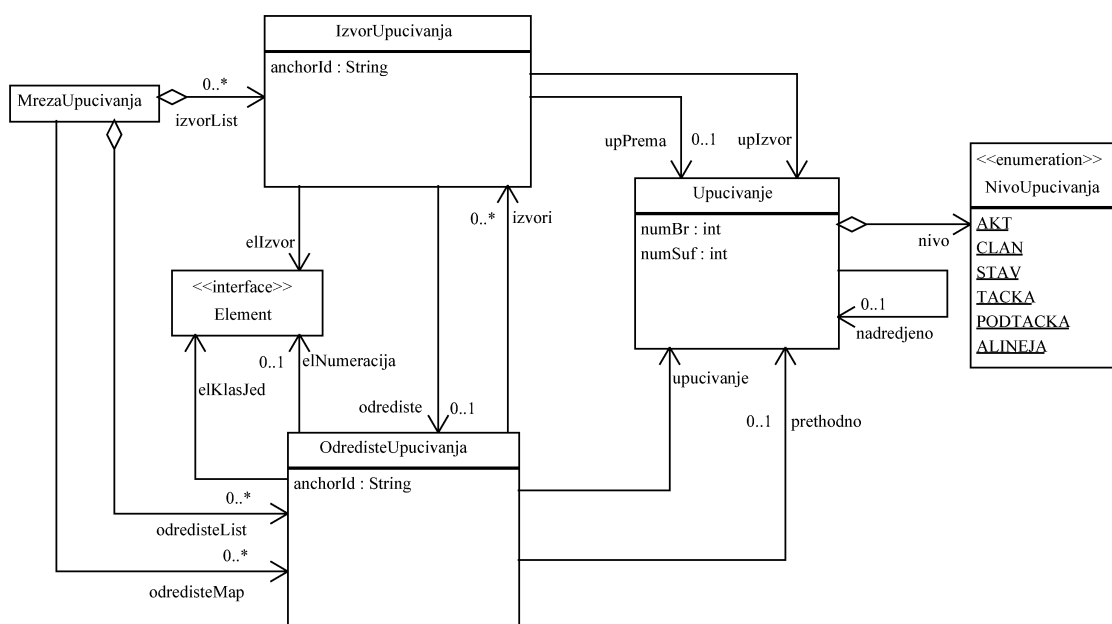


## 8.5.1 Конструисање мреже упућивања

Након читавања из датотеке у *DOM* структуру *XML* документа који је резултат синтаксне анализе, програм конструише мрежу упућивања. У питању је структура података која изражава релације између упућујућих фраза (извора упућивања) и места у документу на која се може вршити упућивање (одредишта упућивања).

Одредиште упућивања може бити свака класификациона јединица на нивоу члана или на нижем нивоу. Свако одредиште у оквиру једног закона је јединствено идентификовано отвореним упућивањем које почиње од нивоа члана и на крају упућује само на то одредиште, на пример: „члан 3. став 5. тачка 2)”. Овакво упућивање у програму користимо као јединствени идентификатор одредишта и зваћемо га идентификационо упућивање.

Класе у *Java* програмском језику помоћу којих је имплементирана мрежа упућивања шематски су приказане на дијаграму 8.2 коришћењем *UML* нотације [74].



Дијаграм 8.2: *UML* дијаграм класа за мрежу упућивања

Идентификационо упућивање се у структури података појављује у два облика: као вредност типа *String* која представља кључ по коме се проналази

одредиште и као једноструко повезана листа, где сваки елемент листе одговара једном нивоу упућивања, а почетак листе је на најнижем нивоу.

Идентификационо упућивање у облику *String*-а (кључ) је сама та упућујућа фраза.<sup>9</sup> Идентификационо упућивање у облику једноструко повезане листе је објекат класе *Upucivanje*. За сваки објекат те класе је одређен ниво класификационе јединице (веза *nivo*), децимални дело нумеричке ознаке (поље *numBr*), суфикс нумеричке ознаке (поље *numSuf*) и веза према остатку листе, тј. према идентификатору надређене класификационе јединице (веза *nadredjeno*). Суфикс се интерно памти као редни број у листи могућих суфикса.

Корени објекат мреже упућивања је класе *MrezaUpucivanja* и њега чини листа објеката класе *IzvorUpucivanja* (веза *izvorList*) и листа објеката класе *Odre-disteUpucivanja* (веза *odredisteList*). Додатна веза према објектима класе *Odre-disteUpucivanja* (веза *odredisteMap*) омогућава проналажење одредишта према кључу (идентификационом упућивању у облику *String*-а) и та веза је имплементирана коришћењем *HashMap*<sup>10</sup> структуре из стандардне библиотеке програмског језика *Java*.

У примерима из одељка 7.2 видимо да су у контролном документу код сваке упућујуће фразе уоквирене све нумеричке ознаке и да се приликом позиционирања курсора миша изнад неке од тих нумеричких ознака приказује облачић са хипервезом према одговарајућем одредишту. Стога сваку нумеричку ознаку у упућујућој фрази третирамо као извор упућивања. Извор упућивања је представљен објектом класе *IzvorUpucivanja* који је повезан са *XML* елементом одговарајуће нумеричке ознаке у оквиру *DOM* структуре (веза *ellzvor*). Објекат класе *IzvorUpucivanja* је даље повезан са два објекта класе *Upucivanje*, и то: са идентификатором одредишта на које се упућује (веза *upPrema*) и идентификатором класификационе јединице у којој се налази извор упућивања (веза *upIzvor*).

---

<sup>9</sup>Разматрана је и могућност да кључ има компактнији запис, на пример „3-5-2” уместо „члан 3. став 5. тачка 2)”, али је закључено да разлика у перформансама коју бисмо на тај начин добили није критичка (барем не за примену у оквиру овог истраживања), а увели бисмо додатну сложеност разликовањем интерне и екстерне форме идентификационог упућивања.

<sup>10</sup><http://docs.oracle.com/javase/7/docs/api/java/util/HashMap.html>

Објекат класе *OdredisteUpucivanja* је повезан са *XML* елементом одговарајуће класификационе јединице (веза *elKlasJed*) и са *XML* елементом који садржи нумеричку ознаку за ту класификациону јединицу, уколико је за тај тип класификационе јединице предвиђено нумеричко означавање (веза *elNumeracija*). Објекат класе *OdredisteUpucivanja* је поред тога повезан са два објекта класе *Upucivanje*, и то: са својим идентификационим упућивањем (веза *upucivanje*) и са идентификатором класификационе јединице истог нивоа која претходи у тексту прописа (веза *pretходно*).

На крају, објекти класе *IzvorUpucivanja* и *OdredisteUpucivanja* су и непосредно повезани (везе *odrediste* и *izvori*). Сваки од тих објеката такође садржи поље *anchorId* које се користи у формирању хипервеза.

У приказаној структури података постоји редундантност која је последица чињенице да је структура прилагођена и поступку конструисања и ономе чему после треба да служи, а у поступку конструисања се неки делови структуре конструишу на основу других и самим тим су редундантни. На пример, веза *odredisteMap* се конструише на основу веза према идентификационим упућивањима која се из објеката класе *Upucivanje* претварају у кључеве, док се у следећем кораку конструишу непосредне везе између објеката класа *IzvorUpucivanja* и *OdredisteUpucivanja* (везе *odrediste* и *izvori*).

## 8.5.2 Ажурирање *DOM* структуре

На основу конструисане мреже упућивања ажурира се *XML* документ који је садржан у *DOM* структури. Прво се ажурирају извори упућивања, да би се постигло појављивање облачића када курсором миша пређемо преко тог извора (слика 8.6) и да би се извор упућивања означио као место на које може упутити хиперлинк (*anchor*).

- 1) са њиме повезано лице, ако је оно бил → члан 65. став 3. тачка 4) овог закона ако је према њему предузета правна радња;
- 2) треће лице из члана 65. став 3. тач. 3) и 4) овог закона, ако је знало или морало знати за постојање личног интереса у време закључења правног посла, односно предузимања правне

Слика 8.6: Пример облачића који се појављује изнад извора упућивања

У примеру са слике 8.6 извор упућивања је нумерација „4”, што у резултату синтаксне анализе представља *XML* елемент типа *Numeracija*. Садржај тог *XML* елемента пре и после ажурирања дат је на слици 8.7.

```
<Numeracija>4</Numeracija>
      ↓
<Numeracija bubble="1">
  <xhtml:a id="a517"/>
  4
  <BubbleEnv>
    <Bubble>
      <xhtml:span>
        <xhtml:a href="#a477">
          члан 65. став 3. тачка 4) овог закона
        </xhtml:a>
      </xhtml:span>
    </Bubble>
  </BubbleEnv>
</Numeracija>
```

Слика 8.7: Извор упућивања пре и после ажурирања *XML*-а

На слици 8.7 су подебљаним словима означени уметнути делови: атрибут *bubble* који означава да нумерација садржи облачић, елемент *xhtml:a* који дефинише место на које може упутити хиперлинк (*anchor*) и елемент *BubbleEnv* у оквиру кога је садржај облачића. Изглед облачића и функционалност појављивања и склањања одређени су *CSS* стилем.

Након ажурирања извора упућивања, на сличан начин се ажурирају и одређена упућивања. Затим се на почетак додаје списак идентификованих пропуста и испред тога се умеће *CSS* стил.

Можемо приметити да смо приликом ажурирања уметали и поједине *XHTML* елементе. То смо чинили за хипервезе и за уграђивање *CSS* стила у сам документ. Као крајњи резултат имамо специјализован *XML* у који су уметнути поједини *XHTML* елементи и где се *CSS* стилем одређује изглед свега заједно. На тај начин смо успели да истовремено постигнемо два циља:

1. да контролни документ буде једна датотека чији садржај се може приказати

у сваком савременом веб прегледачу;

2. да контролни документ буде допуна *XML* документа добијеног синтаксном анализом, уместо да се прави комплетна трансформација у *HTML*.

Наиме, веб прегледачи стандардно подржавају приказ *XML*-а уз помоћ *CSS* стила, али је за хипервезе и уградњу *CSS* стила у сам документ било потребно укључити *HTML*. Примењен начин комбиновања наведених стандарда добро подржавају савремени веб прегледачи.

У последњем кораку генерисања контролног документа *XML* документ се из *DOM* структуре исписује у датотеку. Назив те датотеке добија наставак *.xhtml* зато што желимо да подразумевана апликација која отвара ту датотеку буде веб прегледач.

## 8.6 Додатне могућности *Unitex*-а

С обзиром да је *Unitex* софтвер отвореног кода, свако ко то жели може да преузме изворни код *Unitex*-а и модификује га [75]. На тај начин су у оквиру овог истраживања имплементирани одређење додатне могућности *Unitex*-а које су мотивисане потребама самог истраживања, а по својој природи имају значаја за целокупну популацију корисника *Unitex*-а. Те додатне могућности су, као допринос аутора ове дисертације, постале саставни део верзије 3.1 *Unitex*-а, која је у време писања овог рада у *beta* издању. Додатне могућности се односе на графички кориснички интерфејс *Unitex*-а који користи *Java SE* платформу и *Swing* компоненте<sup>11</sup>.

### 8.6.1 Генерисање битмапираних слика за графове

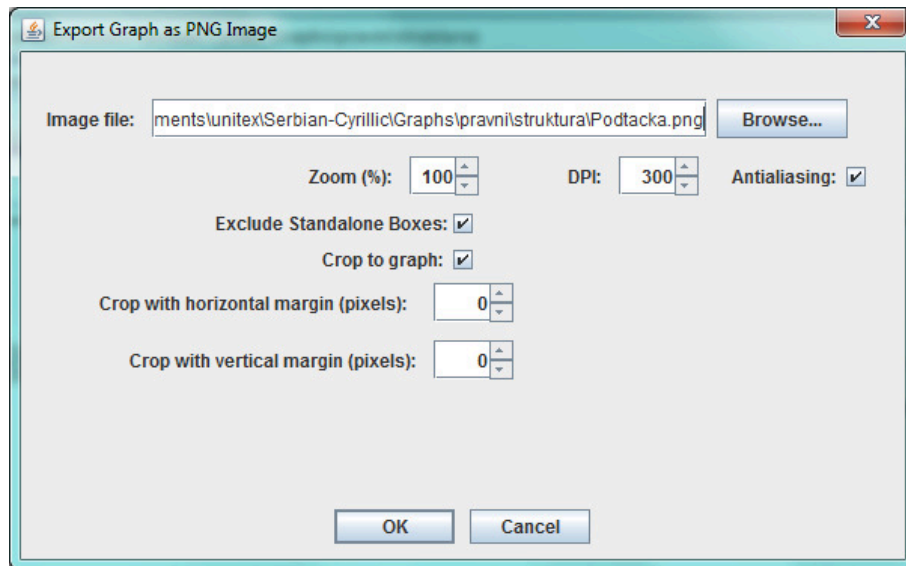
*Unitex* у верзији 3.0 подржава генерисање битмапираних слика за графове у оквиру *Save as* опције. Међутим, том приликом корисник не добија могућност да подешава опције од значаја за генерисање битмапираних слика, као што је резолуција, већ се користе подешавања која су задата за приказ графа на екрану,

---

<sup>11</sup><http://docs.oracle.com/javase/7/docs/technotes/guides/swing/>

а која нису довољно флексибилна за формирање квалитетних битмапираних слика и не омогућавају ефикасан рад у случају када за већи број графова треба формирати слике.

Слике у овом раду на којима су приказани *Unitex* графови, укључујући слике из прилога А, су формиране коришћењем нове опције за извоз графа у облику битмапиране слике (види слику 8.8).



Слика 8.8: Извоз графа у облику слике

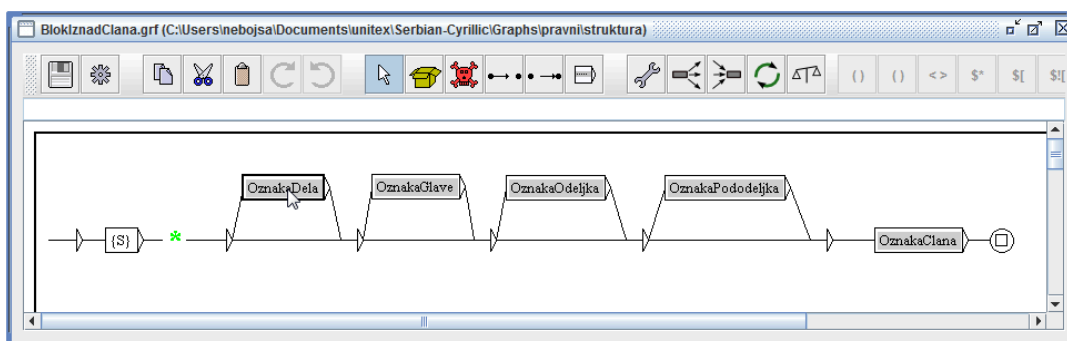
Приликом коришћења ове опције, поред избора датотеке за слику, додатно се може се подесити: фактор увеличавања, густина тачака (*DPI*), омекшавање ивица (*Antialiasing*), искључивање неповезаних кућица (што су најчешће коментари) и ограничење слике само на део који заузима граф (*Crop to graph*). За слике у *JPEG* формату се може подесити и квалитет компресије.

Нова опција омогућава брзу и квалитетну припрему слика са графовима за потребе укључивања у писани рад. Наиме, код припреме за штампу је потребно посебно контролисати увеличавање, а посебно густину тачака, слика која се укључује у писани рад треба да садржи граф без додатног белог простора око њега (или са прецизно дефинисаним маргинама у односу на део који заузима граф), а коментари на графу, који имају сврху када се граф самостално приказује, могу бити сувишни када је граф као слика укључен у писани рад, јер се садржај слике објашњава у пратећем тексту.

## 8.6.2 Унапређења уређивача графова

Од укупног времена које је у оквиру овог истраживања проведено у коришћењу система *Unitex* највише је утрошено на изради и измени графова. Том приликом су уочене могућности за унапређење уређивача графова у оквиру *Unitex*-а, од чега је имплементирано:

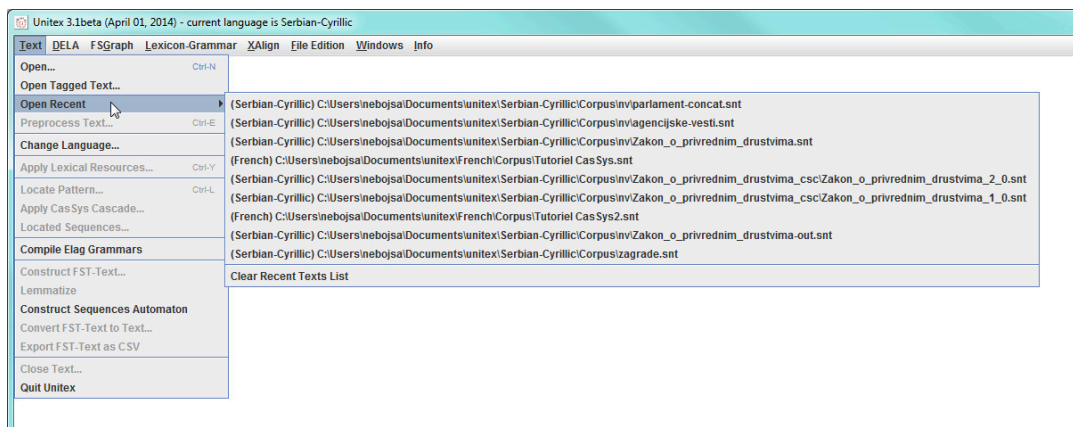
- истицање кућице изнад које је курсор миша, да би корисник лакше „нацртао” кућицу коју жели да изабере или помери (слика 8.9);
- толеранција на незнатно, често нехотично, померање миша приликом клика на кућицу, које је потребно интерпретирати као „чисти клик”, уместо као незнатно померање кућице које корисник доживљава као да рачунар није реаговао на клик;
- када се прекине уређивање текста у оквиру једне кућице (на пример кликом на другу кућицу) измењен текст по правилу треба уписати назад у кућицу, а за опозив измене треба користи *ESC* тастер;
- бројна мања унапређена и исправке у оквиру уређивача графова.



Слика 8.9: Истицање кућице изнад које је курсор миша (стрелица)

## 8.6.3 Остала унапређења

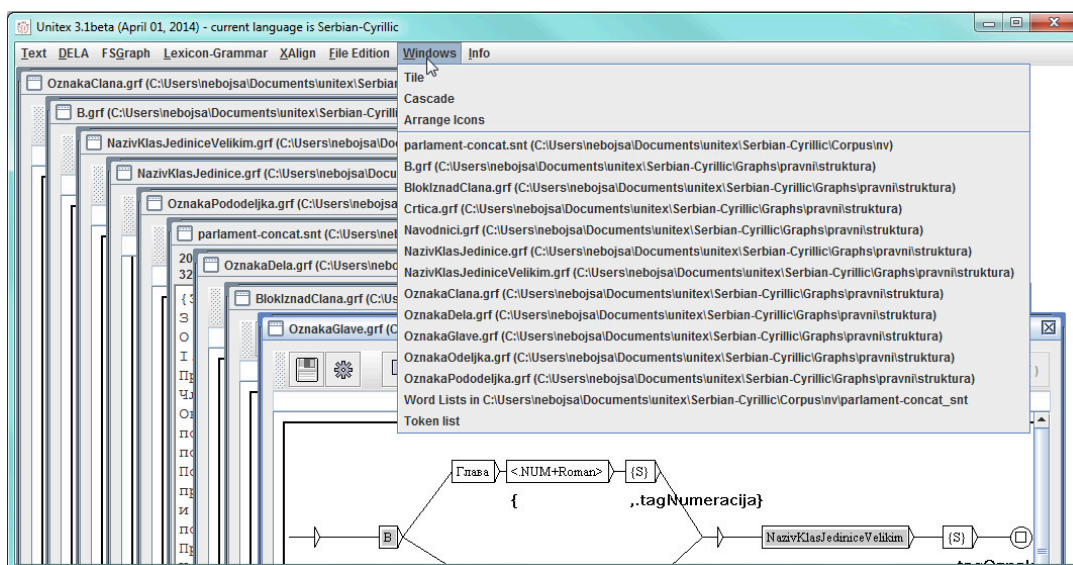
Од осталих унапређења корисничког интерфејса система *Unitex* посебно ћемо истаћи онемогућавање опција менија које у датој ситуацији нису применљиве (уз промену боје слова у сиву) и увођење опције за брзо отварање скоро коришћених датотека (слика 8.10).



Слика 8.10: Онемогућене опције менија и скоро коришћене датотеке

Брзо отварање скоро коришћених датотека је имплементирано за отварање текста и за отварање графова. При томе се при отварању скоро коришћеног текста истовремено поставља и избор језика на онај језик који је био изабран приликом претходног коришћења тог текста.

Уз увођење могућности да се приликом поновног покретања корисничког интерфејса као подразумевана опција у избору језика понуди језик који је претходно коришћен, брзо отварање скоро коришћених датотека знатно поједностављује операције које корисник понавља приликом сваког покретања корисничког интерфејса.



Слика 8.11: Листа отворених прозора у *Windows* менију



Од имплементираних унапређења ћемо још навести и додавање листе тренутно отворених прозора у оквиру *Windows* менија, што олакшава прелазак на жељени прозор у случају када је отворен већи број прозора (слика 8.11).

## 8.7 Транскодовање графова

Систем за препознавање именованих ентитета који је коришћен у одељку 3.6 и описан у [68] у оригиналној верзији користи специфичан начин записивања слова српског језика. Реч је о кодној шеми која се користи у многим језичким ресурсима развијеним у оквиру Групе за језичке технологије Универзитета у Београду, укључујући и Корпус савременог српског језика [50], [51], [53] о коме је било речи у поглављу 3. Кодна шема носи назив Аурора по истоименом програмском систему Душка Витаса описаном у раду из 1981. године [76].

Запис у кодној шеми Аурора користи само слова енглеског алфабета тако што се „ћ”, „ч”, „ш”, „ж”, „ђ”, „џ”, „љ” и „њ” записују као „cx”, „cy”, „sx”, „zx”, „dx”, „dy”, „lx” и „nx”, док се остала ћирилична слова записују као латинични еквиваленти. На пример, реч „кључ” се записује као „klxycy”. Латинични текст се у кодној шеми Аурора записује идентично као и њему еквивалентан ћирилични текст.

Кодна шема Аурора је осмишљена тако да алати који се користе за обраду текста не морају да имају посебну подршку за примењену кодну шему. На пример, ако алат треба да препозна секвенце узастопних слова и то ради добро за текст на енглеском језику, исправно ће препознати секвенце узастопних слова и у ћириличном тексту који је записан у Аурора кодној шеми. Чак је и енглески лексикографски поредак Аурора-записаних речи у високој мери сагласан са латиничним абецедним поретком тих речи у српском језику. Такође се може рећи да није тешко навићи се на читање текста записаног у Аурора кодној шеми, што се, наравно, не може очекивати од ширег круга крајњих корисника, али за потребе истраживачког рада може бити прихватљиво.

Оваква шема је била неопходна у време када је Аурора настала, јер је тада подршка разноврсним алфабетима, осим енглеској латиници, у рачунарским системима била на ниском нивоу. Међутим, Аурора остаје погодна за истражи-

вања у обради природног језика докле год у ланцу алата које користимо може да се појави неки који нема довољно добру подршку за стандардне кодне шеме намењене алфаветима других језика, а не само енглеском, што се у данашње време практично своди на питање подршке за *Unicode*.

На жалост, подршка за *Unicode* још увек увек није присутна у свим значајним алатима за обраду природног језика. На пример, за индексирање Корпуса савременог српског језика се користи *The IMS Open Corpus Workbench (CWB)*<sup>12</sup> који у верзији 3.0 (што је последња стабилна верзија у време писања овог рада) још увек нема подршку за *Unicode*. Подршка за *Unicode* је најављена за следећу стабилну верзију.

С обзиром да основни алати који су коришћени у овом истраживању (*Unitex*, *Java* и алати за *XML*) имају подршку за *Unicode*, одредили смо се да користимо стандардан ћирилични *Unicode* запис за текст на српском језику.

У Систему за препознавање именованих ентитета у оквиру *Unitex*-ових синтаксних графова се користи кодна шема Аурора, што значи да су ти синтаксички графови прилагођени корпусу записаном у кодној шеми Аурора. Како је корпус закона у овом истраживању записан у складу са *Unicode* стандардом за кодирање карактера, потребно је у графовима изменити одређене делове из Аурора записа у стандардне ћириличне *Unicode* карактере, а што је могуће урадити захваљујући томе што *Unitex* подржава *Unicode*. Пример измене графа је дат на слици 8.12.

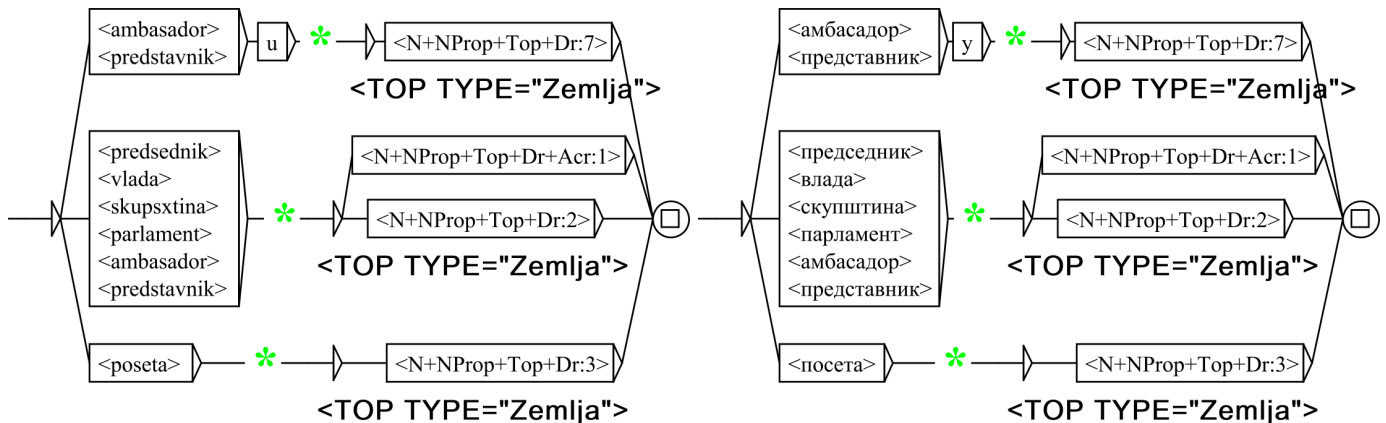
На слици 8.12 је са леве стране приказан граф у Аурора кодној шеми а са десне стране у стандардној ћирилици. Примећујемо да коментар на врху графа није промењен. Коментар заправо и није у Аурора кодној шеми, већ у стандардној латиници, тако да је транслитерација коментара из латинице у ћирилицу посебно питање. С обзиром да функционалност графова не зависи од коментара, питањем транслитерације коментара се нисмо бавили.

Синтаксички и семантички кодови (попут „N+NP<sub>prop</sub>+Top+Dr”) треба да остану непромењени, док леме (попут „<skupsxtina>”) и експлицитан текст (попут „u”) треба транскодovati у ћирилицу (у „<скупштина>” и „у”). У датом примеру трансдукторске излазе (попут „<TOP TYPE="Zemlja">”) није требало транско-

<sup>12</sup><http://cwb.sourceforge.net/>

prepoznaje drzavu na osnovu pojavljivanje ispred nje nekog okidaca  
predsednik Francuske

prepoznaje drzavu na osnovu pojavljivanje ispred nje nekog okidaca  
predsednik Francuske



Слика 8.12: Граф *drzava* у Аурора (лево) и ћириличној (десно) варијанти

довати, јер су то посебна обележја која не представљају текст на српском језику. Неки делови обележја потичу од речи из енглеског језика (попут „TYPE”), неки су само почеци речи (попут „TOP”), док неки делови обележја потичу од речи из српског језика (попут „Zemlja”), али без примене Аурора кодне шеме (у Аурори би било „Zemlxa”).

С обзиром да Систем за препознавање именованих ентитета садржи више од 1200 графова, било је потребно направити алат за аутоматско транскодовање графова. Алат је имплементиран у програмском језику *Java*, а поред тога је направљен кратак *batch script*<sup>13</sup> за једноставно коришћење алата из командне линије на *Windows* оперативном систему. Транскодовање графова се покреће командом:

```
tcgraphs izvorni-dir ciljni-dir
```

Након извршавања команде сви графови из директоријума *izvorni-dir*, укључујући поддиректоријуме, биће транскодовани и уписани у *ciljni-dir*, задржавајући структуру поддиректоријума.

Главни задаци које је требало решити при имплементацији алата су:

1. генерална манипулација графовима (читање из датотеке, модификовање и писање у датотеку);
2. одређивање делова које треба транскодовати;

<sup>13</sup><http://technet.microsoft.com/en-us/library/bb490869.aspx>

3. транскодовање ниске карактера;
4. обрада већег броја датотека у оквиру задатог директоријума.

За манипулацију графовима коришћен је постојећи програмски кôд из дела система *Unitex* који је имплементиран на програмском језику *Java*. За одређивање делова које треба транскодовати имплементирана је додатна синтаксна анализа садржаја кућица и логика која разграничава случајеве у складу са претходном анализом примера са слике 8.12. При томе су сви излази трансдуктора искључени из транскодовања. За транскодовање ниске карактера је имплементиран једноставан коначни аутомат. За обраду већег броја датотека у оквиру задатог директоријума је коришћена подршка стандардне библиотеке програмског језика *Java* за рад са датотекама и директоријумима.

## Поглавље 9

### Закључак и даљи рад

Централни резултат овог истраживања је софтверски алат за формирање контролног документа на основу чистог текста прописа (поглавље 7). Алат представља прототип софтверског решења применљивог у пракси и сублимира друге резултате истраживања: прецизан опис методама рачунарске лингвистике језичких правила за упућујуће фразе (поглавље 4), опис логичке структуре прописа (поглавље 5) и имплементацију синтаксне анализе логичке структуре и упућујућих фраза у тексту прописа (поглавље 6).

Основна сврха контролног документа је да помогне у уочавању техничких пропуста и методолошких несагласности при изради прописа, а може се користити и као подршка анотацији текста приликом припреме за унос у базе података које садрже текстове прописа. Прегледним приказом препознате структуре текста прописа и упућујућих фраза посредно се помаже уочавању пропуста, јер део текста који је на неочекиван начин приказан указује на могућ пропуст који је до тога довео. Поред тога, у контролном документу се одређени пропусти експлицитно истичу, као што су неисправна упућивања, пропусти у редоследу нумерисања и пасуси који нису могли бити класификовани у складу са методолошким правилима.

У тексту Закона о привредним друштвима, који је као пример коришћен у овом раду, на шест места су идентификована неисправна упућивања, што је највероватније пропуст током израде закона у усклађивању упућујућих фраза са изменама нумерација одредби на које се упућује. Поред аутоматског препо-

знавања одређених пропуста у упућујућим фразама, алат помаже и при „ручној” провери упућивања тиме што омогућава хипервезе из упућујуће фразе према одредби на коју се врши упућивање, као и хипервезе у супротном смеру.

То што је алат помогао да се открију пропусти чак и у донетим законима је неспоран показатељ његове употребљивости. Међутим, треба имати у виду да текст закона пролази више фаза израде и нивоа редакције и да би у тим корацима који претходе доношењу закона алат могао имати још већу употребну вредност. Дакле, у процесу израде и доношења прописа, текст прописа пролази кроз већи број редакција, како оних које представљају формалне кораке у процесу (након седнице Владе или седнице Народне скупштине), тако и интерних редакција током израде прописа. Подршка редакцији текста прописа је главни случај примене израђеног алата, односно производа који би се даље развијао на бази датог прототипа.

Прецизан опис језичких правила за упућујуће фразе, поред тога што чини основ за израду аутоматизованих алата, уједно представља једно прецизно тумачење методолошких правила за израду прописа, до кога се дошло кроз анализу бројних специјалних случајеве, уз коришћење корпуса законских текстова који је израђен у оквиру и за потребе овог истраживања (поглавље 3), као и уз помоћ посебно израђених софтверских алата (поглавље 8).

Посебан резултат истраживања представљају опште лингвистичке анализе корпуса закона, уз поређење са Корпусом савременог српског језика. Ове анализе показују да текстови закона имају уобичајене опште карактеристике природно-језичког садржаја уз неке специфичне карактеристике законских текстова, као што је честа употреба појединих иначе не толико учесталих речи (нпр. „односно”), карактеристичан термилошки избор за неке уобичајене појмове (нпр. чешће се користи термин „лице”, а ређе „особа”) или ретко појављивање личних имена наспрам чешћег појављивања институционалних референци.

Можемо рећи да развој аутоматске обраде правних текстова као посебне области обраде природног језика почиње током осамдесетих година, прво у земљама са прецедентним правним системима попут америчког, канадског и британског, а затим и у осталим земљама. Што се тиче српског језика, након

рада Душка Витаса из 1982. године „Ка програму за аутоматско мењање и допуњавање законских текстова” [70], област природно-језичке обраде правних текстова на српском језику није посебно привлачила пажњу истраживача. Поред конкретних приказаних резултата, ова дисертација је покушала да прикаже и потенцијале развоја обраде правних текстова на српском језику.

## 9.1 Даљи рад

Ово истраживање је оставило бројне отворене могућности за даљи рад, како у даљој примени и развоју главних резултата истраживања, тако и у даљем развоју алата који су у оквиру истраживања реализовани.

Анализа упућујућих фраза се може даље развијати у делу упућивања на друге законе, као и обухватањем специфичних форми упућивања које се појављују приликом формулисања одредби о изменама и допунама прописа (нпр. „У члану 7. после става 1. додаје се нови став 2. који гласи:...”), као и у другим специфичним контекстима. Посебну тему могућег истраживања представљају описна упућивања попут „закон којим се уређује...”, „одредбе које се односе на...” или „пропис из члана...”.

Анализа упућивања на друге прописе умногоме улази у проблематику сличну препознавању именованих ентитета, јер је у суштини потребно препознати назив прописа. Препознавање категорија именованих ентитета специфичних за прописе је још једна област могућег даљег рада. Поред назива закона, ту спадају и називи институција који могу бити експлицитни (нпр. „Народна скупштина Републике Србије”) и описни (нпр. „орган надлежан за...”), улоге у систему који је предмет прописа, које опет могу бити експлицитне (нпр. „председник Републике Србије”) и описне (нпр. „министар надлежан за...” или „лице овлашћено за...”), временске одреднице које су такође предмет методолошких правила, употреба бројева и слично.

Неки од наведених случајева именованих ентитета, као што су експлицитни називи институција и улога, датуми и бројеви, су већ обухваћени Системом за препознавање именованих ентитета коришћеном у 3.6, при чему се поставља циљ подизања одзива и прецизности препознавања таквих именованих енти-

тета у оквиру правних текстова (на пример, да се прецизно препознају сви постојећи државни органи). Са друге стране, неки случајеви, као што је описно навођење институција и улога, представљају специфичне форме именованих ентитета карактеристичне за правне текстове који захтевају специфичну технику препознавања.

Повезано са именованим ентитетима је и препознавање прописом утврђених термина. У питању су термини за које се може рећи да у смислу одређеног прописа имају посебно значење, односно на посебан начин прецизирано значење. При томе треба разликовати увођење термина (што се може учинити дефиницијом или просто првим појављивањем у нормативној одредби) и његово касније навођење, као и навођење термина утврђених другим прописима (нпр. „повезано лице у смислу Закона о порезу на добит правних лица”).

Ако се у истраживању такође бавимо својствима и односима појмова који стоје иза препознатих термина, онда се подижемо на семантички ниво. Ту можемо истраживати формалну репрезентацију знања које је исказано одредбама прописа коришћењем одговарајућих формалних језика попут *Web Ontology Language (OWL)* [33], затим могућности екстракције знања, као и могућности аутоматског резонувања, чиме улазимо у област вештачке интелигенције. Прелазак на семантички ниво и укључивање метода вештачке интелигенције отвара простор ка семантичким проверама приликом израде закона, проналажењу одредби које су својим значењем повезане са одређеним питањем, издвајању кључних чињеница из правног текста и на крају према захтевном циљу аутоматског извођења правне аргументације.

У ширем смислу се и аутоматско превођења правних текстова може сматрати могућом даљом активношћу. Аутоматско превођење је специфично по томе што више језика између којих се планира превођење треба обухватити заједничком методологијом и скупом изабраних алата, као и хармонизованим радом експерата за те језике. Због тога се све стране морају прилагодити претходно договореној заједничкој методологији, па се пут до аутоматског превођења не може приказати као потенцијално директан наставак резултата овог истраживања, али се може сматрати могућом даљом активношћу у истој области.



Анализа логичке структуре правног теста се у даљем раду може проширити на препознавање цитата, што је пре свега значајно код измена и допуна. Поред тога се анализа логичке структуре може проширити на разне специјалне делове текста као што су табеле и прилози, а такође и на нове типове аката (одлуке, решења, итд). При томе свака нова специфичност у логичкој структури може значити и одговарајућу специфичност у упућујућим фразама, јер је структура упућујуће фразе повезана са логичком структуром текста.

У даљем раду се по питању измена и допуна може поставити циљ аутоматског формирања пречишћеног текста на основу одредаба о изменама и допунама, уз одговарајућу претходну семантичку анализу одредби, чиме би се реализовала визија Душка Витаса из 1982. године [70].

У делу претходне обраде текста, једна од тема за даљи рад је нежељено појављивање у улазном тексту разних карактера из широког скупа карактера које обухвата *Unicode*. Најчешћи проблем је да се уместо ћириличног слова појављује по изгледу истоветно латинично слово, али представљено другачијим карактером, затим проблем може бити појављивање разних варијанти наводника, цртица и других знакова интерпункције, разних симбола који се могу појавити испред ставки набрајања, карактера који представљају бројеве у експоненту и слично.

Још једна тема у вези са претходном обрадом је сегментација реченица. Поред тога што је у правном тексту посебно изражен значај сегментације на пасусе, имамо и врло специфичну структуру реченице која се може протезати кроз више тачака, подтачака и алинеја, а тиме и кроз више пасуса. Због тога се може размотрити да се уобичајена подела на реченице, као главна сегментација текста у припреми за природно-језичку обраду, раслоји на поделу на пасусе и на реченице.

Функционалност алата за генерисање контролних докумената, поред могућности укључивања резултата других овде описаних даљих активности, може се проширити у смеру да и одређени неисправно исказани елементи логичке структуре и упућујућих фраза буду као такви препознати, уместо да остану непознати. На пример, уколико се алинеја појави непосредно испод става (без тачака и подтачака), што је неисправно, то би ипак могло бити пре-

познато као алинеја, али тако да неисправност буде истакнута. У том случају би требало препознати и потенцијално упућивање на такву алинеју.

Алат за генерисање контролних докумената може да прерасте у алат за проверу текста прописа у току израде и као такав може бити проширен са правописном провером, као и механизмом за сугерисање исправки. Да би се алат користио у свакодневном раду, потребно га је учинити лако доступним крајњим корисницима (било као апликацију која се лако инсталира на локалном рачунару, сервис коме се приступа преко веба или комбинацију та два), затим треба додатно повести рачуна да се алат једноставно и ефикасно уклапа у постојећи процес рада и уобичајена софтверска решења, израдити разумљиву корисничку документацију и алат редовно унапређивати.

Посебан задатак у даљем раду може бити израда и одржавање језичких ресурса за правне текстове. Ту пре свега мислимо на укључивање свих речи које се појављују у прописима у електронске речнике, уз редовно укључивање нових речи које се појављују у новообјављеним прописима. Поред електронских речника, језичке ресурсе представљају и правила за опис логичке структуре правног текста, за препознавање упућивања, именованих ентитета, термина, итд.

Искуство након транскодовања графова из Аурора кодне шеме у стандардну *Unicode* ћирилицу је показало да је језичке ресурсе за српски језик у изворном облику најпогодније развијати користећи стандардну *Unicode* ћирилицу, а затим транслитеровати и транскодовати у латиницу и у Аурора кодну шему. Стога је једна од могућих даљих активности да се у развоју појединих језичких ресурса који су коришћени у овом раду направи прелазак изворне верзије са Аурора кодне шеме на стандардну *Unicode* ћирилицу коришћењем алата за транскодовање, као и да се тај алат прошири тако да може да транскодује у супротном смеру, да би се користио за генерисање латиничне и Аурора варијанте ресурса на основу изворне ћириличне верзије.

Могућа даља унапређења система *Unitex* као основног алата за природнојезичку обраду су у сваком случају бројна, а овом приликом ћемо истаћи потребе које су препознате у оквиру досадашњег истраживања и додатно се намећу из претходно описаног могућег даљег развоја. Кључне потребе за унапре-

ђивање се односе на случајеве када се *Unitex* користи као платформа за специјализована софтверска решења, а те потребе су:

- ефикасно извршавање целокупног циклуса обраде од улазне текстуалне датотеке до крајњег резултата обраде за случајеве примене где се цео циклус обраде заједно са предобрадом често понавља, за разлику од сценарија где се може једном урадити предобрада, а затим више пута понављати даља обрада, чему је *Unitex* тренутно боље прилагођен;
- добро подржана могућност уграђивања платформе у специјализоване апликације и алате, као и израде независних специјализованих модула који раде на платформи, при чему се под добром подршком подразумева комуникација између модула путем добро дефинисаних програмерских интерфејса (*APIs*);
- конзистентан модел аотираног текста као међурезултата између свих фаза природно-језичке обраде уз модуларну архитектуру која омогућава да се промени имплементација модела аотираног текста (меморијска, у датотечком систему, у бази података, прилагођена дистрибуираној платформи и слично) и уз пажљиво пројектован и добро документован програмерски интерфејс.

# Литература

- [1] D. Vitas, L. Popović, C. Krstev, I. Obradović, G. Pavlović–Lažetić, and M. Stanojević, *The Serbian Language in the Digital Age*. Springer Berlin Heidelberg, 2012, ISBN: 978-3-642-30754-6. DOI: 10.1007/978-3-642-30755-3.
- [2] G. Salton, *Automatic Text Processing: the Transformation, Analysis, and Retrieval of Information by Computer*. Reading, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1989, ISBN: 978-0-201-12227-5.
- [3] M. Grobelnik, “Many faces of text processing”, in *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*, ser. WIMS '11, Sogndal, Norway: ACM, 2011, 5:1–5:3, ISBN: 978-1-4503-0148-0. DOI: 10.1145/1988688.1988694.
- [4] D. Vitas, G. Pavlović–Lažetić, and C. Krstev, “Electronic dictionary and text processing in Serbo-Croatian”, in *Sprache – Kommunikation – Informatik: Akten des 26. Linguistischen Kolloquiums*, J. Darski and Z. Vetulani, Eds., vol. 293, De Gruyter, 1993, p. 225.
- [5] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Second Edition*. Prentice-Hall, 2009.
- [6] M. Hess, “Natural and formal language processing”, *Logic Programming in Action*, pp. 127–175, 1992.
- [7] B.-L. Gunnarsson, “Language for special purposes”, in *Encyclopedia of Language and Education*, ser. Encyclopedia of Language and Education, G. Tucker and D. Corson, Eds., vol. 4, Springer Netherlands, 1997, pp. 105–117, ISBN: 978-0-7923-4931-0. DOI: 10.1007/978-94-011-4419-3\_11.

- [8] R. Grishman and R. Kittredge, *Analyzing Language in Restricted Domains: Sublanguage Description and Processing*. Routledge, 1986, ISBN: 978-0-89859-620-5.
- [9] P. M. Tiersma, *Legal Language*. University of Chicago Press, 1999.
- [10] D. Mellinkoff, *The Language of the Law*. Resource Publications, 2004, ISBN: 978-1-59244-690-2.
- [11] T. Endicott, "Law and language", in *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, Ed., 2014.
- [12] M. Nakamura, S. Nobuoka, and A. Shimazu, "Towards translation of legal sentences into logical forms", in *New Frontiers in Artificial Intelligence*, ser. Lecture Notes in Computer Science, K. Satoh, A. Inokuchi, K. Nagao, and T. Kawamura, Eds., vol. 4914, Springer Berlin Heidelberg, 2008, pp. 349–362, ISBN: 978-3-540-78196-7. DOI: 10.1007/978-3-540-78197-4\_33.
- [13] S. Höfler and K. Sugisaki, "From drafting guideline to error detection: automating style checking for legislative texts", in *EACL 2012 Workshop on Computational Linguistics and Writing*, Association for Computational Linguistics, Apr. 2012, pp. 9–18. DOI: 10.5167/uzh-62049.
- [14] М. Павловић, „Кратка историја српског правног језика”, *Зборник радова са научног скупа „Право и језик”*, Крагујевац, 2006.
- [15] С. Табороши, „Језик у праву”, *Зборник радова са научног скупа „Право и језик”*, Крагујевац, 2006.
- [16] Г. Станковић, „Глосе о језику у праву”, *Зборник радова са научног скупа „Право и језик”*, Крагујевац, 2006.
- [17] С. Савић, „Језик закона – карактеристике и родна перспектива”, *Зборник радова са научног скупа „Право и језик”*, Крагујевац, 2006.
- [18] М. Луковић, „Српски правни језик”, *Архив за правне и друштвене науке*, св. 92, бр. 3–4, стр. 1533–1554, 2006.
- [19] N. Visković, *Jezik prava*. Zagreb: Naprijed, 1989.
- [20] З. И. Јелић, *Основи нормативне технике: увод у савремену науку о методама стварања и разумевања права*. Београд: Економика, 1986.

- [21] J. Bates and L. Mader, *Law Drafting and Legislative Process in the Republic of Serbia: An Assessment*. OSCE Office of Democratic Institutions and Human Rights, 2001. [Online]. Available: <http://www.legislationline.org/>.
- [22] Народна Скупштина Републике Србије, „Јединствена методолошка правила за израду прописа”, *Службени гласник Републике Србије*, св. 21/10, 2010.
- [23] Влада Републике Србије, „Методологија за израду подзаконских прописа”, *Службени гласник Републике Србије*, св. 75/10, 2010.
- [24] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.
- [25] M. Gross and D. Perrin, “Electronic dictionaries and automata in computational linguistics”, *Lecture Notes in Computer Science*, vol. 377, 1989.
- [26] M. Gross, “Local grammars and their representation by finite automata”, *Data, Description, Discourse: Papers on the English Language in honour of John McH Sinclair*, pp. 26–38, 1993.
- [27] E. Roche and Y. Shabes, Eds., *Finite-State Language Processing*. Cambridge, MA, USA: MIT Press, 1997, ISBN: 978-0-262-18182-2.
- [28] Unitex website, [Online]. Available: <http://www-igm.univ-mlv.fr/~unitex/>.
- [29] S. Paumier, *Unitex user manual*, Université Paris-Est Marne-la-Vallée. [Online]. Available: <http://www-igm.univ-mlv.fr/~unitex/UnitexManual3.1.pdf>.
- [30] B. Courtois and M. Silberztein, “Dictionnaires électroniques du français”, *Langue française*, vol. 87, no. 1, pp. 3–4, 1990.
- [31] C. Krstev, *Processing of Serbian: Automata, Texts and Electronic Dictionaries*. Faculty of Philology of the University of Belgrade, 2008.
- [32] C. Krstev and D. Vitas, “Processing of corpora of Serbian using electronic dictionaries”, *Prace Filologiczne*, no. LXIII, pp. 279–292, 2012.

- [33] B. Motik, B. Parsia, and P. Patel-Schneider, "OWL 2 Web Ontology Language structural specification and functional-style syntax (second edition)", W3C, W3C Recommendation, Dec. 2012. [Online]. Available: <http://www.w3.org/TR/2012/REC-owl2-syntax-20121211/>.
- [34] E. L. Rissland, K. D. Ashley, and R. P. Loui, "AI and law: a fruitful synergy", *Artif. Intell.*, vol. 150, no. 1-2, pp. 1-15, Nov. 2003, ISSN: 0004-3702. DOI: 10.1016/S0004-3702(03)00122-X.
- [35] B. G. Buchanan and T. E. Headrick, "Some speculation about artificial intelligence and legal reasoning", *Stanford Law Review*, pp. 40-62, 1970.
- [36] D. A. Waterman, J. Paul, and M. Peterson, "Expert systems for legal decision making", *Expert Systems*, vol. 3, no. 4, pp. 212-226, 1986.
- [37] D. Stein, "Machine translation – past, present, and future", *Translation: Computation, Corpora, Cognition*, vol. 3, no. 1, 2013.
- [38] J. R. Pierce, J. B. Carroll, et al., *Language and Machines: Computers in Translation and Linguistics*. Washington, DC, USA: Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council, 1966.
- [39] P. Koehn, A. Birch, and R. Steinberger, "462 machine translation systems for Europe", *Proceedings of MT Summit XII*, pp. 65-72, 2009.
- [40] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation", in *Proceedings of the 40<sup>th</sup> Annual Meeting on Association for Computational Linguistics*, Association for Computational Linguistics, 2002, pp. 311-318.
- [41] E. Chieze, A. Farzindar, and G. Lapalme, "An automatic system for summarization and information extraction of legal information", in *Semantic Processing of Legal Texts*, ser. Lecture Notes in Computer Science, E. Francesconi, S. Montemagni, W. Peters, and D. Tiscornia, Eds., vol. 6036, Springer Berlin Heidelberg, 2010, pp. 216-234, ISBN: 978-3-642-12836-3. DOI: 10.1007/978-3-642-12837-0\_12.

- [42] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst, "Moses: open source toolkit for statistical machine translation", in *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, ser. ACL '07, Prague, Czech Republic: Association for Computational Linguistics, 2007, pp. 177–180. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1557769.1557821>.
- [43] T. Erjavec, C. Ignat, B. Pouliquen, and R. Steinberger, "Massive multi lingual corpus compilation: Acquis Communautaire and totale", *Archives of Control Science*, vol. 15, no. 4, p. 529, 2005.
- [44] D. Gelbart and J. C. Smith, "Beyond Boolean search: FLEXICON, a legal tex-based intelligent system", in *Proceedings of the 3<sup>rd</sup> International Conference on Artificial Intelligence and Law*, ser. ICAIL '91, Oxford, England: ACM, 1991, pp. 225–234, ISBN: 978-0-89791-399-7. DOI: 10.1145/112646.112674.
- [45] M.-F. Moens, "Summarizing court decisions", *Inf. Process. Manage.*, vol. 43, no. 6, pp. 1748–1764, Nov. 2007, ISSN: 0306-4573. DOI: 10.1016/j.ipm.2007.01.005.
- [46] M.-F. Moens and R. De Busser, "First steps in building a model for the retrieval of court decisions", *Int. J. Hum.-Comput. Stud.*, vol. 57, no. 5, pp. 429–446, Nov. 2002, ISSN: 1071-5819. DOI: 10.1006/ijhc.1029.
- [47] M.-F. Moens, E. Boiy, R. M. Palau, and C. Reed, "Automatic detection of arguments in legal texts", in *Proceedings of the 11<sup>th</sup> International Conference on Artificial Intelligence and Law*, ACM, 2007, pp. 225–230.
- [48] E. de Maat, R. Winkels, and T. van Engers, "Automated detection of reference structures in law", *Frontiers in Artificial Intelligence and Applications*, p. 41, 2006.
- [49] L. Bacci, E. Francesconi, and M. T. Sagri, "A rule-based parsing approach for detecting case law references in Italian court decisions", in *Semantic Processing of Legal Texts (SPLeT-2012) Workshop Programme*, 2002, p. 27.



- [50] Д. Витас и Љ. Протић, „Конспект за изградњу референтног корпуса српског стандардног језика”, *Научни састанак слависта у Вукове дане, 12-16.9.2001.*, св. 31, Београд, Нови Сад: Међународни славистички центар, 2003., стр. 221–227.
- [51] М. Утвић, *Изградња референтног корпуса савременог српског језика*, дисертација, Универзитет у Београду, Филолошки факултет, 2013.
- [52] The Unicode Consortium, *The Unicode Standard, Version 6.2.0*, version 6.2.0, The Unicode Consortium, Mountain View, CA, 2012, ISBN: 978-1-936213-07-8. [Online]. Available: <http://www.unicode.org/versions/Unicode6.2.0/>.
- [53] Група за језичке технологије Универзитета у Београду. Корпус савременог српског језика (СрпКор), верзија СрпКор2013, веб адреса: <http://www.korpus.matf.bg.ac.rs>.
- [54] С. Krstev, D. Vitas, and A. Savary, “Prerequisites for a comprehensive dictionary of Serbian compounds”, in *Advances in Natural Language Processing*, Springer, 2006, pp. 552–563.
- [55] М. Gross, “The construction of local grammars”, *Finite-state Language Processing*, p. 329, 1997.
- [56] Д. Витас, *Преводиоци и интерпретатори (Увод у теорију и методе компилације програма језика)*. Београд: Математички факултет, 2006.
- [57] Д. Витас, „О елементарној морфографемској класи”, *Научни састанак слависта у Вукове дане*, св. 26, бр. 2, стр. 195–206, 1997.
- [58] Д. Кликовас, *Језик и моћ: огледи из социолингвистике и стилистике*. Библиотека XX век, 2008.
- [59] САНУ, *Речник српскохрватског књижевног и народног језика*, књ. I–XVIII. Београд: Српска академија наука и уметности, 1959–2008.
- [60] Р. Јовановић и Л. Атанацковић, *Систематски речник српскохрватског језика*. Нови Сад: Матица Српска, 1980.
- [61] С. Fellbaum, *WordNet: An Electronic Lexical Database*, ser. Language, Speech, and Communication. MIT Press, 1998, ISBN: 978-0-262-06197-1.

- [62] S. Koeva, C. Krstev, and D. Vitas, "Morpho-semantic relations in WordNet – a case study for two Slavic languages", in *Proceedings of Global WordNet Conference 2008*, University of Szeged, 2008, pp. 239–253.
- [63] Д. Витас, „О проблему не(пре)познате речи у обради текстова на српском језику”, *Зборник Матице Српске за филологију и лингвистику*, св. 50, бр. 1–2, стр. 111–120, 2007.
- [64] Б. Ћорић, *Творба именица у српском језику*. Београд: Друштво за српски језик и књижевност Србије, 2008.
- [65] R. Grishman and B. Sundheim, "Message understanding conference – 6: a brief history", in *COLING*, vol. 1, Copenhagen, Denmark: Association for Computational Linguistics, 1996, pp. 466–471. DOI: 10.3115/992628.992709.
- [66] S. Tatar and I. Cicekli, "Automatic rule learning exploiting morphological features for named entity recognition in Turkish", *J. Information Science*, vol. 37, no. 2, pp. 137–151, 2011. DOI: 10.1177/0165551511398573.
- [67] G. Rehm and H. Uszkoreit, Eds., *META-NET White Paper Series: Europe's Languages in the Digital Age*. Springer, Heidelberg, New York, Dordrecht, London, 2012. [Online]. Available: <http://www.meta-net.eu/whitepapers>.
- [68] C. Krstev, I. Obradović, M. Utvić, and D. Vitas, "A system for named entity recognition based on local grammars", *Journal of Logic and Computation*, vol. 24, no. 2, pp. 473–489, 2014. DOI: 10.1093/logcom/exs079.
- [69] Publications Office of the European Union, *Interinstitutional Style Guide*. Luxembourg: Publications Office, 2011, ISBN: 978-92-78-40701-8. DOI: 10.2830/36616. [Online]. Available: <http://publications.europa.eu/code/en/en-000100.htm>.
- [70] D. Vitas, „Ka programu za automatsko menjanje i dopunjavanje zakonskih tekstova”, *Računalniška obdelava lingvističnih podatkov*, Bled, 1982., str. 467–478.
- [71] T. Arnold-Moore, "Automatic generation of amendment legislation", in *Proceedings of the 6<sup>th</sup> International Conference on Artificial Intelligence and Law*, ser. ICAIL '97, Melbourne, Australia: ACM, 1997, pp. 56–62, ISBN: 978-0-89791-924-1. DOI: 10.1145/261618.261631.

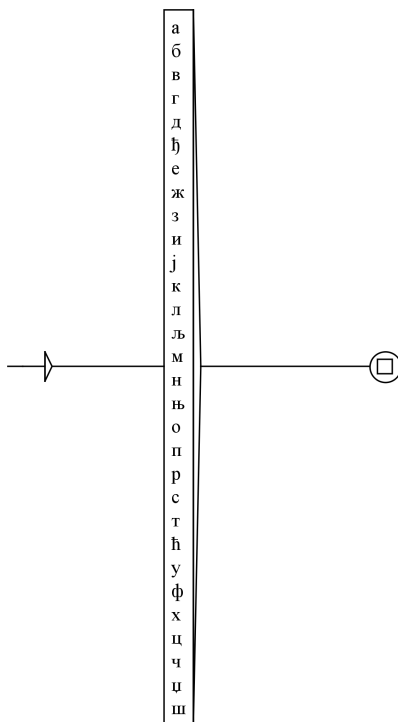
- [72] E. de Maat, S. van de Ven, R. Winkels, and T. van Engers, “Automated handling of amending documents and resulting consolidations”, in *Proceedings of the 2009 Conference on Legal Knowledge and Information Systems: JURIX 2009: The Twenty-Second Annual Conference*, Amsterdam, The Netherlands, The Netherlands: IOS Press, 2009, pp. 116–125, ISBN: 978-1-60750-082-7. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1671082.1671097>.
- [73] N. Friburger and D. Maurel, “Finite-state transducer cascades to extract named entities in texts”, *Theoretical Computer Science*, vol. 313, no. 1, pp. 93–104, 2004. DOI: 10.1016/j.tcs.2003.10.007.
- [74] J. Rumbaugh, I. Jacobson, and G. Booch, *Unified Modeling Language Reference Manual (2<sup>nd</sup> Edition)*. Pearson Higher Education, 2004, ISBN: 978-0-321-24562-5.
- [75] S. Paumier, “Why academic software should be open source”, *Infotheca*, vol. 10, no. 1–2, 70a–74a, 2009.
- [76] Д. Витас, „Генерисање именичких облика у српскохрватском језику”, *Информатика*, св. 3, бр. 81, стр. 49–55, 1981.

# Прилог А

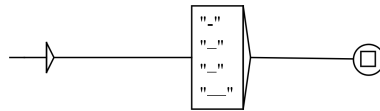
## Синтаксни графови

На бази синтаксних правила која су описана у овој дисертацији, уз помоћ *Unitex* софтверског алата, израђени су синтаксни графови чији приказ следи.

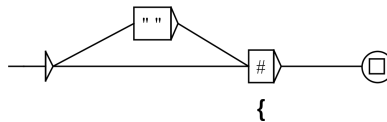
### А.1 Графови који се као подграфови користе на разним местима



Граф 1: Слово (Slovo)

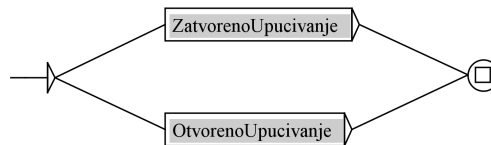


Граф 2: Хоризонтална црта (Crtica)

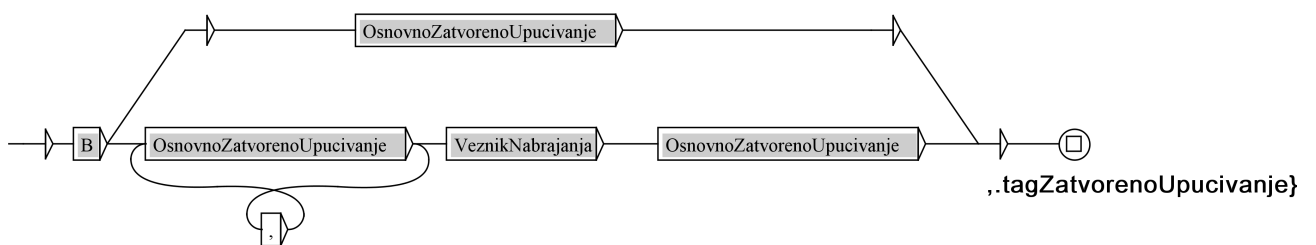


Граф 3: Замена за кућицу „<E>/{}” са контролом размака (B)

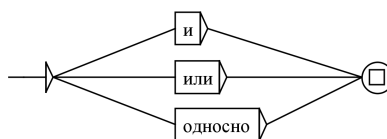
## A.2 Упућујуће фразе



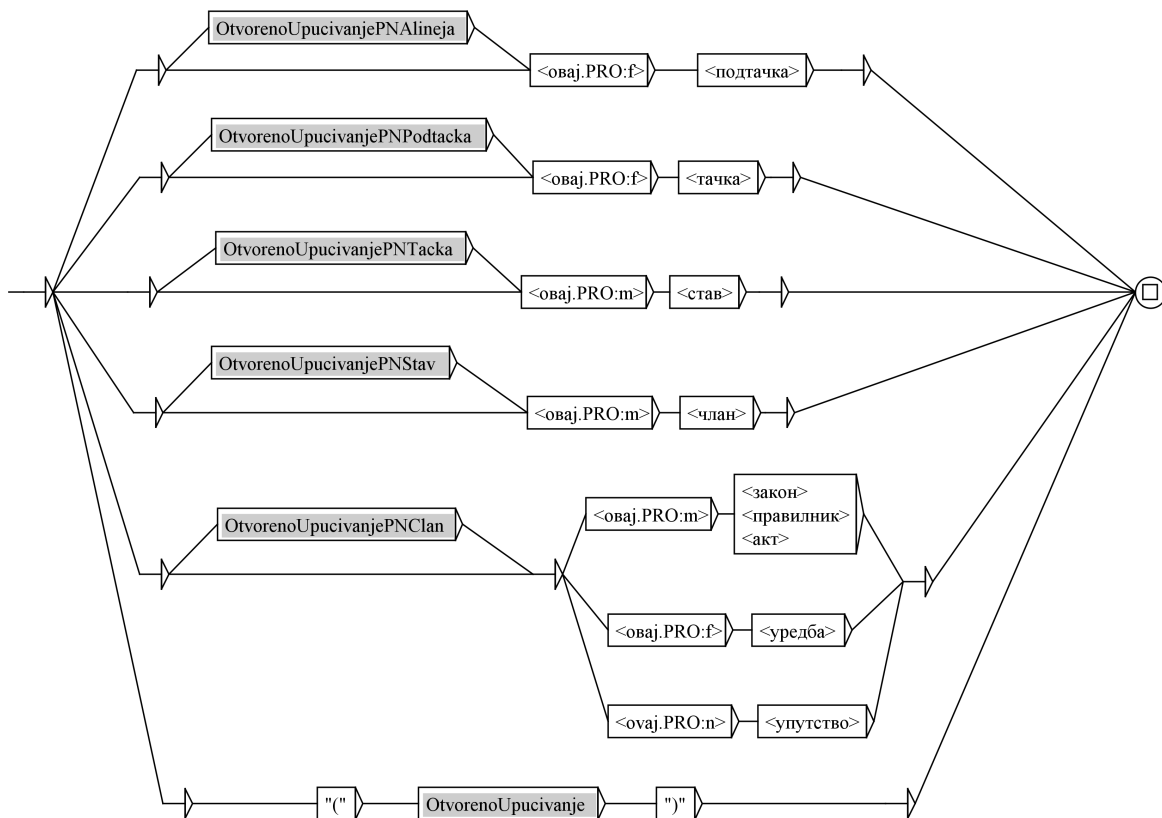
Граф 4: Затворено или отворено упућивање (Upscivanje)



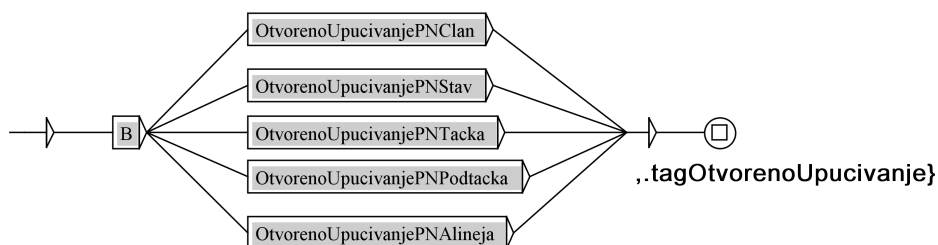
Граф 5: Затворено упућивање (ZatvorenoUpscivanje)



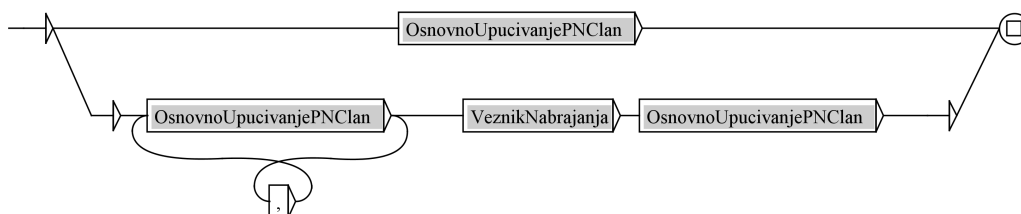
Граф 6: Везник набрајања (VeznikNabrajanja)



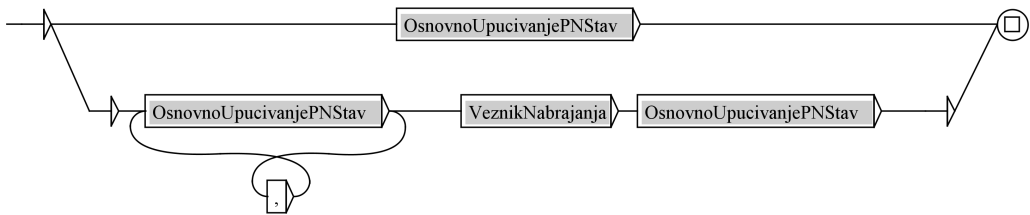
Граф 7: Основно затворено упућивање (OsnovnoZatvorenoUpucivanje)



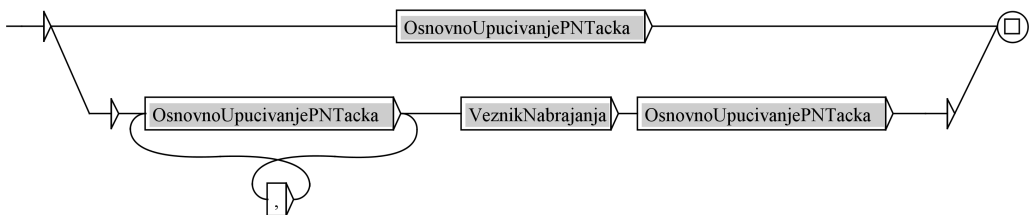
Граф 8: Отворено упућивање (OtvorenoUpucivanje)



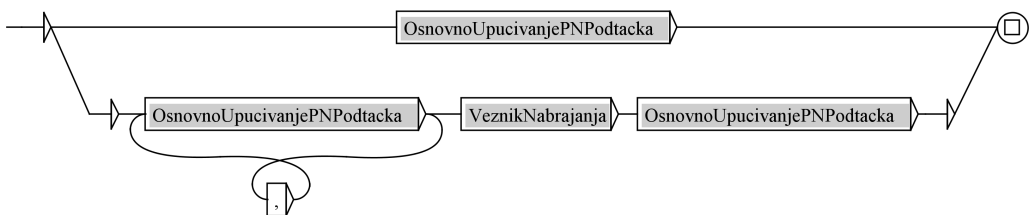
Граф 9: Отворено упућивање почетног нивоа члана (OtvorenoUpucivanjePNClan)



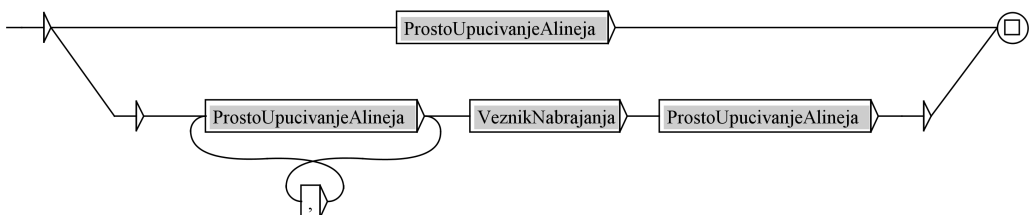
Граф 10: Отворено упућивање почетног нивоа става  
(OtvorenoUpucivanjePNStav)



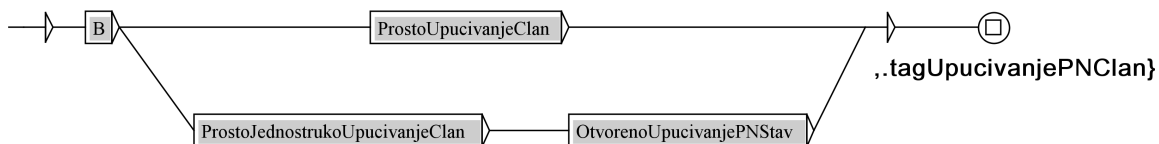
Граф 11: Отворено упућивање почетног нивоа тачке  
(OtvorenoUpucivanjePNTacka)



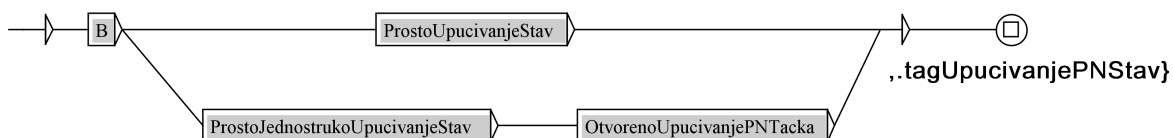
Граф 12: Отворено упућивање почетног нивоа подтачке  
(OtvorenoUpucivanjePNPodtacka)



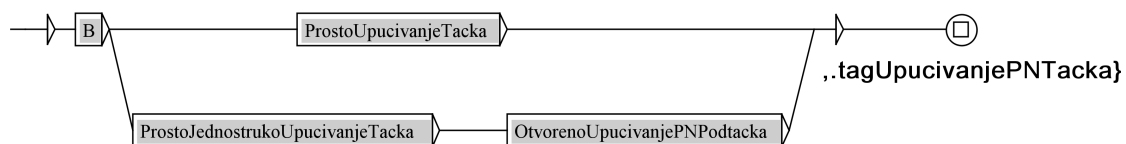
Граф 13: Отворено упућивање почетног нивоа алинеје  
(OtvorenoUpucivanjePNAlineja)



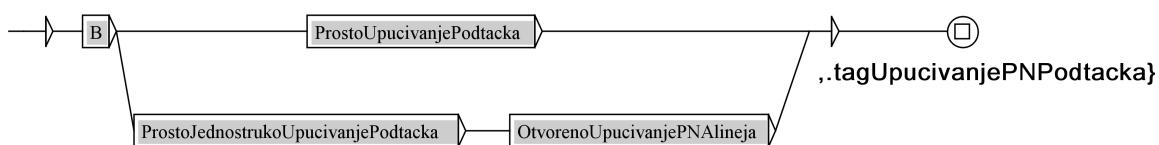
Граф 14: Основно упућивање понтекстног нивоа члана  
(OsnovnoUpucivanjePNClan)



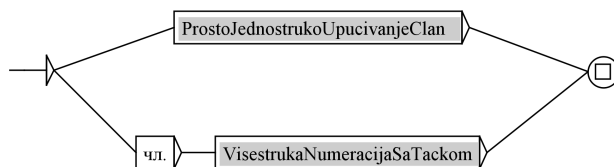
Граф 15: Основно упућивање почетног нивоа става  
(OsnovnoUpucivanjePNStav)



Граф 16: Основно упућивање почетног нивоа тачке  
(OsnovnoUpucivanjePNTacka)

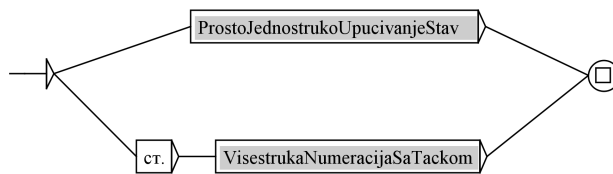


Граф 17: Основно упућивање почетног нивоа подтачке  
(OsnovnoUpucivanjePNPodtacka)

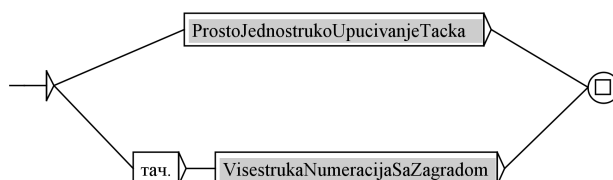


Граф 18: Просто упућивање нивоа члана (ProstoUpucivanjeClan)

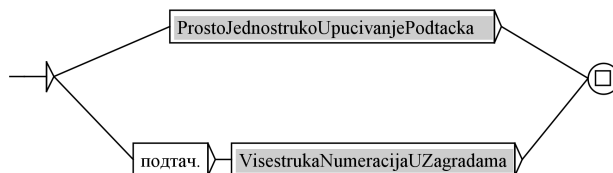




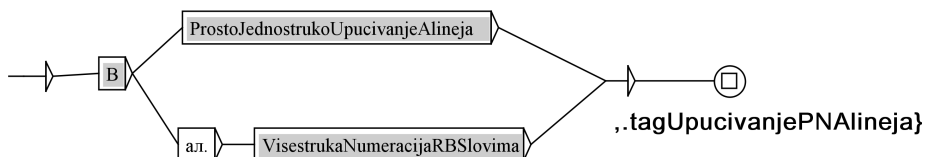
Граф 19: Просто упућивање нивоа става (ProstoUpucivanjeStav)



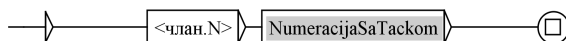
Граф 20: Просто упућивање нивоа тачке (ProstoUpucivanjeTacka)



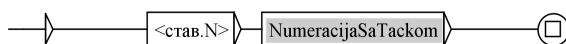
Граф 21: Просто упућивање нивоа подтачке (ProstoUpucivanjePodtacke)



Граф 22: Просто упућивање нивоа алинеје (ProstoUpucivanjeAlineja)

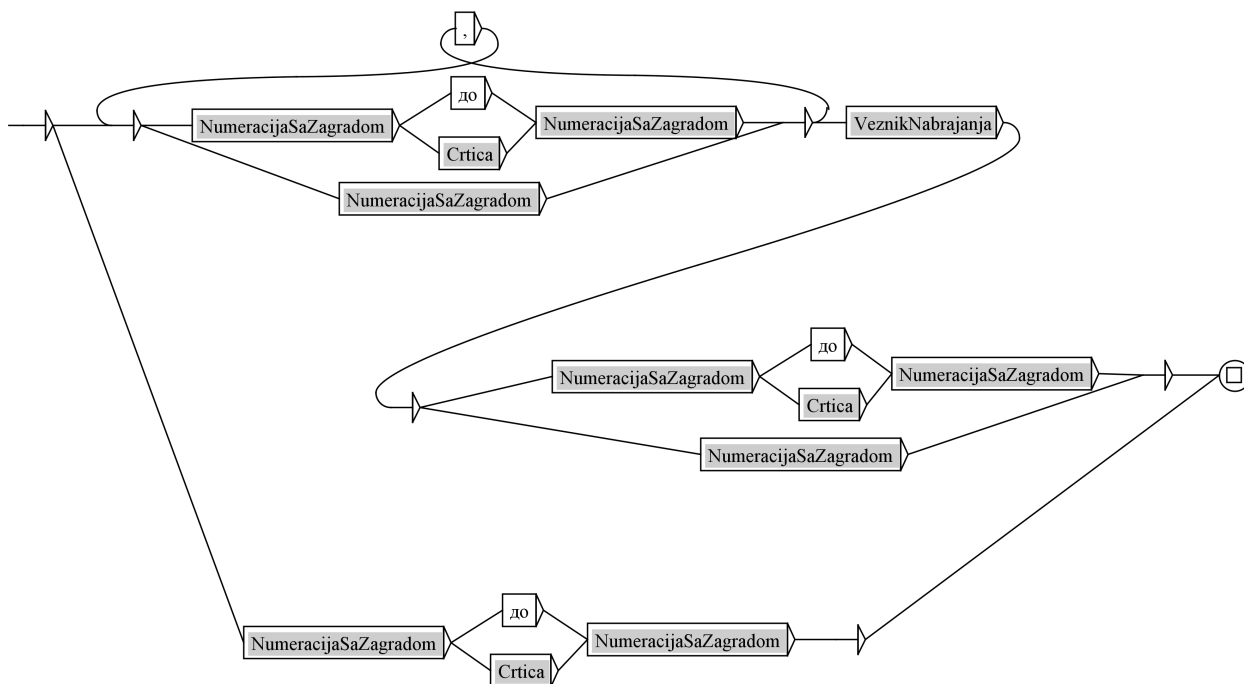


Граф 23: Просто једноструко упућивање на члан  
(ProstoJednostrukoUpucivanjeClan)

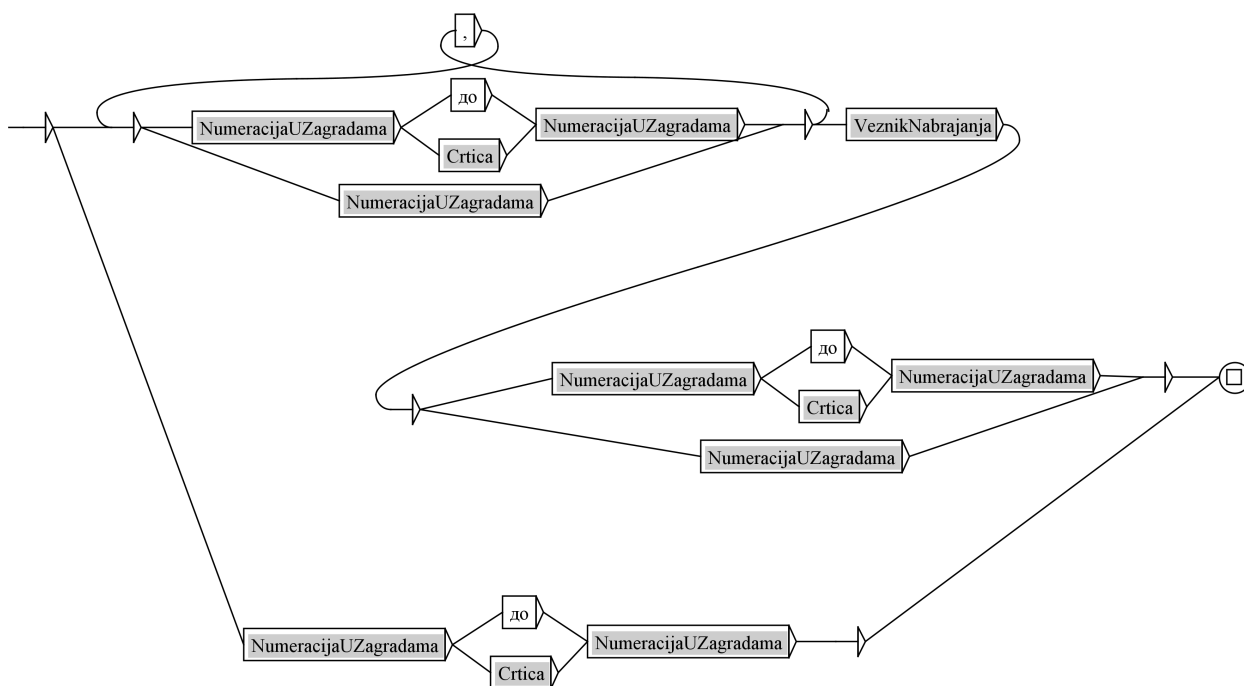


Граф 24: Просто једноструко упућивање на став  
(ProstoJednostrukoUpucivanjeStav)

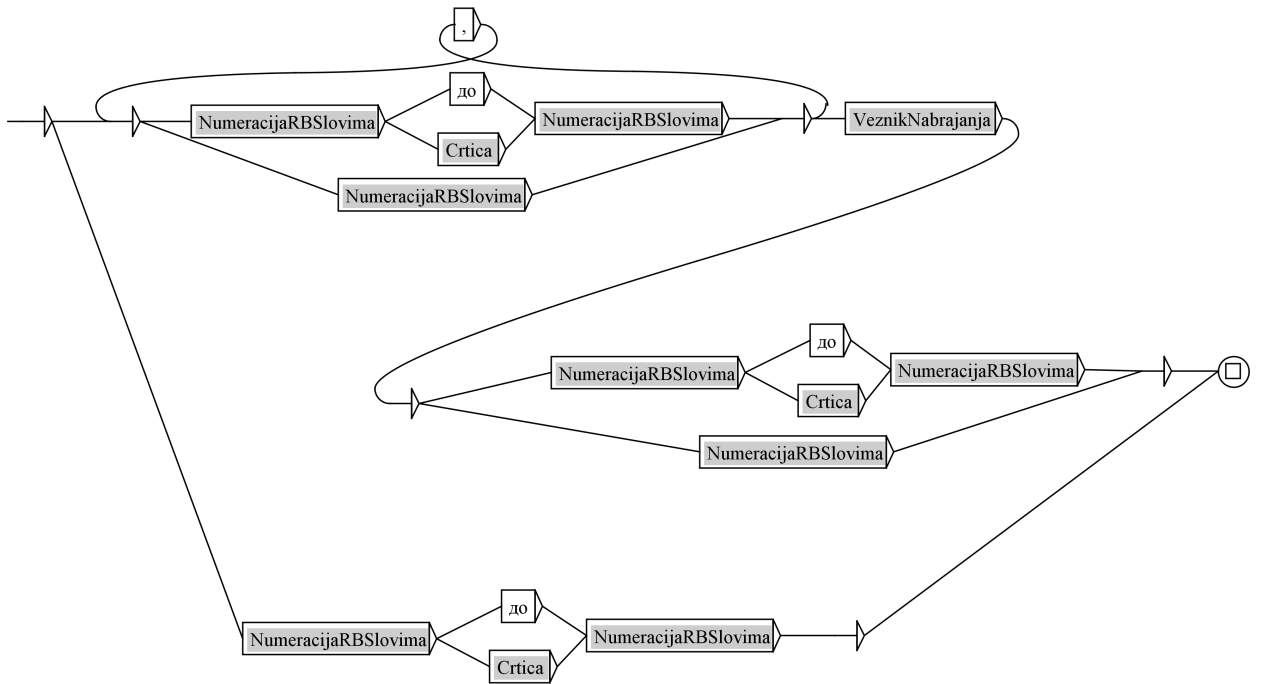




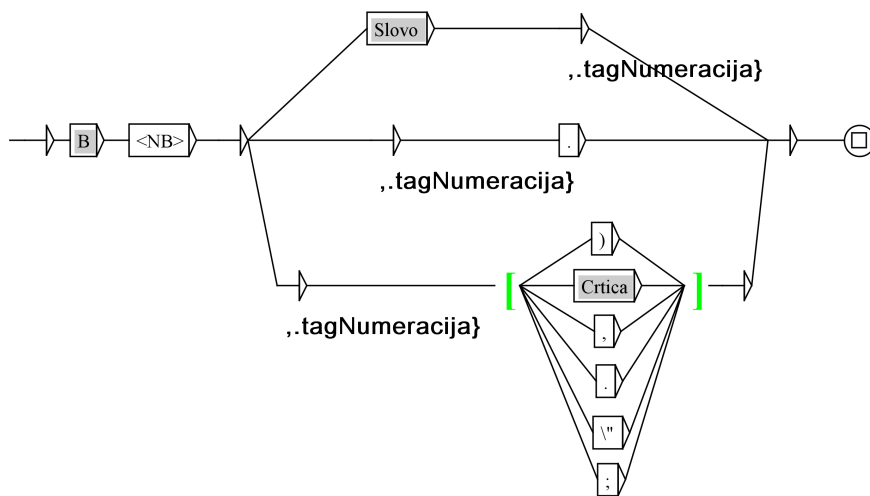
Граф 29: Вишеструка нумерација са десном заградом иза бројева  
(VisestrukaNumeracijaSaZagradom)



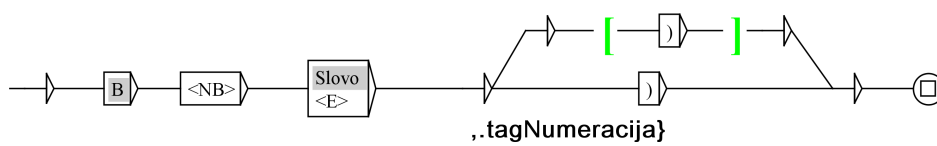
Граф 30: Вишеструка нумерација са бројевима у заградама  
(VisestrukaNumeracijaUZagradama)



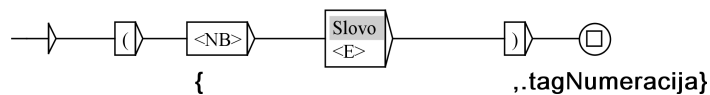
Граф 31: Вишеструка нумерација са редним бројевима исписаним словима  
(VisestrukaNumeracijaRBSlovima)



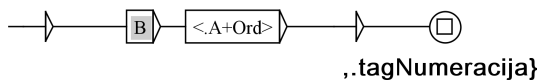
Граф 32: Нумерација са тачком иза броја (NumeracijaSaTackom)



Граф 33: Нумерација са десном заградом иза броја (NumeracijaSaZagradom)

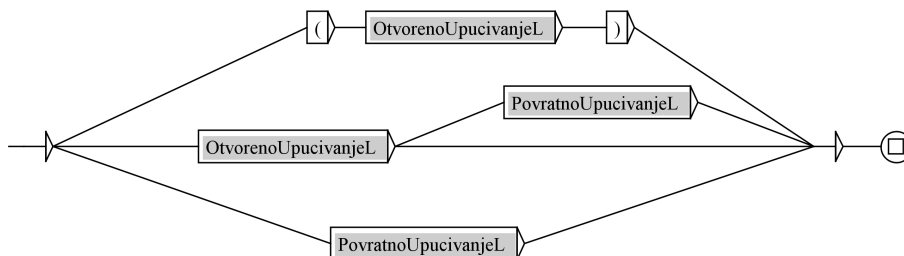


Граф 34: Нумерација са бројем у заградама (NumeracijaUZagradaма)

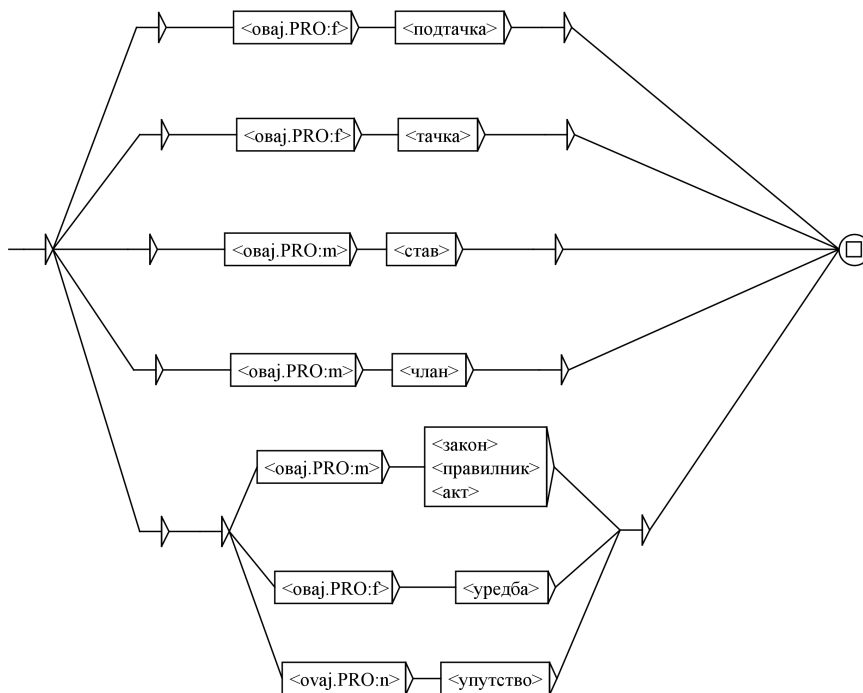


Граф 35: Нумерација са редним бројем исписаним словима (NumeracijaRBSlovима)

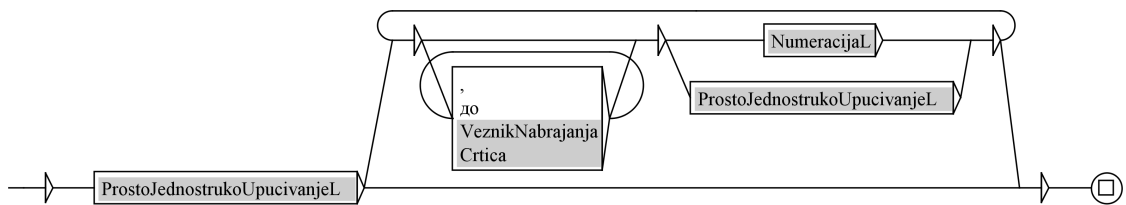
### А.3 Фразе сличне упућујућим



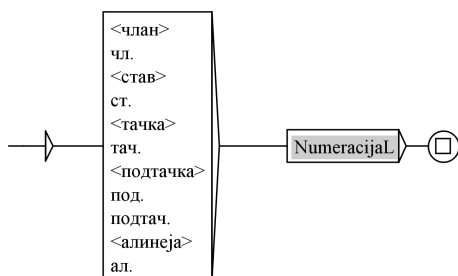
Граф 36: Упућујућа фраза (UpscivanjeL)



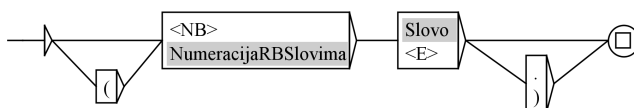
Граф 37: Повратно упућивање (PovratnoUpscivanjeL)



Граф 38: Отворено ушућивање (OtvorenoUpucivanjeL)



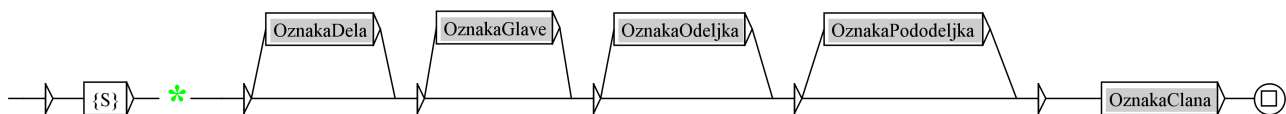
Граф 39: Просто једноструко ушућивање (ProstoJednostrukoUpucivanjeL)



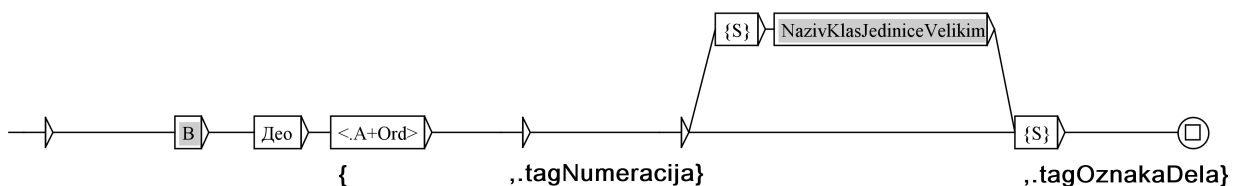
Граф 40: Просто једноструко ушућивање (NumeracijaL)

## A.4 Логичка структура прописа

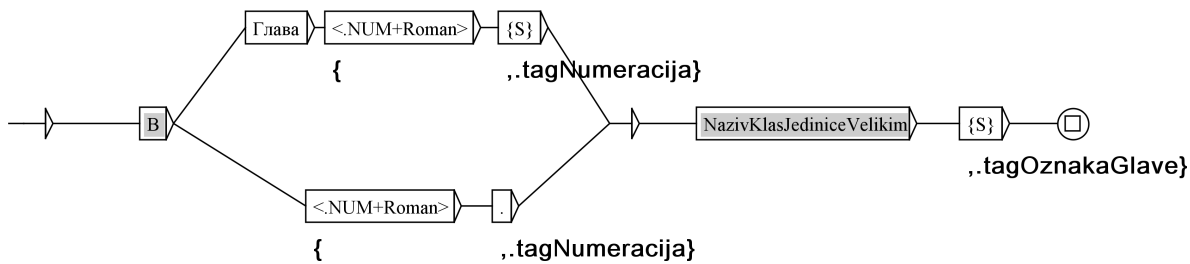
### A.4.1 Блок изнад члана



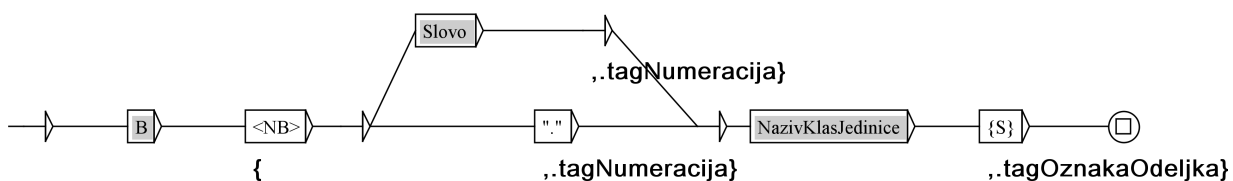
Граф 41: Блок изнад члана (BlokIznadClana)



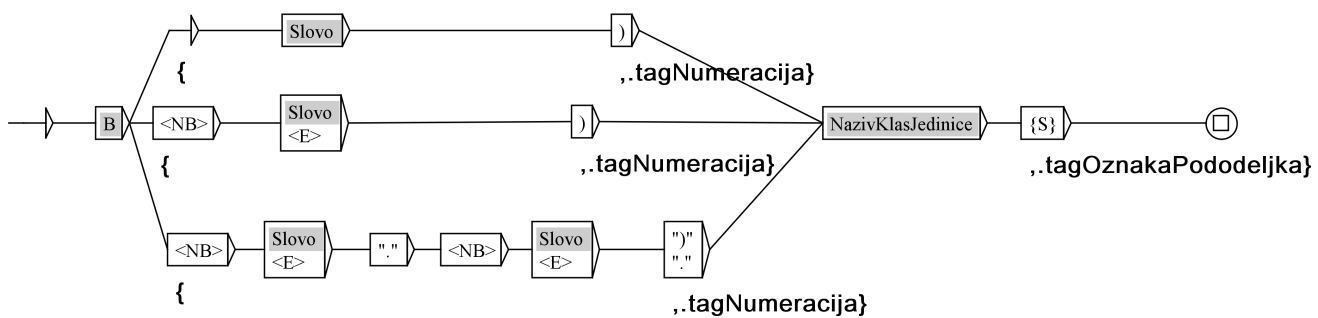
Граф 42: Ознака дѐла (OznakaDela)



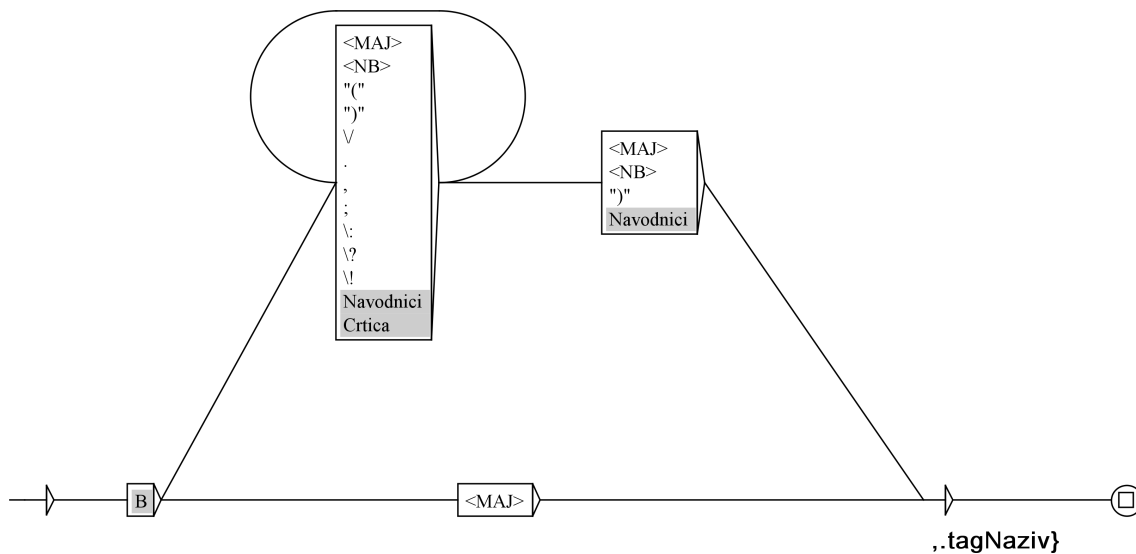
Граф 43: Ознака главе (OznakaGlave)



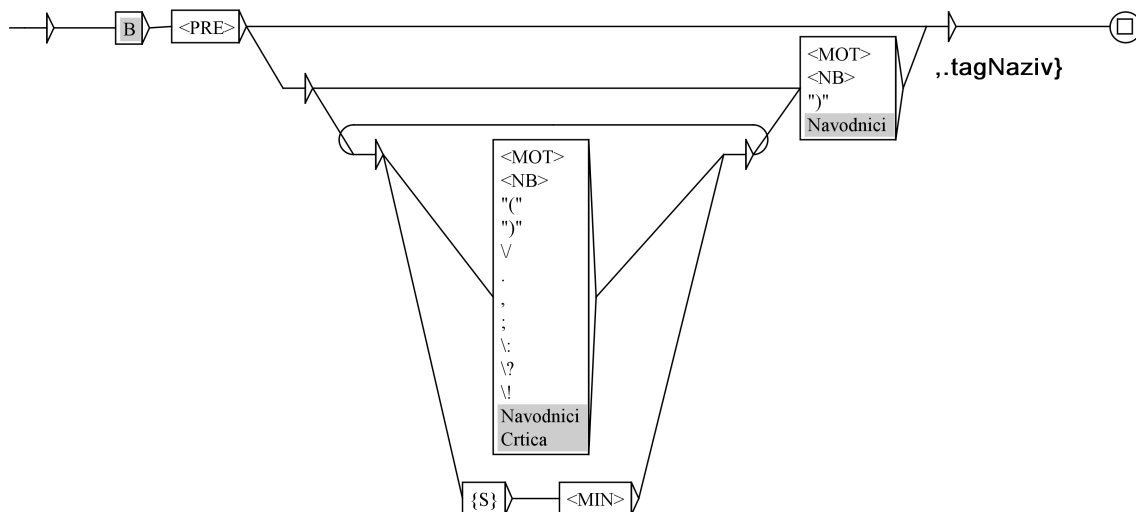
Граф 44: Ознака одељка (OznakaOdeljka)



Граф 45: Ознака подељка (OznakaPododeljka)



Граф 46: Назив класификационе јединице исписан великим словима  
(NazivKlasJediniceVelikim)



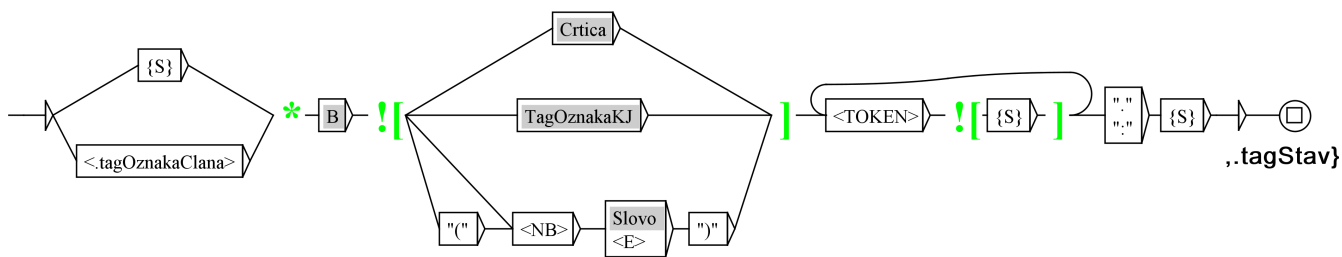
Граф 47: Назив класификационе јединице (NazivKlasJedinice)

#### A.4.2 Унутрашња структура члана

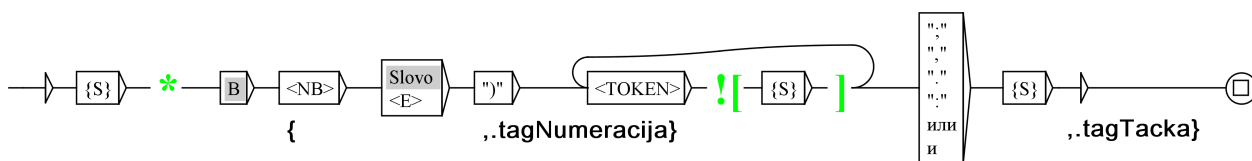


Граф 48: Класификационе јединице унутар члана (UzeKlasJedinice)

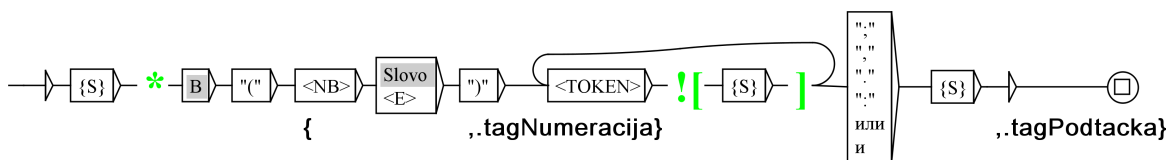




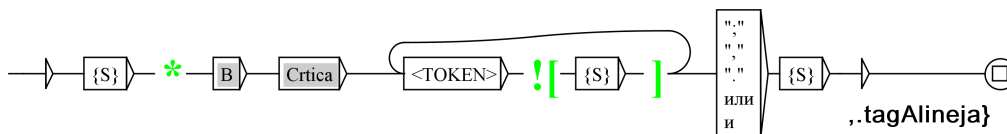
Граф 49: Став (Stav)



Граф 50: Тачка (Tacka)



Граф 51: Подтачка (Podtacka)



Граф 52: Алинеја (Alineja)



Граф 53: Претходно препознате шире класификационе јединице (TagOznakaKJ)



# Прилог Б

## CasSys формат датотеке

У овом прилогу је дата спецификација трансдукторских каскада које су описане у поглављу 6 у CasSys формату датотеке. CasSys користи апсолутне путање у именима датотека за графове које реферише, при чему нуди посебну операцију за увоз датотеке у CasSys формату, ком приликом се почетни део путања прилагоди директоријуму у коме корисник смешта језичке ресурсе за одређен језик.

### Б.1 Садржај датотеке *propisi.csc*

```
"C:\unitex\Serbian-Cyrillic\Graphs\BlokIznadClana.fst2" Merge Enabled 1  
"C:\unitex\Serbian-Cyrillic\Graphs\Upucivanje.fst2" Merge Enabled 1  
"C:\unitex\Serbian-Cyrillic\Graphs\UzeKlasJedinice.fst2" Merge Enabled 1  
"C:\unitex\Serbian-Cyrillic\Graphs\PocetakAkta.fst2" Merge Enabled 1
```

# Прилог В

## XSLT спецификација

У овом прилогу је дата XSLT спецификација путем које се резултат примена каскада из прилога Б, што укључује графове из прилога А, преводи у циљни XML формат синтаксног дрвета (види поглавље 6).

### В.1 Садржај датотеке *propis.xsl*

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet version="1.0"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform">

  <xsl:template match="/">
    <Propis xmlns:xhtml="http://www.w3.org/1999/xhtml">
      <xsl:apply-templates/>
    </Propis>
  </xsl:template>

  <xsl:template match="csc[code[substring(.,1,3)='tag']]">
    <xsl:element name="{substring(code[substring(.,1,3)='tag'],4)}">
      <xsl:apply-templates select="form"/>
    </xsl:element>
  </xsl:template>
```

```
<xsl:template match="csc[form = 's' and not(code)]">
  <br/>
</xsl:template>

<xsl:template match="/Propis/text()[normalize-space()]">
  <Neprepoznato>
    <xsl:value-of select="."/>
  </Neprepoznato>
</xsl:template>

</xsl:stylesheet>
```

# Прилог Г

## Списак закона у корпусу

У табели Г.1 приказани су сви закони чији текстови се налазе у корпусу. За сваки закон је приказано седам статистичких показатеља (за више детаља погледајте одељак 3.2):

1. колико пута се у закону појављују речи — овде се под речју подразумева секвенца узастопних слова, без обзира да ли она представља реч српског језика или не;
2. колико има различитих речи;
3. колико пута се појављују бројеви — овде се под бројем подразумева секвенца узастопних цифара;
4. колико има различитих бројева;
5. колико пута се појављују знакови — овде се под знаком подразумева појединачни карактер који није ни слово ни цифра ни размак;
6. колико има различитих знакова;
7. колико пута се појављују размаци — овде се под размаком подразумева карактер за размак или карактер за нови ред.

Табела Г.1: Списак закона у корпусу са статистикама

редни број, назив и година доношења закона	речи	бројева	знакова	разм.
1. Законик о кривичном поступку (2006)	79003	7063	4287 562	13778 10 83237
2. Законик о кривичном поступку (2011)	67002	6179	3098 619	9912 12 70046
3. Закон о привредним друштвима (2011)	66042	5082	3516 613	9281 12 69405
4. Закон о безбедности саобраћаја на путевима (2009)	55090	5505	5448 365	11513 14 60355
5. Закон о тржишту капитала (2010)	50820	5290	2477 317	7411 12 53305
6. Закон о парничном поступку (2011)	37862	4157	1444 513	4862 10 39261
7. Закон о лековима и медицинским средствима (2010)	35114	4009	1539 240	5104 11 36658
8. Царински Закон (2009)	34298	4736	1643 321	5019 17 36025
9. Закон о извршењу и обезбеђењу (2010)	33333	4248	1136 370	4305 11 34457
10. Закон о прекршајима (2013)	33110	4557	1064 350	3928 10 34133
11. Закон о тржишту хартија од вредности и других финансијских инструмената (2006)	32410	3377	1924 276	5327 13 34055
12. Закон о јавним набавкама (2012)	31291	4258	1595 191	4718 15 32849
13. Закон о безбедности и интероперабилности железнице (2013)	31029	4617	2296 187	6268 14 32931
14. Закон о здравственој заштити (2005)	30272	4176	1410 287	5020 9 31616
15. Закон о планирању и изградњи (2009)	29315	3925	1567 234	4937 16 30656
16. Закон о здравственом осигурању (2005)	29241	3716	1586 267	4841 12 30831
17. Закон о основама система образовања и васпитања (2009)	28837	4066	1339 204	5169 11 30037
18. Закон о енергетици (2011)	28295	3573	1877 222	4690 14 30112
19. Закон о пловидби и лукама на унутрашњим водама (2010)	27002	4236	1597 302	4640 14 28378
20. Закон о стечају (2009)	26943	3809	1043 219	3241 11 27960
21. Закон о прекршајима (2005)	26473	4024	780 314	3205 12 27193
22. Закон о водама (2010)	26375	3920	1774 236	5065 11 28019
23. Закон о ваздушном саобраћају (2010)	24915	3923	920 287	3575 14 25565
24. Закон о патентима (2011)	24010	3192	1051 192	3020 13 25015
25. Закон о спорту (2010)	23706	3887	1137 205	3948 13 24818
26. Закон о електронским комуникацијама (2010)	22739	3595	1081 161	3731 11 23614
27. Закон о поморској пловидби (2011)	22003	3829	1232 230	3481 14 23092
28. Закон о изменама и допунама Закона о републичким административним таксама (2008)	21698	3280	4044 422	4742 14 24669
29. Закон о банкама (2005)	21535	2955	1280 161	3163 13 22795
30. Закон о социјалној заштити (2010)	21535	3349	892 240	2991 12 22400

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

редни број, назив и година доношења закона	речи	бројева	знакова	разм.
31. Закон о рударству и геолошким истраживањима (2011)	21339	3314	1166 184	3689 11 22384
32. Закон о заштити потрошача (2010)	21205	3452	986 161	2920 12 22181
33. Закон о државној припадности и упису пловила (2012)	21157	2937	833 221	2879 13 21959
34. Закон о извршењу кривичних санкција (2014)	20902	3767	665 288	2690 12 21534
35. Закон о облигационим и основама својинско-правних односа у ваздушном саобраћају (2011)	20330	3095	856 194	3001 12 20856
36. Закон о полицији (2005)	19994	3870	997 205	2956 13 20944
37. Закон о заштити природе (2009)	19807	3968	955 141	3610 12 20671
38. Закон о безбедности хране (2008)	19429	2315	1395 114	3621 12 20740
39. Закона о ветеринарству (2005)	18562	3562	1410 176	3897 12 19952
40. Закон о ванредним ситуацијама (2009)	18495	3366	953 163	3228 13 19388
41. Закон о јавном бележништву (2011)	18367	3298	653 184	2650 11 19028
42. Закон о јавним набавкама (2008)	17826	3244	921 141	2655 12 18689
43. Закон о ауторском и сродним правима (2009)	17421	3283	1172 230	3019 12 18416
44. Закон о шумама (2010)	17167	3516	1137 128	4130 12 18246
45. Закон о буџетском систему (2009)	17094	2642	890 125	3061 13 17960
46. Закон о управљању отпадом (2008)	16888	2848	950 121	3076 11 17760
47. Закон о спречавању прања новца и финансирања тероризма (2008)	16777	2257	1510 104	3529 12 18213
48. Закон о државном премеру и катастру (2009)	16667	2961	1508 218	4099 11 18002
49. Закон о војсци Србије (2007)	16549	2957	784 213	2470 14 17378
50. Закон о изменама и допунама Законика о кривичном поступку (2009)	15775	2694	1651 310	4580 13 17246
51. Закон о трансплантацији ћелија и ткива (2009)	15705	2378	618 115	2493 11 16301
52. Закон о железници (2013)	15244	2881	779 120	2398 11 15996
53. Закон о дувану (2005)	15213	1807	1164 122	2826 14 16296
54. Закон о високом образовању (2005)	15099	2649	711 143	2500 14 15777
55. Закон о играма на срећу (2011)	14694	2103	992 150	2606 12 15628
56. Закон о туризму (2008)	14675	2665	1170 148	2985 12 15725
57. Закон о здрављу биља (2008)	14475	2234	914 120	2791 10 15317
58. Закон о ревизији (2013)	14217	2300	896 126	2422 10 15095
59. Закон о средствима за заштиту биља (2008)	14133	2207	905 103	2373 10 14998
60. Закон о психоактивним контролисаним супстанцама (2010)	13982	2289	813 124	2287 13 14747
61. Закон о јавним путевима (2005)	13834	2311	1138 111	3203 10 14929

*наставак на следећој страни*



Табела Г.1 – наставак са претходне стране

редни број, назив и година доношења закона	речи	бројева	знакова	разм.
62. Закон о обавезном осигурању у саобраћају (2009)	13134	2181	896 135	2134 11 13964
63. Закон о хемикалијама (2009)	12893	2019	751 115	1830 12 13604
64. Закон о дивљачи и ловству (2009)	12816	2416	681 121	2034 12 13447
65. Закон о заштити ваздуха (2009)	12597	2399	853 105	2174 10 13288
66. Закон о ученичком и студентском стандарду (2009)	12515	2338	549 110	2220 14 13004
67. Закон о изменама и допунама Закона о планирању и изградњи (2011)	12491	2418	916 154	2798 15 13331
68. Закон о транспорту опасног терета (2010)	12442	1652	1185 113	2410 11 13537
69. Закон о ефикасном коришћењу енергије (2013)	12399	2424	755 106	2056 12 13087
70. Закон о заштити од јонизујућих зрачења и о нуклеарној сигурности (2008)	12279	2054	692 106	1945 13 12923
71. Закон о државним службеницима (2005)	12007	2254	400 198	1472 14 12396
72. Закон о добробити животиња (2009)	12002	2335	848 94	2473 10 12844
73. Закон о јавно – приватном партнерству и концесијама (2011)	11944	2491	425 80	1715 14 12311
74. Закон о националним саветима националних мањина (2009)	11856	2246	568 144	1782 14 12360
75. Закон о војној, радној и материјалној обавези (2009)	11851	2305	574 161	1643 12 12361
76. Закон о изменама Закона којима су одређене новчане казне за привредне преступе и прекршаје (2005)	11562	488	7537 219	10879 9 16042
77. Закон о сточарству (2008)	11494	2203	878 104	2297 10 12291
78. Закон о трансплантацији органа (2009)	11362	2025	486 103	1747 10 11816
79. Закон о одбрани (2007)	11285	2056	648 133	1959 12 11839
80. Закон о јавној својини (2011)	10857	1696	439 105	1683 10 11262
81. Закон о ознакама географског порекла (2009)	10812	1512	460 97	1422 11 11258
82. Закон о оглашавању (2005)	10736	2158	663 122	2065 11 11363
83. Закон о приватном обезбеђењу (2013)	10718	1984	826 105	2289 11 11467
84. Закон о средњем образовању и васпитању (2013)	10513	2342	380 117	1742 13 10846
85. Закон о изменама и допунама Кривичног законика (2009)	10416	1564	1286 324	4575 12 11664
86. Закон о инвестиционим фондовима (2005)	10406	2006	557 98	1639 10 10940

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

редни број, назив и година доношења закона	речи	бројева	знакова	разм.
87. Закон о основном образовању и васпитању (2013)	10372	2443	394 116	1740 11 10695
88. Закон о добровољним пензијским фондовима и пензијским плановима (2005)	10203	2047	609 84	1708 11 10810
89. Закон о научноистраживачкој делатности (2005)	10124	1850	486 120	1851 11 10575
90. Закон о утврђивању надлежности Аутономне Покрајине Војводине (2009)	10054	2010	570 97	1765 12 10622
91. Закон о вину (2009)	10044	1728	594 91	1952 12 10600
92. Закон о јавном тужилаштву (2008)	9700	1957	359 151	1298 12 10034
93. Закон о пољопривредном земљишту (2006)	9670	1959	597 104	1728 11 10234
94. Закон о заштити корисника финансијских услуга (2011)	9634	1688	538 64	1703 11 10199
95. Закон о министарствима (2014)	9574	2151	150 55	1639 11 9685
96. Закон о жиговима (2009)	9566	1672	439 100	1495 12 9979
97. Закон о тајности података (2009)	9534	1831	513 117	1683 13 10031
98. Закон о биоцидним производима (2009)	9481	1512	474 72	1331 9 9947
99. Закон о министарствима (2011)	9319	2124	102 45	1516 11 9395
100. Закон о лечењу неплодности поступцима биомедицински потпомогнутог оплођења (2009)	9187	1867	411 86	1405 12 9586
101. Закон о министарствима (2012)	8974	2043	98 43	1451 11 9043
102. Закон о изменама и допунама Закона о ветеринарству (2010)	8821	1632	1055 146	3514 13 9777
103. Закон о странцима (2008)	8742	1622	658 104	1633 11 9349
104. Закон о изменама и допунама Закона о заштити животне средине (2009)	8737	1835	815 113	2184 13 9466
105. Закон о рачуноводству и ревизији (2006)	8664	1881	443 88	1471 11 9087
106. Закон о министарствима (2008)	8657	2159	151 60	1445 11 8779
107. Закон о заштити од пожара (2009)	8504	2092	531 104	1421 11 8978
108. Закон о адвокатури (2011)	8410	2324	395 98	1283 11 8818
109. Закон о преузимању акционарских друштава (2005)	8402	1575	467 76	1305 14 8847
110. Закон о девизном пословању (2006)	8260	1228	835 85	1851 11 9159
111. Закон о трговини (2010)	8237	2197	456 99	1665 14 8617
112. Закон о изменама и допунама Закона о девизном пословању (2010)	8211	1292	958 97	2069 14 9213
113. Закон о враћању одузете имовине и обештећењу (2011)	8189	1981	571 102	1694 12 8684

наставак на следећој страни

Табела Г.1 – наставак са претходне стране

<b>редни број, назив и година доношења закона</b>	<b>речи</b>	<b>бројева</b>	<b>знакова</b>	<b>разм.</b>
114. Закон о изменама и допунама Закона о буџетском систему (2010)	8140	2055	608 79	1753 16 8699
115. Закон о заштити конкуренције (2009)	8119	2076	331 86	1194 10 8450
116. Закон о изменама и допунама Закона о поморској пловидби (2013)	8099	1907	594 115	1755 11 8609
117. Закон о правној заштити индустријског дизајна (2009)	7960	1626	321 89	1145 10 8260
118. Закон о министарствима (2007)	7898	2045	164 66	1409 10 8026
119. Закон о изменама и допунама Закона о добровољним пензијским фондовима и пензијским плановима (2010)	7796	1530	725 84	1850 15 8467
120. Закон о заштити и одрживом коришћењу рибљег фонда (2009)	7674	1695	566 75	1461 11 8222
121. Закон о заштити конкуренције (2005)	7585	1860	396 79	1194 10 7976
122. Закон о јавним скијалиштима (2005)	7573	1902	447 87	1332 8 8099
123. Закон о спољнотрговинском пословању (2005)	7466	1725	624 92	1804 13 8063
124. Закон о рачуноводству (2013)	7394	1699	405 71	1416 12 7797
125. Закон о запошљавању и осигурању за случај незапослености (2009)	7388	1672	507 122	1352 13 7844
126. Закон о локалној самоуправи (2007)	7327	1663	294 108	1102 12 7587
127. Закон о државној ревизорској институцији (2005)	7286	1849	292 67	1084 10 7576
128. Закон о метеоролошкој и хидролошкој делатности (2010)	7255	1615	370 62	1190 10 7556
129. Закон о изменама и допунама Закона о заштити природе (2010)	7250	1844	553 98	1893 12 7781
130. Закон о изменама и допунама Закона о извршењу кривичних санкција (2009)	7219	1715	597 211	2161 13 7843
131. Закон о амбалажи и амбалажном отпаду (2009)	7193	1642	388 60	1286 12 7543
132. Закон о Уставном суду (2007)	7151	1406	266 122	1032 9 7407
133. Закон о култури (2009)	7110	1574	343 90	1294 12 7456
134. Закон о цивилној служби (2009)	6989	1547	327 89	993 12 7288
135. Закон о Агенцији за борбу против корупције (2008)	6934	1506	320 90	1108 11 7280
136. Закон о удружењима (2009)	6905	1610	345 100	1078 12 7207

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

редни број, назив и година доношења закона	речи	бројева	знакова	разм.
137. Закон о изменама и допунама Закона о порезу на додату вредност (2012)	6897 1179	858 81	1998 15	7650
138. Закон о заштити државне границе (2008)	6863 1560	331 81	983 9	7181
139. Закон о изменама и допунама Закона о девизном пословању (2012)	6839 968	883 92	1953 11	7756
140. Закон о судијама (2008)	6783 1640	351 116	1098 12	7099
141. Закон о трансфузиолошкој делатности (2009)	6778 1428	398 70	1158 10	7158
142. Закон о међународној правној помоћи у кривичним стварима (2009)	6745 1380	417 114	1017 12	7155
143. Закон о изменама и допунама Закона о Народној Србије (2010)	6696 1388	537 106	1628 12	7207
144. Закон о изменама и допунама Закона о порезу на доходак грађана (2013)	6656 1447	718 126	1637 11	7306
145. Закон о јавним предузећима (2012)	6654 1444	328 83	1091 12	6958
146. Закон о изменама и допунама Закона о превозу у друмском саобраћају (2011)	6593 1086	995 70	2149 13	7442
147. Закон о средствима за исхрану биља и оплемењивачима земљишта (2008)	6551 1385	432 61	1120 13	6951
148. Закон о комуналним делатностима (2011)	6540 1693	326 65	951 12	6843
149. Закон о изменама и допунама Закона о порезима на имовину (2013)	6475 1219	624 62	1698 14	7055
150. Закон о уређењу судова (2008)	6318 1669	291 110	982 11	6564
151. Закон о заштити података о личности (2008)	6290 1543	355 70	1174 10	6632
152. Закон о изменама и допунама Закона о пореском поступку и пореској администрацији (2012)	6167 1192	576 106	1466 12	6555
153. Закон о одузимању имовине проистекле из кривичног дела (2013)	6161 1387	557 141	1207 10	6662
154. Закон о изменама и допунама Закона о националним саветима националних мањина (2014)	6063 1223	423 109	1220 12	6462
155. Закон о образовању одраслих (2013)	6044 1661	265 72	1117 12	6301
156. Закон о забрани развоја, производње, складиштења и употребе хемијског оружја и о његовом уништавању (2008)	6030 1231	822 133	1818 16	6567
157. Закон о азилу (2007)	6005 1539	262 84	836 12	6250
158. Закон о изменама и допунама Закона о порезу на добит предузећа (2009)	5954 1151	577 94	1990 12	6475

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

<b>редни број, назив и година доношења закона</b>	<b>речи</b>	<b>бројева</b>	<b>знакова</b>	<b>разм.</b>
159. Закон о Војнобезбедносној агенцији и Војно-обавештајној агенцији (2009)	5902	1423	302 67	1040 10 6185
160. Закон о заштити права оплемењивача биљних сорти (2009)	5875	1296	284 64	856 10 6139
161. Закон о изменама и допунама Закона о финансијском лизингу (2010)	5829	1311	379 45	1037 12 6108
162. Закон о јавним складиштима за пољопривредне производе (2008)	5818	1475	353 77	966 11 6148
163. Закон о матичним књигама (2008)	5790	1228	272 104	940 12 6045
164. Закон о задужбинама и фондацијама (2010)	5783	1372	227 82	855 10 6001
165. Закон о заштити лица са менталним сметњама (2013)	5767	1337	228 76	787 13 5982
166. Закон о подстицајима у пољопривреди и руралном развоју (2012)	5743	1272	374 65	976 12 6093
167. Закон о хипотеци (2005)	5723	1405	281 70	1024 10 6004
168. Закон о изменама и допунама Закона о пензијском и инвалидском осигурању (2010)	5721	1022	936 162	1886 13 6596
169. Закон о изменама и допунама Закона о научноистраживачкој делатности (2010)	5577	1254	407 93	1358 12 5892
170. Закон о одговорности правних лица за кривична дела (2008)	5533	1255	210 82	691 8 5744
171. Закон о библиотечко-информационој делатности (2011)	5528	1497	158 64	1026 12 5578
172. Закон о управним споровима (2009)	5411	1334	214 84	746 12 5622
173. Закон о супстанцама које се користе у недозвољеној производњи опојних дрога психотропних супстанци (2005)	5335	1061	320 58	954 9 5651
174. Закон о правима пацијената (2013)	5328	1276	270 66	944 12 5566
175. Закон о тржишном надзору (2011)	5324	1394	206 36	857 11 5570
176. Закон о изменама и допунама Закона о ванпарничном поступку (2014)	5305	1247	464 123	1411 11 5727
177. Закон о државној управи (2005)	5294	1298	157 97	633 14 5447
178. Закон о Народној скупштини (2010)	5216	1366	189 80	809 13 5426
179. Закон о локалним изборима (2007)	5216	1303	196 78	714 12 5389
180. Закон о седиштима и подручјима судова и јавних тужилаштава (2013)	5205	703	342 70	1384 8 5535
181. Закон о спољним пословима (2007)	5193	1465	164 72	832 11 5282

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

<b>редни број, назив и година доношења закона</b>	<b>речи</b>	<b>бројева</b>	<b>знакова</b>	<b>разм.</b>
182. Закон о изменама и допунама Закона о здравственом осигурању (2011)	5192	1121	541 136	1358 12 5707
183. Закон о старој и реткој библиотечкој грађи (2011)	5185	1277	277 80	863 11 5452
184. Закон о метрологији (2009)	5175	1236	250 56	783 11 5414
185. Закон о изменама и допунама Закона о туризму (2010)	5163	970	734 113	2115 11 5855
186. Закон о изменама и допунама Закона о банкама (2010)	5137	1179	513 101	1147 11 5589
187. Закон о техничким захтевима за производе и оцењивању усаглашености (2008)	5119	1088	355 54	905 12 5447
188. Закон о извршењу казне затвора за кривична дела организованог криминала (2009)	5092	1291	225 65	797 10 5312
189. Закон о арбитражи (2005)	5033	1220	191 90	689 9 5200
190. Закон о правобранилаштву (2014)	5008	1156	120 67	557 11 5126
191. Закон о контроли предмета од драгоцених метала (2011)	5000	1122	328 63	884 10 5302
192. Закон о здравственој исправности предмета опште употребе (2011)	4974	1251	313 71	836 11 5238
193. Закон о изменама и допунама Закона о пореском поступку и пореској администрацији (2007)	4947	1023	685 119	1627 11 5536
194. Закон о одузимању имовине проистекле из кривичног дела (2008)	4927	1239	402 116	901 10 5280
195. Закон о ракији и другим алкохолним пићима (2008)	4893	1202	283 64	910 11 5151
196. Закон о изменама и допунама Закона о пореском поступку и пореској администрацији (2008)	4892	1226	389 78	1178 13 5226
197. Закон о извршењу ванзаводских санкција и мера (2014)	4804	1247	169 69	625 9 4970
198. Закон о финансирању политичких активности (2011)	4782	1160	309 63	802 12 5071
199. Закон о заштити од буке у животној средини (2009)	4755	1164	317 62	841 10 5031
200. Закон о органској производњи (2009)	4743	1223	325 64	842 12 5042
201. Закон о заштити становништва од изложености дуванском диму (2009)	4709	1143	283 43	830 11 4983
202. Закон о коморама здравствених радника (2005)	4708	1124	364 69	953 10 5047

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

редни број, назив и година доношења закона	речи	бројева	знакова	разм.
203. Закон о предшколском васпитању и образовању (2009)	4670	1132	227 74	899 12 4888
204. Закон о забрани дискриминације (2009)	4664	1397	274 67	813 9 4887
205. Закон о Правосудној академији (2009)	4663	1142	134 66	693 13 4876
206. Закон о робним резервама (2013)	4659	1271	248 57	770 12 4892
207. Закон о изменама и допунама Закона о преузимању акционарских друштава (2011)	4637	1111	320 55	1013 11 4947
208. Закон о генетички модификованим организима (2008)	4553	1055	255 63	687 11 4791
209. Закон о изменама и допунама Закона о здравственој заштити (2011)	4482	1084	339 101	1015 12 4781
210. Закон о сарадњи са Међународним кривичним судом (2008)	4477	1142	134 50	515 9 4611
211. Закон о званичној статистици (2009)	4443	1328	172 73	660 11 4610
212. Закон о равноправности полова (2009)	4441	1485	186 60	624 10 4614
213. Закон о седиштима и подручјима судова и јавних тужилаштава (2008)	4413	633	260 50	1256 10 4682
214. Закон о посебним овлашћењима ради ефикасне заштите права интелектуалне својине (2005)	4378	1049	462 61	1044 13 4768
215. Закон о изменама и допунама Закона о буџетском систему (2012)	4376	883	529 81	1717 13 4847
216. Закон о Државном већу тужилаца (2008)	4367	1064	140 72	545 11 4528
217. Закон о јавном дугу (2005)	4355	1155	215 65	701 10 4559
218. Закон о пољопривреди и руралном развоју (2008)	4342	1280	241 56	785 12 4573
219. Закон о изменама и допунама Закона о међународном превозу у друмском саобраћају (2009)	4264	931	422 63	993 11 4573
220. Закон о регионалном развоју (2009)	4218	1106	281 65	858 11 4502
221. Закон о изменама и допунама Закона о пореском поступку и пореској администрацији (2013)	4207	947	747 102	1763 12 4753
222. Закон о признавању сорти пољопривредног биља (2010)	4203	1016	247 57	692 10 4428
223. Закон о финансирању локалне самоуправе (2006)	4201	1216	260 78	762 11 4464
224. Закон о уџбеницима и другим наставним средствима (2009)	4198	1127	252 65	842 12 4413
225. Закон о професионалној рехабилитацији и запошљавању особа са инвалидитетом (2008)	4167	1041	267 70	690 12 4412

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

<b>редни број, назив и година доношења закона</b>	<b>речи</b>	<b>бројева</b>	<b>знакова</b>	<b>разм.</b>
226. Закон о Комуналној полицији (2009)	4144	1304	162 53	587 10 4301
227. Закон о изменама и допунама Закона о избе- глицама (2010)	4076	1109	239 39	773 15 4283
228. Закон о управној инспекцији (2011)	4076	1156	176 60	561 11 4256
229. Закон о изменама и допунама Закона о порезу на добит правних лица (2012)	4049	957	531 84	1275 12 4485
230. Закон о Изменама и допунама Закона о рачу- новодству и ревизији (2009)	4031	1042	427 71	1114 10 4414
231. Закон о Високом савету судства (2008)	3967	1065	149 73	553 10 4133
232. Закон о изменама и допунама Закона о пло- видби и лукама на унутрашњим водама (2012)	3957	1159	363 93	1121 10 4280
233. Закон о детективској делатности (2013)	3955	1030	280 47	765 10 4219
234. Закон о изменама и допунама Закона о буџет- ском систему (2005)	3912	955	353 59	1125 13 4258
235. Закон о изменама и допунама Закона о буџет- ском систему (2006)	3910	1122	323 61	1022 10 4218
236. Закон о пиву (2009)	3909	800	319 44	842 10 4220
237. Закон о општој безбедности производа (2008)	3893	1016	267 36	721 9 4145
238. Закон о употреби Војске Србије и других снага одбране у мултинационалним операцијама ван граница Републике Србије (2009)	3850	992	172 48	555 10 4029
239. Закон о изменама и допунама закона о порезу на доходак грађана (2006)	3846	825	475 78	1269 12 4258
240. Закон о путним исправама (2007)	3770	1037	206 71	562 10 3959
241. Закон о изменама и допунама Закона о Устав- ном суду (2011)	3703	940	265 74	824 11 3936
242. Закон о изменама и допунама Закона о Репу- бличким административним таксама (2005)	3689	1014	685 121	1346 10 4196
243. Закон о извозу и увозу робе двоструке намене (2013)	3685	1077	212 34	680 10 3893
244. Закон о посредовању у решавању спорова (2014)	3660	1149	132 54	526 10 3795
245. Закон о поступку регистрације у Агенцији за привредне регистре (2011)	3623	921	217 72	617 10 3826
246. Закон о платама државних службеника и на- мештеника (2006)	3617	849	382 105	583 12 3908

*наставак на следећој страни*



Табела Г.1 – наставак са претходне стране

редни број, назив и година доношења закона	речи	бројева	знакова	разм.
247. Закон о спољнотрговинском пословању (2008)	3577	1139	206 60	602 11 3765
248. Закон о дијаспори и Србима у региону (2009)	3548	1095	135 47	641 11 3706
249. Закон о цевоводном транспорту гасовитих и течних угљоводоника и дистрибуцији гасовитих угљоводоника (2008)	3517	940	240 51	613 10 3718
250. Закон о изменама и допунама Закона о државним службеницима (2009)	3512	896	319 107	833 12 3790
251. Закон о обављању саветодавних и стручних послова у области пољопривреде (2009)	3484	989	206 49	554 11 3674
252. Закон о изменама и допунама Закона о полицији (2011)	3443	1100	313 90	819 14 3749
253. Закон о изменама и допунама Закона о основама система образовања и васпитања (2013)	3431	1038	507 110	1276 12 3865
254. Закон о главном граду (2007)	3416	1007	124 57	528 10 3538
255. Закон о иновационој делатности (2005)	3405	1064	167 57	665 10 3528
256. Закон о јавном здрављу (2009)	3398	1176	169 32	602 10 3560
257. Закон о црквама и верским заједницама (2006)	3376	1111	92 60	460 12 3454
258. Закон о изменама и допунама Закона о пензијском и инвалидском осигурању (2005)	3364	659	572 141	1270 13 3928
259. Закон о изменама и допунама Кривичног законика (2012)	3363	942	340 100	958 11 3681
260. Закон о политичким странкама (2008)	3357	797	213 63	564 9 3535
261. Закон о кинематографији (2011)	3344	1174	240 66	743 14 3578
262. Закон о Заштитнику грађана (2005)	3336	1056	92 45	429 12 3428
263. Закон о изменама и допунама Закона о пореском поступку и пореској администрацији (2005)	3332	883	315 82	884 10 3630
264. Закон о изменама и допунама Закона о високом образовању (2010)	3305	749	395 90	1089 15 3665
265. Закон о легализацији објеката (2013)	3290	905	188 52	546 11 3466
266. Закон о праву на бесплатне акције и новчану накнаду коју грађани остварују у поступку приватизације (2007)	3266	890	203 59	492 12 3449
267. Закон о изменама и допунама Закона о пореском поступку и пореској администрацији (2011)	3255	712	469 112	1334 13 3633
268. Закон о Централном регистру обавезног социјалног осигурања (2010)	3221	903	236 65	557 11 3401

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

<b>редни број, назив и година доношења закона</b>	<b>речи</b>	<b>бројева</b>	<b>знакова</b>	<b>разм.</b>
269. Закон о изменама и допунама Закона о Војсци Србије (2009)	3209	888	392 96	961 11 3592
270. Закон о слободним зонама (2006)	3194	954	165 49	509 10 3380
271. Закон о заштити топографија полупроводничких производа (2013)	3135	904	173 46	455 10 3290
272. Закон о изменама и допунама Закона о акцизама (2012)	3122	734	754 95	1033 14 3701
273. Закон о изменама и допунама Закона о буџетском систему (2013)	3118	876	264 47	771 11 3372
274. Закон о органској производњи и органским производима (2005)	3116	968	232 46	612 11 3320
275. Закон о изменама и допунама Закона о ауторском и сродним правима (2011)	3113	933	327 86	897 11 3408
276. Закон о изменама и допунама Закона о регистрацији привредних субјеката (2005)	3073	625	447 91	1160 9 3445
277. Закон о младима (2011)	3027	894	107 32	440 9 3143
278. Закон о изменама и допунама Закона о порезима на употребу, држање и ношење добара (2009)	2984	570	651 95	1256 11 3455
279. Закон о јединственом бирачком списку (2009)	2972	745	94 46	320 10 3043
280. Закон о посредовању у промету и закупу непокретности (2013)	2961	856	185 41	498 9 3133
281. Закон о спречавању злостављања на раду (2009)	2954	804	151 46	439 10 3099
282. Закон о оптичким дисковима (2011)	2920	829	163 33	557 10 3053
283. Закон о изменама и допунама Закона о иновационој делатности (2010)	2889	882	230 52	755 12 3102
284. Закон о изгледу и употреби грба, заставе и химне Републике Србије (2009)	2838	909	145 56	563 13 2982
285. Закон о изменама и допунама Закона о извршењу кривичних санкција (2010)	2830	865	183 65	586 11 2984
286. Закон о председнику Републике (2007)	2820	799	81 56	309 8 2899
287. Закон о рехабилитацији (2011)	2815	802	223 58	539 11 3018
288. Закон о волонтирању (2009)	2803	876	188 38	526 12 2984
289. Закон о факторингу (2013)	2801	845	173 46	489 10 2971
290. Закон о изменама и допунама Закона о рударству (2005)	2773	873	263 69	775 13 3014

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

редни број, назив и година доношења закона	речи	бројева	знакова	разм.
291. Закон о изменама и допунама Закона о спре- чавању прања новца и финансирању тероризма (2010)	2758	682	338 56	760 11 3045
292. Закон о изменама и допунама Закона о поре- зима на имовину (2010)	2707	651	384 59	1052 14 3040
293. Закон о изменама и допунама Закона о уре- ђењу судова (2013)	2701	916	201 61	588 12 2887
294. Закон о изменама и допунама Закона о пољо- привредном земљишту (2009)	2686	662	248 55	651 12 2914
295. Закон о заштити од нејонизујућих зрачења (2008)	2668	716	185 36	447 10 2840
296. Закон о изменама и допунама Закона о др- жавном премеру и катастру (2013)	2642	672	257 83	749 12 2831
297. Закон о Српској академији наука и уметности (2010)	2629	902	116 53	450 10 2742
298. Закон о заштити топографија интегрисаних кола (2009)	2616	824	160 37	455 10 2771
299. Закон о изменама и допунама Закона о пла- нирању и и зградњи (2005)	2603	786	300 93	847 13 2874
300. Закон о изменама и допунама Закона о поре- ском поступку и пореској администрацији (2010)	2587	583	586 108	1275 12 2998
301. Закон о изменама и допунама Закона о по- резима на употребу, држање и ношење добара (2010)	2561	584	606 71	1198 10 2995
302. Закон о изменама и допунама Закона о поре- зима на имовину (2007)	2556	590	354 57	905 13 2857
303. Закон о изменама и допунама Закона о играма на срећу (2010)	2542	646	385 91	1016 11 2905
304. Закон о Влади (2005)	2528	794	65 53	308 10 2584
305. Закон о Попису становништва, домаћинства и станова 2011. године (2009)	2516	847	159 60	451 9 2647
306. Закон о изменама и допунама Закона о мини- старствима (2013)	2485	934	72 23	445 11 2544
307. Закон о изменама и допунама Закона о праву на бесплатне акције и новчану накнаду коју гра- ђани остварују у поступку приватизације (2010)	2481	787	220 41	601 12 2667
308. Закон о изменама и допунама Закона о при- ватизацији (2007)	2463	817	252 59	630 10 2691

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

<b>редни број, назив и година доношења закона</b>	<b>речи</b>		<b>бројева</b>		<b>знакова</b>		<b>разм.</b>
309. Закон о објављивању закона и других прописа и аката (2013)	2456	667	72	46	379	11	2498
310. Закон о изменама и допунама Закона о основама система образовања и васпитања (2011)	2407	822	227	65	687	13	2606
311. Закон о изменама и допунама Закона о порезу на добит правних лица (2013)	2389	710	241	52	633	11	2609
312. Закон о изменама и допунама Закона о финансијском лизингу (2005)	2373	725	191	28	401	8	2512
313. Закон о пребивалишту и боравишту грађана (2011)	2349	652	135	49	386	11	2472
314. Закон о Фонду за заштиту животне средине (2009)	2315	800	146	44	445	11	2456
315. Закон о контроли државне помоћи (2009)	2291	718	98	29	358	9	2386
316. Закон о враћању (реституцији) имовине црквама и верским заједницама (2006)	2290	871	68	42	329	8	2363
317. Закон о изменама и допунама Закона о Народној банци Србије (2012)	2278	659	194	47	549	12	2458
318. Закон о стечају и ликвидацији банака и друштава за осигурање (2005)	2269	573	204	54	470	9	2423
319. Закон о изменама и допунама Закона о порезу на додатну вредност (2007)	2210	578	375	64	820	11	2539
320. Закон о изменама и допунама Закона о рударству (2009)	2209	675	316	74	705	12	2505
321. Закон о управљању миграцијама (2012)	2191	801	66	32	339	12	2250
322. Закон о стандардизацији (2008)	2187	746	129	38	457	10	2315
323. Закон о попису пољопривреде 2011. године (2009)	2186	835	130	53	379	10	2293
324. Закон о етанолу (2008)	2181	663	210	51	511	12	2374
325. Закон о личној карти (2006)	2170	689	161	55	407	11	2321
326. Закон о електронској трговини (2008)	2141	805	114	26	362	10	2255
327. Закон о изменама и допунама Закона о процени утицаја на животну средину (2009)	2137	528	236	42	649	9	2355
328. Закон о изменама и допунама Закона о експропријацији (2008)	2129	561	161	47	591	11	2285
329. Закон о Изменама и допунама Закона о спречавању насиља и недоличног понашања на спортиским приредбама (2009)	2114	634	298	41	631	10	2364

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

редни број, назив и година доношења закона	речи	бројева	знакова	разм.
330. Закон о изменама и допунама Закона о Агенцији за приватизацију (2010)	2081	734	163 33	381 10 2202
331. Закон о изменама и допунама Закона о здравственом осигурању (2012)	2058	562	158 40	407 12 2185
332. Закон о закључивању и извршавању међународних уговора (2013)	2030	569	60 34	250 12 2095
333. Закон о изменама и допунама Закона о одбрани (2009)	2022	619	220 62	699 11 2205
334. Закон о Црвеном крсту Србије (2005)	2018	737	130 40	381 10 2131
335. Закон о Изменама и допунама Закона о парничном поступку (2009)	1987	576	288 125	873 9 2260
336. Закон о осигурању депозита (2005)	1968	653	121 40	332 10 2081
337. Закон о изменама и допунама Закона о ученичком и студентском стандарду (2013)	1962	723	86 26	410 12 2020
338. Закон о изменама и допунама Закона о жиговима (2012)	1957	547	280 60	833 12 2158
339. Закон о изменама и допунама Закона о спољнотрговинском пословању (2011)	1927	653	246 42	611 10 2129
340. Закон о изменама и допунама Закона о инвестиционим фондовима (2010)	1911	576	233 59	497 10 2118
341. Закон о електронском документу (2009)	1898	700	65 29	285 10 1973
342. Закон о акредитацији (2010)	1897	595	97 32	324 10 1972
343. Закон о изменама и допунама Закона о организацији и надлежности државних органа у сузбијању организованог криминала (2009)	1881	557	240 64	799 9 2096
344. Закон о роковима измирења новчаних обавеза у комерцијалним трансакцијама (2012)	1873	529	142 32	301 11 2006
345. Закон о заштити пословне тајне (2011)	1860	701	88 25	274 8 1941
346. Закон о обавезном примерку публикација (2011)	1846	696	103 30	413 10 1939
347. Закон о одликовањима Републике Србије (2008)	1838	636	92 43	317 10 1920
348. Закон о печату државних органа (2007)	1831	594	96 46	326 12 1918
349. Закон о изменама и допунама Закона о Агенцији за борбу против корупције (2010)	1818	487	231 56	705 11 2030
350. Закон о изменама и допунама Закона о дувану (2013)	1813	361	395 58	821 12 2097

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

<b>редни број, назив и година доношења закона</b>	<b>речи</b>	<b>бројева</b>		<b>знакова</b>		<b>разм.</b>
351. Закон о изменама и допунама закона о доприносима за обавезно социјално осигурање (2006)	1806	481	178	48	455	11 1960
352. Закон о изменама и допунама Закона о ваздушном саобраћају (2011)	1798	685	125	60	449	10 1913
353. Закон о изменама и допунама Закона о поштанским услугама (2008)	1780	541	243	64	678	11 1991
354. Закон о изменама и допунама Закона о Војнобезбедносној агенцији и Војнообавештајној агенцији (2013)	1771	564	181	33	449	11 1945
355. Закон о изменама и допунама Закона о финансирању локалне самоуправе (2012)	1767	536	243	44	493	12 1981
356. Закон о изменама и допунама Закона о прекршајима (2009)	1758	572	282	71	611	11 1956
357. Закон о судским вештацима (2010)	1752	607	70	34	274	10 1822
358. Закон о Агенцији за осигурање депозита (2005)	1748	603	114	39	338	9 1861
359. Закон о изменама и допунама Закона о здравственој заштити (2012)	1736	572	88	29	304	11 1816
360. Закон о основама уређења служби безбедности Републике Србије (2007)	1733	647	32	27	325	11 1814
361. Закон о изменама и допунама Закона о осигурању (2005)	1724	548	271	82	636	10 1966
362. Закон о социјалном становању (2008)	1698	596	93	22	295	9 1792
363. Закон о изменама и допунама Закона о заштити конкуренције (2013)	1685	605	162	50	449	12 1846
364. Закон о потврђивању Споразума између Владе Републике Србије и Владе Државе Израел о трговини и економској сарадњи (2009)	1678	741	61	24	250	9 1741
365. Закон о изменама и допунама Закона о платном промету (2011)	1671	492	158	45	383	10 1811
366. Закон о изменама и допунама Закона о ауторском и сродним правима (2012)	1671	595	149	36	353	12 1799
367. Закон о изменама и допунама Закона о акцизама (2012)	1664	405	574	77	678	13 2147
368. Закон о изменама и допунама Закона о Агенцији за лицензирање стечајних управника (2009)	1661	556	121	19	294	9 1752
369. Закон о изменама и допунама Закона о акцизама (2013)	1644	373	243	55	549	10 1847

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

редни број, назив и година доношења закона	речи	бројева	знакова	разм.
370. Закон о изменама и допунама Закона о акцизама (2007)	1615	358	325 64	740 12 1881
371. Закон о изменама и допунама Закона о јавним предузећима и обављању делатности од општег интереса (2005)	1613	511	123 31	399 9 1729
372. Закон о изменама и допунама Закона о доприносима за обавезно здравствено осигурање (2005)	1608	485	182 49	392 9 1768
373. Закон о Фонду за развој Републике Србије (2009)	1584	637	88 39	287 11 1668
374. Закон о изменама и допунама Закона о организацији и надлежности државних органа у поступку за ратне злочине (2009)	1583	480	123 35	420 9 1686
375. Закон о посебним условима за упис права својине на објектима изграђеним без грађевинске дозволе (2013)	1563	541	74 29	224 10 1619
376. Закон о изменама и допунама Закона о акцизама (2008)	1560	431	321 66	683 12 1785
377. Закон о одређивању максималног броја запослених у локалној администрацији (2009)	1550	441	112 23	236 9 1654
378. Закон о споразумном финансијском реструктурирању привредних друштава (2011)	1548	624	51 18	223 9 1595
379. Закон о изменама и допунама Закона о судским таксама (2009)	1535	244	655 99	1583 9 2084
380. Закон о изменама и допунама Закона о трговини (2012)	1527	639	129 55	437 11 1649
381. Закон о изменама и допунама Закона буџетском систему (2011)	1523	480	181 47	511 13 1719
382. Закон о Агенцији за осигурање и финансирање извоза Републике Србије (2005)	1521	633	92 29	278 8 1614
383. Закон о посебним мерама за спречавање вршења кривичних дела против полне слободе према малолетним лицима (2013)	1444	506	128 37	259 8 1564
384. Закон о изменама и допунама Закона о репродуктивном материјалу шумског дрвећа (2008)	1416	394	209 39	437 11 1573
385. Закон о изменама и допунама Закона о ванредним ситуацијама (2011)	1409	442	180 45	513 10 1566
386. Закон о изменама и допунама Закона о судијама (2013)	1407	477	181 65	474 11 1584

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

редни број, назив и година доношења закона	речи	бројева	знакова	разм.
387. Закон о избору председника Републике (2007)	1389	506	61 39	169 8 1444
388. Закон о изменама и допунама Закона о осигурању депозита (2010)	1341	441	137 30	421 11 1468
389. Закон о изменама и допунама Закона о заштити ваздуха (2012)	1341	381	89 41	284 10 1405
390. Закон о изменама и допунама Закона о управљању отпадом (2010)	1333	474	171 46	427 10 1495
391. Закон о изменама и допунама Закона о порезима на имовину (2008)	1324	372	228 45	656 12 1523
392. Закон о уплати доприноса за пензијско и инвалидско осигурање за поједине категорије осигураника-запослених (2005)	1315	412	95 24	241 10 1414
393. Закон о изменама и допунама закона о финансијској подршци породици са децом (2005)	1311	459	102 28	299 10 1418
394. Закон о преузимању имовине и обавеза одређених банака ради очувања стабилности финансијског система Републике Србије (2012)	1307	445	129 45	306 10 1399
395. Закон о Развојном фонду Аутономне Покрајине Војводине (2012)	1296	565	65 26	238 11 1353
396. Закон о изменама и допунама Закона о судским таксама (2005)	1281	238	556 84	1336 10 1746
397. Закон о изменама и допунама Закона о порезу на доходак грађана (2009)	1251	398	199 55	431 14 1431
398. Закон о изменама и допунама Закона о порезу на добит правних лица (2011)	1241	436	149 42	343 11 1371
399. Закон о условном отпису камата и мировању пореског дуга (2012)	1239	371	92 21	252 9 1350
400. Закон о изменама и допунама Закона о Агенцији за осигурање депозита (2010)	1238	446	99 20	246 10 1326
401. Закон о утврђивању јавног интереса и посебним поступцима експропријације и прибављања документације ради реализације изградње система за транспорт природног гаса "Јужни ток" (2013)	1236	464	74 35	260 9 1283
402. Закон о изменама и допунама Закона о јавном информисању (2009)	1227	419	164 30	289 10 1326
403. Закон о изменама и допунама Закона о пореском поступку и пореској администрацији (2013)	1225	402	141 47	321 12 1309

*наставак на следећој страни*



Табела Г.1 – наставак са претходне стране

редни број, назив и година доношења закона	речи	бројева	знакова	разм.
404. Закон о изменама и допунама Закона о регионалном развоју (2010)	1225	391	159 47	553 11 1393
405. Закон о мировању и отпису дуга по основу доприноса за обавезно здравствено осигурање (2008)	1208	358	103 24	241 10 1319
406. Закон о изменама и допунама закона о социјалној заштити и обезбеђивању социјалне сигурности грађана (2005)	1205	407	125 55	300 9 1313
407. Закон о одређивању максималног броја запослених у републичкој администрацији (2009)	1184	457	52 27	168 12 1230
408. Закон о обележавању дана жалости на територији Републике Србије (2005)	1166	468	91 24	226 8 1242
409. Закон о изменама и допунама Закона о заложном праву на покретним стварима уписаним у регистар (2005)	1155	450	83 24	221 8 1230
410. Закон о Изменама и допунама Закона о агенцији за привредне регистре (2009)	1155	482	112 19	288 10 1243
411. Закон о изменама и допунама Закона о енергетици (2012)	1140	420	77 26	225 11 1212
412. Закон о изменама и допунама Закона о парничном поступку (2014)	1113	423	161 57	436 11 1278
413. Закон о изменама и допунама Закона о слободном приступу информацијама од јавног значаја (2009)	1113	414	135 40	403 10 1250
414. Закон о изменама и допунама Закона о акцизама (2011)	1109	248	155 47	352 11 1221
415. Закон о изменама и допунама Закона о буџетском систему (2013)	1102	342	149 41	297 12 1211
416. Закон о изменама и допунама Закона о радиодифузији (2006)	1076	437	103 41	325 11 1165
417. Закон о изменама и допунама Закона о порезу на доходак грађана (2009)	1073	468	142 51	385 12 1200
418. Закон о изменама и допунама Закона о доприносима за обавезно социјално осигурање (2013)	1069	322	179 50	366 11 1231
419. Закон о изменама и допунама закона о Републичким административним таксама (2009)	1063	295	202 85	419 12 1224
420. Закон о изменама и допунама Закона о привредним друштвима (2011)	1056	341	170 58	445 11 1230

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

редни број, назив и година доношења закона	речи	бројева	знакова	разм.
421. Закон о умањењу нето прихода лица у јавном сектору (2013)	1055	372	91 21	166 10 1127
422. Закон о изменама и допунама Закона о финансијској подршци породици са децом (2009)	1053	373	100 27	251 10 1151
423. Закон о изменама и допунама Закона о јавном тужилаштву (2013)	1028	414	105 44	274 11 1127
424. Закон о допунама закона о ванпарничном поступку (2012)	1025	399	50 18	144 10 1059
425. Закон о изменама и допунама Закона о Агенцији за привредне регистре (2011)	1006	412	140 49	303 12 1109
426. Закон о обуци судија, јавних тужилаца, заменика јавних тужилаца и судијских и тужилачких помоћника (2006)	978	353	82 41	185 8 1042
427. Закон о амнестији (2012)	937	337	141 76	258 12 1047
428. Закон о изменама и допунама Закона о избору народних посланика (2011)	924	338	107 45	349 12 1021
429. Закон о изменама и допунама Закона о буџетском систему (2010)	915	256	136 49	412 12 1036
430. Закон о изменама и допунама Закона о дугану (2010)	908	259	114 38	389 9 1020
431. Закон о изменама и допунама Закона о инвестиционим фондовима (2009)	895	316	136 45	341 11 1022
432. Закон о изменама и допунама Закона о заштити права оплемењивача биљних сорти (2011)	887	309	123 40	340 10 1007
433. Закон о изменама и допунама Закона о стецају и ликвидацији банака и друштава (2008)	858	336	58 19	156 9 907
434. Закон о утврђивању максималне зараде у јавном сектору (2012)	842	340	53 20	123 11 892
435. Закон о изменама и допунама Закона о привредним коморама (2009)	838	369	70 25	230 11 900
436. Закон о јавном дугу Републике Србије по основу преузимања обавеза Републичког фонда за пензијско и инвалидско осигурање пољопривредника насталих по основу неисплаћених пензија и новчаних накнада (2005)	829	321	91 26	185 13 914
437. Закон о изменама и допунама Закона о осигурању (2007)	828	330	133 52	318 9 952

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

редни број, назив и година доношења закона	речи	бројева	знакова	разм.
438. Закон о организацији и надлежности државних органа за борбу против високотехнолошког криминала (2005)	824	357	32 18	125 7 857
439. Закон о изменама и допунама Закона о дувану (2007)	824	294	197 54	350 9 1009
440. Закон о привременом смањењу плата, односно зарада, нето накнада и других примања у државној администрацији и јавном сектору (2009)	823	370	42 17	138 10 864
441. Закон о изменама и допунама Закона о порезу на доходак грађана (2013)	813	304	123 47	260 12 912
442. Закон о изменама и допунама Закона о биоцидним производима (2011)	801	326	86 30	231 12 880
443. Закон о изменама и допунама Закона о личној карти (2011)	796	350	74 27	205 11 867
444. Закон о изменама и допунама Закона о безбедности саобраћаја на путевима (2011)	788	342	101 39	260 12 878
445. Закон о изменама и допунама Закона о здравственој заштити (2010)	787	305	51 26	158 10 832
446. Закон о изменама и допунама Закона о Народној банци Србије (2012)	786	238	89 36	241 11 868
447. Закон о задуживању Републике Србије код КВС BANKE A.D. BEOGRAD, NOVE KREDITNE BANKE MARIBOR...које су дале заједничку понуду (2010)	780	217	67 22	231 11 799
448. Закон о изменама и допунама Закона о спречавању насиља и недоличног понашања на спорским приредбама (2007)	780	350	96 29	280 8 861
449. Закон о изменама и допунама Закона о безбедности саобраћаја на путевима (2010)	775	333	109 43	243 10 875
450. Закон о изменама и допунама Закона о мирном решавању радних спорова (2009)	763	297	93 30	282 12 855
451. Закон о јавном дугу Републике Србије по основу преузимања обавеза Републичког фонда за пензијско и инвалидско осигурање запослених насталих по основу неисплаћених пензија и новчаних накнада (2005)	762	325	71 26	159 13 828
452. Закон о изменама и допунама Закона о порезу на доходак грађана (2012)	759	334	100 43	228 11 847
453. Закон о Класификацији делатности (2008)	758	294	37 21	133 10 794

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

<b>редни број, назив и година доношења закона</b>	<b>речи</b>		<b>бројева</b>		<b>знакова</b>		<b>разм.</b>
454. Закон о изменама и допунама Закона о биоцидним производима (2010)	742	232	107	37	303	9	851
455. Закон о изменама и допунама Царинског закона (2005)	721	331	103	32	228	10	809
456. Закон о изменама и допунама Закона о прекршајима (2008)	709	189	104	66	353	8	813
457. Закон о изменама и допунама Закона о превозу у друмском саобраћају (2005)	709	238	141	25	311	10	808
458. Закон о подстицању грађевинске индустрије Републике Србије у условима економске кризе (2010)	693	331	39	15	122	11	730
459. Закон о остваривању права на здравствену заштиту деце, трудница и породиља (2013)	685	257	52	18	111	8	726
460. Закон о изменама и допунама закона о уређењу судова (2005)	684	228	148	65	289	8	779
461. Закон о изменама и допунама Закона о стратешкој процени утицаја на животну средину (2010)	679	245	82	22	240	10	757
462. Закон о изменама и допунама Закона о доприносима за обавезно социјално осигурање (2011)	669	223	94	32	234	10	752
463. Закон о изменама и допунама Закона о доприносима за обавезно социјално осигурање (2014)	649	269	96	37	195	12	725
464. Закон о изменама и допунама Закона о локалним изборима (2011)	646	294	65	31	190	11	717
465. Закон о изменама и допунама Закона о јавном тужилаштву (2010)	642	228	54	22	130	8	693
466. Закон о изменама и допунама Закона о енергетици (2012)	638	244	125	45	271	10	759
467. Закон о изменама и допунама Закона о Влади (2007)	628	256	62	30	205	8	691
468. Закон о изменама Закона о општем управном поступку (2009)	626	165	113	60	435	9	738
469. Закон о изменама и допунама Закона о финансирању локалне самоуправе (2011)	621	252	69	27	176	13	691
470. Закон о изменама и допунама Кривичног законика (2009)	613	305	50	20	132	10	657
471. Закон о изменама и допунама Закона о царинској тарифи (2007)	611	283	49	22	160	11	656

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

редни број, назив и година доношења закона	речи	бројева	знакова	разм.
472. Закон о изменама и допунама Закона о судијама (2010)	607	226	57 21	141 9 661
473. Закон о изменама и допунама Закона о заштити од буке у животној средини (2010)	605	221	121 34	286 9 719
474. Закон о отпису камате на доспеле обавезе по основу одређених пореза и доприноса за обавезно социјално осигурање (2008)	599	229	65 18	136 10 671
475. Закон о забрани манифестација неонацистичких или фашистичких организација и удружења и забрани употребе неонацистичких или фашистичких симбола и обележја (2008)	596	253	62 14	129 7 633
476. Закон о изменама и допунама Закона о хемикалијама (2011)	593	305	55 19	156 10 639
477. Закон о преузимању обавеза здравствених установа према веледрогеријама по основу набавке лекова и медицинског материјала и претварању тих обавеза у јавни дуг Републике Србије (2012)	588	226	104 29	160 11 695
478. Закон о изменама и допунама Закона о професионалној рехабилитацији и запошљавању особа са инвалидитетом (2013)	588	209	40 28	113 9 628
479. Закон о расподели трансферних средстава из буџета Републике Србије и учешћу општина, градова и града Београда у порезу на зараде у 2006. години (2005)	583	365	667 403	576 9 941
480. Закон о забрани усавршавања, производње и стварања залиха бактериолошког (биолошког) и токсичког оружја и о њиховом уништавању (2010)	577	296	54 19	179 11 623
481. Закон о изменама и допунама Закона о хемикалијама (2010)	567	226	77 32	269 9 644
482. Закон о одговорности произвођача ствари са недостатком (2005)	566	270	13 13	75 9 585
483. Закон о затезној камати (2012)	554	242	32 17	109 10 585
484. Закон о изменама и допунама Закона о осигурању депозита (2008)	551	255	61 21	183 12 612
485. Закон о изменама и допунама Закона о порезу на доходак грађана (2014)	549	248	97 38	162 12 601

наставак на следећој страни

Табела Г.1 – наставак са претходне стране

<b>редни број, назив и година доношења закона</b>	<b>речи</b>	<b>бројева</b>	<b>знакова</b>	<b>разм.</b>
486. Закон о изменама и допунама Закона о платама у државним органима и јавним службама (2010)	541	239	62 15	149 12 594
487. Закон о изменама и допунама Закона о порезу на доходак грађана (2011)	540	202	75 38	174 10 600
488. Закон о регулисању односа Републике Србије и зајмопримаца или гараната за преузете обавезе по зајмовима међународне банке за обнову и развој (2008)	528	217	99 53	169 10 570
489. Закон о изменама и допунама Закона о држављанству Републике Србије (2007)	516	186	92 37	214 7 607
490. Закон о изменама и допунама Закона о фискалним касама (2012)	515	236	118 57	252 10 618
491. Закон о изменама и допунама Закона о службеној употреби језика и писама (2008)	513	242	35 21	97 11 538
492. Закон о изменама и допунама Закона о војној, радној и материјалној обавези (2010)	508	205	73 21	193 9 581
493. Закон о изменама и допунама Закона о Фонду за развој Републике Србије (2012)	505	243	53 20	139 10 554
494. Закон о изменама Закона о хемикалијама (2012)	498	162	245 73	367 9 738
495. Закон о изменама и допунама Закона о доприносима за обавезно социјално осигурање (2013)	491	194	76 38	174 8 560
496. Закон о допунама Закона о коморама здравствених радника (2010)	485	175	28 14	69 9 507
497. Закон о амнестији (2005)	480	209	75 49	127 8 534
498. Закон о изменама и допунама Закона о електронској трговини (2013)	477	250	72 26	199 9 539
499. Закон о изменама и допунама Закона о доприносима за обавезно социјално осигурање (2008)	474	196	49 27	122 12 509
500. Закон о допунама Закона о Царинској тарифи (2008)	468	222	170 44	226 10 661
501. Закон о изменама и допунама Закона о државним службеницима (2008)	462	170	60 33	194 12 520
502. Закон о амнестији (2009)	462	205	41 25	86 10 501
503. Закон о изменама и допунама Закона о судским таксама (2012)	446	161	80 41	189 10 514

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

редни број, назив и година доношења закона	речи	бројева	знакова	разм.
504. Закон о изменама и допунама закона о јавном дугу Р.Србије по основу преузимања обавеза републичког фонда за ПИО запослених насталих по основу неисплаћених пензија и новчаних накнада (2005)	445	148	74 18	133 12 520
505. Закон о изменама и допунама Закона о организацији и надлежности државних органа за борбу против високотехнолошког криминала (2009)	442	211	52 18	232 9 492
506. Закон о изменама и допунама Царинског закона (2012)	442	198	72 31	219 11 517
507. Закон о изменама и допунама Закона о иновационој делатности (2013)	439	184	57 28	147 8 494
508. Закон о изменама и допунама Закона о Фонду за развој Републике Србије (2005)	439	232	35 16	144 8 474
509. Закон о изменама и допунама Закона о Фонду за развој Републике Србије (2010)	437	216	58 16	168 10 495
510. Закон о изменама и допунама Закона о обавезном осигурању у саобраћају (2011)	433	189	86 29	213 10 516
511. Закон о изменама и допунама Закона о Фонду за заштиту животне средине (2011)	433	211	65 19	196 12 490
512. Закон о изменама и допунама Законика о кривичном поступку (2013)	431	225	49 23	133 8 475
513. Закон о задуживању Републике Србије код ERSTE GCIB FINANCE I.B.V. Amsterdam, са ERSTE BANK а.д. Нови Сад као агентом, MARFIN BANK а.д. Београд и Војвођанске банке а.д. Нови Сад (2009)	424	170	38 18	117 9 440
514. Закон о изменама и допунама Закона о високом образовању (2013)	422	185	43 21	132 10 458
515. Закон о изменама и допунама закона о уређењу судова (2006)	422	218	45 22	154 8 462
516. Закон о изменама и допунама Закона о извршењу казне затвора за кривична дела организованог криминала (2010)	418	215	50 13	135 9 466
517. Закон о Српској енциклопедији (2005)	417	226	26 14	84 7 443
518. Закон о изменама и допунама Закона о јединственом бирачком списку (2011)	411	176	42 18	108 10 450
519. Закон о изменама и допунама закона о радиодифузији (2005)	405	187	59 31	156 8 451

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

<b>редни број, назив и година доношења закона</b>	<b>речи</b>	<b>бројева</b>	<b>знакова</b>	<b>разм.</b>
520. Закон о изменама и допунама Закона о државној управи (2007)	400	205	25 15	110 7 423
521. Закон о изменама и допунама Закона о Високом савету судства (2010)	399	170	50 22	163 10 450
522. Закон о изменама и допунама Закона о порезима на имовину (2011)	399	160	101 33	189 12 475
523. Закон о допунама Закона о превозу у друмском саобраћају (2006)	396	165	67 23	128 11 453
524. Закон о изменама и допунама Закона о Државном већу тужилаца (2010)	389	155	40 18	140 8 430
525. Закон о рехабилитацији (2005)	373	217	14 12	47 3 388
526. Закон о изменама и допунама Закона о јавном дугу Р. Србије по основу преузимања обавеза Републичког фонда за ПИО пољопривредника насталих по основу неисплаћених пензија и новчаних накнада (2005)	365	146	69 19	118 11 434
527. Закон о изменама и допунама Законика о кривичном поступку (2011)	362	143	116 48	201 10 436
528. Закон о државном печату Републике Србије (2007)	359	197	10 9	57 9 370
529. Закон о изменама и допунама Закона о спољним пословима (2009)	351	212	39 15	102 12 385
530. Закон о изменама и допунама Закона о туризму (2012)	346	135	79 33	191 10 421
531. Закон о изменама и допунама Закона о објављивању закона и других прописа и општих аката и о издавању „Службеног гласника Републике Србије“ (2008)	340	154	36 15	126 11 371
532. Закон о измени Закона о приватизацији (2014)	339	166	50 22	90 10 375
533. Закон о задуживању Републике Србије код Комерцијалне банке а.д. Београд за потребе куповине зграде Аероинжењеринга и извођење неопходних радова на објекту за потребе смештаја Првог основног суда у Београду (2011)	329	181	27 14	57 10 354
534. Закон о изменама и допунама Закона о Државној ревизорској институцији (2010)	326	158	46 18	86 9 369
535. Закон о изменама и допунама Законика о кривичном поступку (2014)	326	187	46 22	136 9 368

*наставак на следећој страни*



Табела Г.1 – наставак са претходне стране

редни број, назив и година доношења закона	речи	бројева	знакова	разм.
536. Закон о задуживању Републике Србије код ALPHA BANK SRBIJA а.д. Нови Сад и PIRAEUS BANK а.д. Београд (2009)	320	160	30 13	76 9 341
537. Закон о изменама и допунама Закона о доприносима за обавезно социјално осигурање (2011)	319	147	50 29	99 8 355
538. Закон о измени и допуни Закона о државној управи (2010)	317	185	21 12	58 9 335
539. Закон о изменама и допунама Закона о Државном већу тужилаца (2011)	313	133	25 11	51 8 331
540. Закон о преузимању обавеза Јавног предузећа „Путеви Србије” према привредним друштвима по основу неизмирених обавеза и претварању тих обавеза у јавни дуг Републике Србије (2013)	310	157	32 11	68 11 344
541. Закон о регулисању обавеза Републике Србије према Међународном монетарном фонду по основу коришћења средстава алокације специјалних права вучења одобрених Резолуцијама Одбора гувернера... (2009)	310	176	36 22	67 8 338
542. Закон о изменама и допунама Закона о високом образовању (2012)	308	157	62 35	112 10 354
543. Закон о изменама и допунама Закона о платама државних службеника и намештеника (2010)	298	177	186 72	130 10 413
544. Закон о допунама Закона о промету непокретности (2009)	294	150	16 8	46 8 308
545. Закон о изменама и допунама Закона о радиодифузији (2009)	288	151	48 28	112 9 331
546. Закон о изменама и допунама Закона о порезима на имовину (2011)	287	151	62 32	132 9 336
547. Закон о изменама Закона о организацији и надлежности државних органа у сузбијању организованог криминала, корупције и других посебно тешких кривичних дела (2013)	278	154	60 33	136 9 323
548. Закон о изменама и допунама Закона о порезу на добит правних лица (2013)	278	172	52 25	99 10 306
549. Закон о изменама и допунама Кривичног законика (2013)	277	137	42 20	112 9 316

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

редни број, назив и година доношења закона	речи	бројева	знакова	разм.
550. Закон о изменама и допунама Закона о банкама и другим финансијским организацијама (2005)	275	146	56 39	114 9 322
551. Закон о изменама и допунама Закона о судијама (2005)	274	147	62 33	122 8 329
552. Закон о изменама и допунама Закона о државном премеру и катастру (2010)	274	113	69 35	202 10 340
553. Закон о изменама и допунама Закона о јавним путевима (2013)	273	137	51 22	127 9 323
554. Закон о изменама и допунама Закона о дувану (2012)	272	149	31 20	61 9 296
555. Закон о изменама Закона о извршењу и обезбеђењу (2014)	271	157	19 13	65 9 290
556. Закон о регулисању односа Републике Србије и дужника – корисника јавних средстава, за прузете обавезе по иностраним кредитима, односно зајмовима (2011)	267	151	54 43	93 8 299
557. Закон о изменама и допунама Закона о ванредним ситуацијама (2012)	265	147	64 26	133 9 328
558. Закон о изменама и допунама Закона о здравственој заштити (2010)	265	131	33 17	84 9 295
559. Закон о изменама и допунама Закона о Влади (2011)	262	147	30 18	67 9 288
560. Закон о изменама и допунама Закона о Високом савету судства (2011)	261	125	25 11	49 8 279
561. Закон о изменама Закона о забрани развоја, производње, складиштења и употребе хемијског оружја и о његовом уништавању (2013)	261	152	40 12	120 13 300
562. Закон о изменама и допунама Закона о јавном дугу (2011)	258	156	28 13	77 10 282
563. Закон о задужењу Републике Србије код ЕУ-ROBANK EFG A.D. BEOGRAD (2009)	257	144	23 11	50 9 277
564. Закон о изменама Закона о порезу на доходак грађана (2009)	252	167	25 18	93 11 273
565. Закон о изменама Закона о водама (2012)	251	94	83 27	176 9 332
566. Закон о изменама и допунама Закона о Заштитнику грађана (2007)	250	119	42 17	119 7 291

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

редни број, назив и година доношења закона	речи	бројева	знакова	разм.
567. Закон о утврђивању јавног интереса за експропријацију непокретности ради изградње магистралног гасовода МГ-11 (2009)	250	154	32 18	79 8 264
568. Закон о измени и допуни Закона о високом образовању (2008)	246	127	44 20	81 11 282
569. Закон о изменама и допунама Закона о државним и другим празницима у Републици Србији (2007)	245	111	33 12	109 8 276
570. Закон о изменама и допунама Закона о организацији и надлежности државних органа у поступку за ратне злочине (2007)	240	158	32 15	68 9 267
571. Закон о изменама и допунама Закона о раду (2013)	239	147	23 12	46 8 258
572. Закон о задуживању Републике Србије код Societe Generale уз гаранцију Светске банке (2011)	237	135	28 13	52 9 262
573. Закон о изменама и допунама закона о судијама (2005)	235	127	65 41	99 7 274
574. Закон о изменама и допунама Закона о Агенцији за осигурање и финансирање извоза Републике Србије (2010)	233	135	22 10	93 9 262
575. Закон о изменама и допунама Закона о финансирању политичких странака (2008)	231	101	38 16	117 8 269
576. Закон о изменама и допунама Закона о висини стопе затезне камате (2011)	225	114	27 12	70 8 250
577. Закон о изменама и допунама Закона о спречавању прања новца и финансирања тероризма (2009)	224	95	55 19	95 9 269
578. Закон о измени Закона о платама државних службеника и намештеника (2013)	224	135	16 13	42 10 238
579. Закон о престанку важења закона о Полицијској академији (2006)	223	139	15 12	56 8 232
580. Закон о изменама и допунама Закона о државним празницима у Републици Србији (2011)	223	110	28 18	87 9 255
581. Закон о издавању доплатне поштанске марке (2005)	219	143	10 7	36 7 230
582. Закон о изменама и допунама Закона о јавном тужилаштву (2005)	216	131	37 24	93 8 250
583. Закон о изменама и допунама Закона о уџбеницима и другим наставним средствима (2006)	214	137	24 13	66 9 234

наставак на следећој страни

Табела Г.1 – наставак са претходне стране

<b>редни број, назив и година доношења закона</b>	<b>речи</b>	<b>бројева</b>	<b>знакова</b>	<b>разм.</b>
584. Закон о привременом изузимању од опорезивања порезом на доходак грађана одређених врста прихода (2008)	212	108	23 17	42 8 228
585. Закон о изменама и допунама Закона о Влади (2008)	208	128	22 12	52 9 228
586. Закон о изменама Закона о јавном бележничтву (2013)	202	134	28 20	64 8 229
587. Закон о престанку важења Закона о оснивању Новинско-издавачке установе “Братство” (2006)	199	95	7 7	67 7 193
588. Закон о изменама и допунама Закона о уређењу судова (2009)	193	116	23 13	105 9 217
589. Закон о измени и допуни Закона о пензијском и инвалидском осигурању (2008)	193	101	55 29	100 10 228
590. Закон о престанку важења Закона о оснивању Јавног предузећаза истраживање, производњу, прераду и промет нафте и природног гаса (2005)	192	109	36 12	68 9 219
591. Закон о измени Закона о пензијском и инвалидском осигурању (2013)	189	91	42 26	86 11 229
592. Закон о изменама и допунама Закона о одликовањима Републике Србије (2010)	180	103	25 10	85 9 212
593. Закон о Речнику Српске академије наука и уметности (2005)	179	122	8 7	38 6 188
594. Закон о измени Закона о пензијском и инвалидском осигурању (2013)	177	92	50 26	93 11 220
595. Закон о допуни Закона о јавном тужилаштву (2012)	173	95	29 19	60 9 195
596. Закон о допуни Закона о судијама (2012)	171	92	25 16	54 9 190
597. Закон о изменама и допунама Закона о преузимању акционарских друштава (2009)	159	103	28 13	56 9 185
598. Закон о изменама и допунама Закона о јавном дугу (2009)	157	109	18 10	41 9 175
599. Закон о допунама Закона о задругама (2006)	153	107	23 11	56 9 170
600. Закон о изменама и допунама законика о кривичном поступку (2005)	152	75	22 14	85 8 164
601. Закон о допунама Закона о слободном приступу информацијама од јавног значаја (2007)	150	99	16 12	59 7 166
602. Закон о изменама и допуни Закона о јавним путевима (2012)	147	84	36 20	97 8 184

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

<b>редни број, назив и година доношења закона</b>	<b>речи</b>	<b>бројева</b>		<b>знакова</b>		<b>разм.</b>
603. Закон о престанку важења Закона о Фонду за осигурање и финансирање спољнотрговинских послова (2008)	147	92	8	6	35	8 153
604. Закон о допунама Закона о заштити података о личности (2012)	139	81	24	11	49	12 156
605. Закон о изменама и допунама закона о телекомуникацијама (2006)	138	87	23	10	70	8 159
606. Закон о измени Закона о слободном приступу информацијама од јавног значаја (2010)	135	100	11	9	34	9 145
607. Закон о изменама Закона о вину (2012)	134	91	17	10	56	9 149
608. Закон о изменама Закона о становању (2011)	134	57	39	28	105	9 166
609. Закон о изменама Закона о попису пољопривреде 2011. године (2011)	132	55	52	18	107	8 185
610. Закон о изменама и допунама Закона о обележавању дана жалости на територији Републике Србије (2009)	131	93	22	7	50	9 148
611. Закон о изменама Закона о порезима на употребу, држање и ношење добара (2011)	131	75	46	26	74	8 162
612. Закон о изменама и допунама Законика о кривичном поступку (2005)	129	97	16	14	43	7 142
613. Закон о изменама и допунама Закона о Агенцији за осигурање депозита (2008)	124	91	14	9	49	8 138
614. Закон о престанку важења Закона о Фонду за заштиту животне средине (2012)	124	88	10	8	35	8 132
615. Закон о измени и допуни Закона о лековима и медицинским средствима (2012)	121	86	9	8	28	8 130
616. Закон о изменама и допунама Закона о јавном тужилаштву (2009)	121	89	25	12	60	8 146
617. Закон о изменама Закона о државним службеницима (2007)	119	76	28	13	65	8 147
618. Закон о изменама Закона о шумама (2012)	118	81	20	15	44	9 138
619. Закон о допунама Закона о одржавању стамбених зграда (2011)	114	62	28	16	52	9 137
620. Закон о допунама Закона о Влади (2012)	113	58	20	15	77	10 130
621. Закон о допуни закона о раду (2009)	113	85	12	8	39	9 124
622. Закон о изменама Закона о заштити од јонизујућих зрачења и о нуклеарној сигурности (2012)	108	48	64	17	102	6 173

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

редни број, назив и година доношења закона	речи	бројева	знакова	разм.
623. Закон о измени и допуни Закона о осигурању (2013)	108	82	33 25	65 9 129
624. Закон о престанку важења Закона о удруживању у Заједницу Југословенске електропривреде (2009)	107	73	6 6	38 7 113
625. Закон о изменама и допунама Закона о путним исправама (2009)	106	68	16 12	37 8 121
626. Закон о измени Закона о гаранцијском фонду (2005)	104	80	13 10	34 8 116
627. Закон о престанку важења Закона о организацији и пословању привредних јединица установа за извршење кривичних санкција (2009)	103	78	7 7	21 7 110
628. Закон о изменама Закона о планирању и изградњи (2012)	103	79	16 13	31 10 117
629. Закон о допуни Закона о платама државних службеника и намештеника (2007)	99	80	11 7	23 9 104
630. Закон о допунама Законика о кривичном поступку (2010)	95	64	24 20	50 9 113
631. Закон о измени и допунама Закона о ветеринарству (2012)	93	56	17 13	36 8 109
632. Закон о измени и допуни Закона о путним исправама (2010)	92	75	18 15	39 8 108
633. Закон о изменама Закона о порезу на додату вредност (2013)	92	71	22 17	48 10 108
634. Закон о изменама Законика о кривичном поступку (2007)	91	57	30 21	53 7 117
635. Закон о допуни Закона о платама у државним органима и јавним службама (2012)	91	71	15 12	40 9 105
636. Закон о измени Закона о судским таксама (2011)	90	63	26 24	44 10 106
637. Закон о изменама Закона о републичким административним таксама (2012)	88	52	66 51	88 9 138
638. Закон о допуни Закона о платама у државним органима и јавним службама (2013)	86	64	16 14	39 10 100
639. Закон о допуни Закона о судијама (2009)	86	72	16 11	34 8 101
640. Закон о изменама Закона о враћању одузете имовине и обештећењу (2013)	86	66	11 9	32 9 97

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

<b>редни број, назив и година доношења закона</b>	<b>речи</b>	<b>бројева</b>		<b>знакова</b>		<b>разм.</b>
641. Закон о повлачењу резерве дате на Протокол о забрани употребе у рату загушљивих, отровних и сличних гасова и бактериолошких средстава (2009)	85	62	5	5	19	7 89
642. Закон о допуни Закона о пензијском и инвалидском осигурању (2012)	85	62	30	20	56	12 110
643. Закон о измени Закона о преузимању обавеза здравствених установа према веледрогеријама по основу набавке лекова и медицинског материјала и претварању тих обавеза у јавни дуг Републике Србије (2013)	84	70	14	10	27	8 92
644. Закон о изменама Закона о безбедносно - информативној агенцији (2009)	84	49	12	9	44	9 96
645. Закон о изменама Закона о организацији и надлежности државних органа у поступку за ратне злочине (2005)	83	63	9	7	40	7 92
646. Закон о допунама Закона о уређењу судова (2011)	83	49	19	16	57	10 100
647. Закон о измени и допуни Закона о приватизацији (2012)	82	61	23	17	48	9 101
648. Закон о изменама Закона о јавним путевима (2011)	82	63	15	11	33	9 96
649. Закон о допуни Закона о платном промету (2006)	82	65	11	8	25	7 91
650. Закон о изменама Закона о обавезном осигурању у саобраћају (2012)	80	56	19	13	43	8 97
651. Закон о изменама Закона о Државној ревизорској институцији (2007)	79	63	14	8	41	7 93
652. Закон о изменама Закона о попису становништва, домаћинства и станова 2011. године (2011)	77	55	19	11	50	8 100
653. Закон о измени Закона о мировању и отпису дуга по основу доприноса за обавезно здравствено осигурање (2009)	76	64	9	8	24	8 85
654. Закон о измени Законика о кривичном поступку (2013)	73	61	12	10	24	10 82
655. Закон о изменама Закона о приватизацији (2012)	71	57	17	16	35	8 83
656. Закон о измени Закона о Влади (2014)	70	58	20	17	37	9 87

*наставак на следећој страни*

Табела Г.1 – наставак са претходне стране

<b>редни број, назив и година доношења закона</b>	<b>речи</b>		<b>бројева</b>		<b>знакова</b>		<b>разм.</b>
657. Закон о изменама Закона о организацији и надлежности државних органа у сузбијању организованог криминала (2005)	70	58	16	12	41	7	81
658. Закон о изменама Закона о дувану (2011)	68	48	20	15	37	8	86
659. Закон о допуни Закона о пензијском и инвалидском осигурању (2009)	67	47	22	17	49	9	78
660. Закон о измени Закона о јавним путевима (2007)	66	49	21	11	33	8	85
661. Закон о измени Закона о подстицању грађевинске индустрије Републике Србије у условима економске кризе (2012)	60	50	9	9	24	8	68
662. Закон о изменама Законика о кривичном поступку (2012)	60	43	18	10	40	8	77
663. Закон о измени Закона о подстицању грађевинске индустрије Републике Србије у условима економске кризе (2011)	59	49	7	7	20	8	66
664. Закон о изменама и допунама Закона о запошљавању и осигурању за случај незапослености (2010)	54	49	6	6	21	8	60
665. Закон о измени Закона о осигурању (2009)	53	43	22	16	43	9	65
666. Закон о измени Закона о осигурању (2012)	53	43	26	19	47	9	67
667. Закон о измени Закона о осигурању (2011)	53	43	24	17	45	9	66
668. Закон о измени Закона о безбедности саобраћаја на путевима (2014)	52	46	12	12	28	9	62
669. Закон о изменама и допунама Закона о уређењу судова (2010)	52	43	12	9	29	8	63
670. Закон о измени Закона о јавном информисању (2005)	51	44	6	5	20	7	55
671. Закон о измени Закона о Високом савету правосуђа (2005)	50	44	14	12	27	7	60
672. Закон о изменама Закона о ваздушном саобраћају (2012)	49	38	11	10	21	7	59
673. Закон о допуни Закона о јавном тужилаштву (2011)	48	42	14	13	31	9	60
674. Закон о изменама Закона о радиодифузији (2006)	47	41	14	13	34	7	57
675. Закона о измени закона о превозу у друмском саобраћају (2005)	47	40	10	10	19	6	55

*наставак на следећој страни*



Табела Г.1 – наставак са претходне стране

<b>редни број, назив и година доношења закона</b>	<b>речи</b>		<b>бројева</b>		<b>знакова</b>		<b>разм.</b>
676. Закон о измени закона о јавном бележништву (2012)	45	40	9	7	24	8	54
677. Закон о измени Законика о кривичном поступку (2008)	45	39	11	10	29	8	55
678. Закон о измени Закона о јавној својини (2013)	45	40	6	6	21	8	51
679. Закон о измени Законика о кривичном поступку (2007)	44	39	9	8	24	7	53
680. Закон о измени Закона о путним исправама (2008)	43	37	7	7	20	8	50
681. Закон о измени Закона о сточарству (2012)	41	36	8	8	18	7	49

## Биографија аутора

Небојша Васиљевић је рођен 16. децембра 1968. године у Београду, где је завршио Основну школу „Ђорђе Крстић” и Математичку гимназију. Дипломирао је 1993. године на Математичком факултету Универзитета у Београду, смер Рачунарство и информатика, са средњом оценом током студија 9,73. Звање магистра рачунарства је стекао 1998. године на Математичком факултету одбра- ном магистарског рада са темом „Формалне методе и модели података”.

Током школовања је освојио више награда и признања, као што су: трећа награда на Међународној математичкој Олимпијади у Хавани (1987), друга награда на Балканијади у математици у Атини (1987), Октобарска награда Београда за постигнуте резултате у математици (1987) и прва награда за израђени научно-истраживачки рад која је додељена поводом Дана Универзитета у Београду (1991). Такође је шест пута учествовао на средњошколским савезним такмичењима из математике и физике и освојио више награда и признања.

Од 1993. до 2004. године ради на Математичком факултету Универзитета у Београду у звањима асистента приправника и асистента, где је држао вежбе из предмета у области рачунарства, као што су основи програмирања, преводиоци и интерпретатори и програмски системи. Од почетка је био активно укључен у рад Рачунарске лабораторије Математичког факултета, а једно време је био и заменик шефа рачунарске лабораторије.

У том периоду је такође био стални сарадник часописа *PC* (од 1995. године), сарађивао је са Интернет одељењем Радија Б92 (1996-1999), где је реализовао пионирски подухват преноса радија преко интернета, био је саветник за информационе технологије председника Станком Холдинг компаније (1998-2000), координатор Радне групе за информациони систем у Министарству за

социјална питања (2001-2003) и био је један од оснивача предузећа *Software for Internet* (2001), које касније мења назив у *Inbox*.

Од 2004. године прелази као стално запослен у предузеће *Inbox* и ради као директор развоја до 2007. године, где руководи главним пројектима развоја софтвера и укупним технолошким развојем предузећа.

У оквиру сарадње са Центром за слободне изборе и демократију (CeSID, 2000-2007) био је координатор паралелног бројања гласова, а затим и члан управног одбора CeSID-а. Био је одговоран за методологију паралелног бројања гласова, пројектовање и развој одговарајућег софтвера, одређивање узорка, статистичку обраду и презентацију података.

Од 2007. до 2014. године ради као помоћник министра почевши од положаја помоћника министра у Сектору за информатичко друштво Министарства за телекомуникације и информатичко друштво и закључно са положајем помоћника министра у Сектору за електронске комуникације, информационо друштво и поштански саобраћај Министарства туризма, трговине и телекомуникација. Учествовао је у изради већег броја закона, подзаконских аката и стратегија, а израдом многих је и руководио. Руководио је реализацијом више пројеката који су имали изузетан значај за развој електронске управе и примену информационо-комуникационих технологија у образовању.

Од августа 2014. године, као виши саветник у Групи за развој Дигиталне агенде Министарства трговине, туризма и телекомуникација, наставља са активностима на кључним пројектима у области развоја електронских комуникација и информационог друштва, а пре свега у вези дигитализације терестричког емитовања телевизијског програма, примене информационо-комуникационих технологија у образовању и легислативе у области информационог друштва.

Објавио је већи број научних и стручних радова. Говори енглески језик.

Прилог 1.

## Изјава о ауторству

Потписани-а Небојша Васиљевић

број индекса \_\_\_\_\_

### Изјављујем

да је докторска дисертација под насловом

Аутоматска обрада правних текстова на српском језику

- резултат сопственог истраживачког рада,
- да предложена дисертација у целини ни у деловима није била предложена за добијање било које дипломе према студијским програмима других високошколских установа,
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио интелектуалну својину других лица.

**Потпис докторанда**

У Београду, \_\_\_\_\_

\_\_\_\_\_

Прилог 2.

## Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора Небојша Васиљевић

Број индекса \_\_\_\_\_

Студијски програм \_\_\_\_\_

Наслов рада Аутоматска обрада правних текстова на српском језику

Ментор Др Цветана Крстев, ванредни професор

Потписани/а Небојша Васиљевић

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла за објављивање на порталу **Дигиталног репозиторијума Универзитета у Београду**.

Дозвољавам да се објаве моји лични подаци везани за добијање академског звања доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

**Потпис докторанда**

У Београду, \_\_\_\_\_

\_\_\_\_\_

### Прилог 3.

## Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

Аутоматска обрада правних текстова на српском језику

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигитални репозиторијум Универзитета у Београду могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство

2. Ауторство - некомерцијално

3. Ауторство – некомерцијално – без прераде

4. Ауторство – некомерцијално – делити под истим условима

5. Ауторство – без прераде

6. Ауторство – делити под истим условима

(Молимо да заокружите само једну од шест понуђених лиценци, кратак опис лиценци дат је на полеђини листа).

**Потпис докторанда**

У Београду, \_\_\_\_\_

\_\_\_\_\_

1. Ауторство - Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.

2. Ауторство – некомерцијално. Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.

3. Ауторство - некомерцијално – без прераде. Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.

4. Ауторство - некомерцијално – делити под истим условима. Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.

5. Ауторство – без прераде. Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.

6. Ауторство - делити под истим условима. Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.