

УНИВЕРЗИТЕТ У БЕОГРАДУ
РУДАРСКО-ГЕОЛОШКИ ФАКУЛТЕТ

Александра Ђ. Томашевић

**РАЗВОЈ МОДЕЛА ЗА УПРАВЉАЊЕ
РУДАРСКОМ ПРОЈЕКТНОМ
ДОКУМЕНТАЦИЈОМ**

Докторска дисертација

Београд, 2018

UNIVERSITY OF BELGRADE
FACULTY OF MINING AND GEOLOGY

Aleksandra Đ. Tomašević

**MODEL DEVELOPMENT FOR
MINING PROJECT DOCUMENTATION
MANAGEMENT**

Doctoral Dissertation

Belgrade, 2018

Ментори:

Др Божо Колоња, редовни професор

Ужа научна област: *Пројектовање и планирање површинских копова*

Универзитет у Београду - Рударско-геолошки факултет

Др Ранка Станковић, ванредни професор

Ужа научна област: *Математика и информатика*

Универзитет у Београду - Рударско-геолошки факултет

Чланови комисије:

Др Никола Лилић, редовни професор

Ужа научна област: *Заштита на раду и заштита животне средине*

Универзитет у Београду - Рударско-геолошки факултет

Др Динко Кнежевић, редовни професор

Ужа научна област: *Заштита на раду и заштита животне средине*

Универзитет у Београду - Рударско-геолошки факултет

Др Милош Утвић, доцент,

Ужа научна област: *Библиотека информатика*

Универзитет у Београду - Филолошки факултет

Датум одбране: _____

РАЗВОЈ МОДЕЛА ЗА УПРАВЉАЊЕ РУДАРСКОМ ПРОЈЕКТНОМ ДОКУМЕНТАЦИЈОМ

САЖЕТАК

Рударство је једна од економски најзначајнијих привредних грана, чије се пословање организује и изводи по одговарајућим стандардима домаће и међународне рударске праксе. Током животног века једног рудника, који може трајати и више од сто година, генерише се обимна документација (дозволе, уговори, извештаји, елаборати, решења, инвестиционо-техничка документација и др.) у различитим видовима (текстуална, нумеричка, графичка, картографска и др.). Управљање документацијом у оваквим условима веома је важан и захтеван задатак. Ако се у обзир узме чињеница да се послови у рударству обављају у дистрибуираном окружењу: на руднику, у радионицама и канцеларијама, то управљање документацијом добија још више на значају.

Сходно томе, циљ докторске дисертације "Развој модела за управљање рударском пројектном документацијом" био је да се развије систем који би омогућио ефикасно управљање рударском пројектном документацијом у електронској форми, засновано на језичким технологијама, за претраживање и екстракцију информација употребом различитих језичких ресурса.

У истраживању су примењене дескриптивне, компаративне и аналитичке методе истраживања, које подразумевају упоређивање, вредновање и интерпретацију добијених резултата и анализу података из истраживања других аутора, затим статистичка обрада података, квалитативне и квантитативне методе за истраживање текста пројектне документације, управљање терминолошким базама и базама знања, као и адекватни ресурси и алати за моделирање знања и екстракцију структурираних података из неструктурираног текста пројектне документације уз обезбеђивање високог квалитета излазних информација.

Систем управљања рударском пројектном документацијом је развијен да интегрише текстуалне, лексичке, семантичке и терминолошке ресурсе, омогућавајући напредније претраживање докумената и екстракцију информација. Ови ресурси су интегрисани са скупом веб сервиса и апликација, за различите корисничке профиле и случајеве употребе.

За потребе прикупљања, описивања и систематизације рударске пројектне документације по фазама пројекта, развијена је рударска доменска онтологија RuDokOnto, осмишљена на начин да се повезује са другим сродним онтологијама, као што су на пример *EarthResource* (EarthResourceML, 2018), *MinExOnt* (Mentes, 2012) и др.

Евалуација модела развијеног система показала је различите могућности претраживања рударске пројектне документације. Такође, показано је да се коришћењем семантичке анализе може извршити формална провера сагласности рударског пројектног документа са стандардима и законском регулативом.

Резултати истраживања описаног система, који представља прво софтверско решење за имплементацију језичких технологија, показали су да је он применљив не само у рударству, већ и у другим инжењерским струкама. Систем је независан од врсте алфабета (ћирилица и латиница), што га чини примењивим на остале језике балканског региона сличне српском језику, а његова подршка морфолошким речницима може се применити на већину морфолошки сложених језика, као што су словански језици. Значајна побољшања претраживања и ефикасност екстракције информација заснивају се на семантичким мрежама и терминолошким речницима, уз подршку локалних граматике.

Кључне речи: рударство, пројектна документација, дигитална библиотека, природно језичке технологије, проналажење информација, екстракција информација

Научна област:

Рударско инжењерство

Ужа научна област:

Пројектовање и планирање површинских копова

УДК:

004.8:005.1:311

519.87

622:622.271:626.877

(043.3)

MODEL DEVELOPMENT FOR MINING PROJECT DOCUMENTATION MANAGEMENT

ABSTRACT

Mining is one of the most significant economic branches. Its business is organized and performed according to domestic and international mining practice. During a mine lifecycle, which may be more than hundred years, vast documentation is being generated (permits, contracts, reports, papers, technical documentation, etc.) in various aspects (textual, numerical, graphical, cartographic, etc.). Documentation management is very important and requiring task in those conditions. If the fact is taken into account that mining business is performed in the distributed environment: in the mine, in the workshops and offices, documentation management becomes more valuable.

Consequently, the aim of the PhD Thesis "Model development for managing mining project documentation" was to develop the system which could enable efficient managing of mining project documentation in electronic form, based on human language technology, for information retrieval and information extraction by using different language resources.

Descriptive, comparative and analytical research methods are applied in the research and they assume comparison, evaluation and interpretation of obtained results and data analysis out of other author research, statistical data processing, qualitative and quantitative methods for the research of the project documentation text, managing of terminological database, and knowledge database as well as adequate resources and tool for knowledge modeling and extraction of structured data out of non-structured project documentation text with having high quality of output data.

System for mining project documentation management is developed to integrate textual, lexical, semantical and terminological resources enabling advance document research and information extraction. These resources are integrated with a group of web services and applications for different user profiles and usages.

The mining domain ontology *RuDokOnto* is developed for the purpose of collecting, describing, and systematization of mining project documentation throughout the phases of the mining project's life cycle in a way that links other related ontologies, such as for example *EarthResource* (EarthResourceML, 2018), *MinExOnt* (Mentes, 2012) etc.

Model evaluation of the developed system has shown a variety of search options for mining project documentation. A formal check of mining project documents with standards and legislation using semantic analysis is also shown.

Research results of described system which represents the first software solution for implementation of language technologies have shown that its usage can be applied in other engineering branches, not just in mining. The system is independent of the type of alphabet (Cyrillic and Latin) which results in its usage on other languages throughout Balkan region which are similar to Serbian language. Its support to the morphological dictionaries may be applied on the most of morphological complex languages, as Slav languages are. Significant improvements of research and efficiency in information extraction are based on semantic networks and terminology dictionaries, with the support of local grammars.

Key words: mineral processing, project documentation, digital library, human language technologies, information retrieval, information extraction

Scientific field:

Mining Engineering

Scientific subfield:

Mine Planning and Design

UDC:

004.8:005.1:311

519.87

622:622.271:626.877

(043.3)

*"Није довољно стећи знање,
треба га знати и употребити."*

*Марко Тулије Цицерон
(Marcus Tullius Cicero, 106. - 43. пре н.е.)
римски државник, књижевник и беседник*

САДРЖАЈ

Списак слика.....	v
Списак табела	ix
Скраћенице	x
1. Увод	1
1.1. Предмет истраживања	1
1.2. Циљ истраживања	2
1.3. Полазне хипотезе	3
1.4. Научне методе истраживања	4
1.5. Структура рада	5
2. Преглед досадашњих истраживања	7
3. Модели за управљање документима.....	21
3.1. Основе теорије управљања знањем.....	21
3.1.1. Хијерархија знања	23
3.1.2. Модел управљање знањем	25
3.2. Управљање документима	29
3.2.1. Предности система за управљање пројектном документацијом.....	30
3.2.2. Општи модел система за управљање документима	32
3.3. Дигиталне библиотеке	34
3.3.1. Дефинисање концепта дигиталне библиотеке.....	35
3.3.2. Архитектура дигиталних библиотека.....	37
3.3.3. Дигитални објекти	38
3.3.4. Метаподаци.....	40
3.3.4.1. Категорије метаподатака.....	40
3.3.4.2. Формати метаподатака	41
3.3.4.3. Даблинско језгро.....	42
3.3.5. Примери дигиталних библиотека из домена рударства	44
3.3.6. Дигитализација	47
3.3.7. Омека као платформа за креирање дигиталне библиотеке	48
3.3.7.1. Развојно окружење Омеке	48

3.3.7.2. Програмски додаци Омеке	50
3.3.7.3. Креирање дигиталне библиотеке	54
3.4. Термилошки ресурси као ресурси знања.....	61
3.5. Онтологије	64
3.5.1. Дефиниција и основни појмови	64
3.5.2. Врсте онтологија.....	68
3.5.3. Принципи развоја онтологије.....	69
3.5.4. Језици за моделирање онтологија.....	70
3.5.5. Семантички веб и веб-сервиси	73
3.5.6. SWEET онтологија.....	76
4. Рударска пројектна документација.....	81
4.1. Инвестиционо-техничка документација у Републици Србији.....	89
4.2. Инвестиционо-техничка документација према светским стандардима у рударству	95
4.3. Пратећа рударска документација.....	99
5. Модел управљања рударском пројектном документацијом	101
5.1. RОмека@RGF – рударска дигитална библиотека	104
5.2. Корпус текстова из домена рударства.....	108
5.2.1. Припрема корпуса.....	108
5.2.2. Анотација и управљање корпусом.....	111
5.3. Лексички и термилошки ресурси.....	113
5.3.1. Обележавање речничких одредница	116
5.3.2. Рударски доменски маркери.....	116
5.3.3. Рударски семантички маркери.....	119
5.4. Претраживање, прелиставање, проналажење и екстракција информација	121
5.5. Изградња онтологије	125
5.5.1. Protégé – алат за изградњу онтологија	125
5.5.2. Доменска онтологија <i>RuDokOnto</i>	126
5.5.2.1. Класе	128
5.5.2.2. Примерци класа.....	136
5.5.2.3. Релације.....	137

5.5.2.4. Провера онтологије.....	140
5.6. Веза онтологија-документ.....	141
6. Евалуација модела развијеног система.....	146
6.1. Претраживање докумената у дигиталној библиотеци.....	146
6.2. Претраживање корпуса локалним граматикама.....	156
6.3. Претраживање корпуса регуларним изразима.....	161
6.4. Тестирање упита над онтологијом.....	163
6.5. Обележавање докумената према Смерницама TEI P5.....	167
7. Закључак и препоруке за даљи рад.....	178
Литература.....	181
Биографија	

СПИСАК СЛИКА

Слика 2.1. Модел структуре документа (Shreve, 2002)	8
Слика 2.2. Реализације модела портала за интелигентно управљање е-документима.....	10
Слика 2.3. Портал GeoSciML	15
Слика 2.4. Портал EarthResourceML	16
Слика 2.5. Owl Viz визуелизација онтологије <i>MinExOnt</i>	17
Слика 2.6. Онтологија <i>RudOnto</i> : панел за експортовање одредница речника са хијерархијом и садржајем OWL структуре (<i>горе</i>) и прелиставање концепата на српском или енглеском језику (<i>доле</i>)	18
Слика 3.1. Пирамида знања.....	24
Слика 3.2. Компоненте управљања знањем (Bhatt, 2003).....	27
Слика 3.3. Модел управљања знањем	28
Слика 3.4. Општи модел система за управљање документима	32
Слика 3.5. Метех.....	34
Слика 3.6. Пример употребе Даблинског језгра за описивање дигиталних објеката	44
Слика 3.7. Дигиталних библиотека LibraryMine.....	45
Слика 3.8. Дигитална библиотека универзитета <i>Penn State</i>	46
Слика 3.9. Дигитална библиотека <i>United Nations</i> и <i>University of North Texas</i>	46
Слика 3.10. Дијаграм елемената дигиталне библиотеке у Омеки.....	55
Слика 3.11. Описивање објеката метаподацима	57
Слика 3.12. Панел за увоз докумената.....	57
Слика 3.13. Унос геопросторних одредница, додељивање ознака и успостављања релација између објеката	58
Слика 3.14. Панели за креирање колекција.....	59
Слика 3.15. Креирање веб странице	60
Слика 3.16. Семантика у термилошким ресурсима	62
Слика 3.17. Класификација онтологија према (Gruninger <i>et al.</i> , 2008)	69

Слика 3.18. RDF искази у облику тројка: субјекат, предикат и објекат	71
Слика 3.19. Међусобни однос онтолошких језика: OWL Full, OWL DL и OWL Lite	72
Слика 3.20. Програмски језици Семантичког веба (Petrušić, 2016)	73
Слика 3.21. Архитектура семантичког веба	75
Слика 3.22. SWEET онтологије	77
Слика 3.23. Основне онтологије и њихове међусобне релације (Raskin, 2006)	79
Слика 4.1. Животни циклус рударских пројеката	83
Слика 4.2. Веза између минералних ресурса и минералних резерви	88
Слика 4.3. Структура, редослед израде и везе међу рударским пројектима	90
Слика 4.4. Земље чланице CRIRSCO асоцијације	97
Слика 5.1. Система управљања документацијом заснован на језичким технологијама	103
Слика 5.2. Централни репозиторијум за складиштење рударске пројектне документације	104
Слика 5.3. Почетни панел дигиталне библиотеке ROmeka@RGF	105
Слика 5.4. Рударски корпус	108
Слика 5.5. Расподела корпусних речи према типу текстова	110
Слика 5.6. Лексичка разноврсност текстова за рударски корпус	110
Слика 5.7. Лексички и термилошки ресурси	113
Слика 5.8. Подела скупа докумената упитом	123
Слика 5.9. Позиција онтологије у систему	125
Слика 5.10. Визуелизација онтологије <i>RuDokOnto</i>	127
Слика 5.11. Креирање онтологије	128
Слика 5.12. Креирање класа	129
Слика 5.13. Таксономија онтологије <i>RuDokOnto</i>	129
Слика 5.14. Таксономија класе <i>PravniSubjekt</i>	129
Слика 5.15. Таксономија класе <i>Basen</i>	130
Слика 5.16. Таксономија класе <i>Leziste</i>	131
Слика 5.17. Таксономија класе <i>Rudnik</i>	131

Слика 5.18. Таксономија класе <i>Dokument</i>	132
Слика 5.19. Таксономија класе <i>ZakonskaRegulativa</i>	132
Слика 5.20. Таксономија класе <i>PravniAkt</i>	133
Слика 5.21. Таксономија класе <i>InvTehDokumentacija</i>	134
Слика 5.22. Таксономија класе <i>Projekat</i>	135
Слика 5.23. Креирање примерака класа	136
Слика 5.24. Примерци класе <i>TehnickiRudProj</i>	137
Слика 5.25. Панели за креирање својства објекта и својстава типова података	138
Слика 5.26. Релације између докумената и фаза пројекта.....	140
Слика 5.27. Екстракција информација заснована на онтологијама (Wimalasuriya & Dou, 2010).....	141
Слика 5.28. XML приказ текста са анотираним именованим ентитетима.....	144
Слика 5.29. Приказ текста са анотираним именованим ентитетима.....	144
Слика 6.1. Претраживање докумената коришћењем стандардног упита.....	147
Слика 6.2. Претраживања докумената коришћење експанзије упита.....	149
Слика 6.3. Пример проширења упита за кључну реч "хомогенизација"	151
Слика 6.4. Резултати претраживања једночланих и вишечланих термина у дигиталној библиотеци R0meка@RGF	152
Слика 6.5. Графички приказ прецизности, одзива и Ф-мере.....	154
Слика 6.6. Синтаксички граф који препознаје ортографске варијанте назива <i>Tamnava Zapadno polje (горе)</i> са извученим конкорданцама (доле).....	157
Слика 6.7. Синтаксички граф који у рударском корпусу препознаје ортографске варијанте назива тамнавских поља (<i>горе</i>), са извученим конкорданцама (доле).....	159
Слика 6.8. Синтаксички граф који у рударском корпусу препознаје назив руде (<i>горе</i>), са извученим конкорданцама (доле).....	160
Слика 6.9. Граф и резултат његове примене за екстракцију примера обрасца из домена припреме минералних сировина.....	160
Слика 6.10. Пример конкорданци (<i>горе</i>) генерисаних у CQPweb систему са фреквенцијама (доле)	162

Слика 6.11. Структура SPARQL упита над онтологијом <i>RuDokOnto</i>	163
Слика 6.12. Онтолошке лексикализације.....	165
Слика 6.13. Пример упита који препознаје све ортографске варијанте назива примерка <i>TamnavaZapad</i>	165
Слика 6.14. Пример упита који препознаје инвестиционо-техничку документацију по фазама пројекта	166
Слика 6.16. Нивои обележавања текста закона и студије	167
Слика 6.17. Анотирање законске регулативе у складу са Смерницама ТЕИ Р5.....	169
Слика 6.18. Анотирање рударске пројектне документације у складу са Смерницама ТЕИ Р5.....	170
Слика 6.19. Вертикализоваан и анотиран сегмент текста закона (<i>лево</i>) и студије (<i>десно</i>)	174
Слика 6.20. Резултат добијен извршавањем CQP упита коришћењем структурних атрибута корпуса.....	176
Слика 6.21. Упућивање на документ и део документа.....	177

СПИСАК ТАБЕЛА

Табела 3.1. Елементи Даблинског језгра (ISO 15836-1:2017, Milenković, 2003).....	43
Табела 4.1. Стандарди за процену ресурса и резерви.....	96
Табела 4.2. Стандарди за евалуацију рударских пројеката	98
Табела 5.1. Структура колекција у дигиталној библиотеци.....	106
Табела 5.2. Резултати обраде текстова за рударски корпус	109
Табела 5.3. Доменски маркери из области рударства	117
Табела 5.4. Поддоменски маркери из области рударства	118
Табела 5.5. Семантички маркери за област рударства	120
Табела 5.6. Повезивање докумената са фазама пројекта.....	139
Табела 6.1. Резултати претраживања у дигиталној библиотеци ROmeKa@RGF	152
Табела 6.2. Прорачун вредности одзива, прецизности и Ф-мере.....	153
Табела 6.3. Препознати изрази са фреквенцијама добијени применом синтаксичког графа на рударски корпус.....	156
Табела 6.1. Могуће вредности атрибута <i>type</i> етикете <i><head></i>	171

СКРАЋЕНИЦЕ

AusIMM.....	Australasian Institute of Mining and Metallurgy; Институт за рударство и металургију Аустралазије
AVI	Audio Video Interleave
CBRR	Comissão Brasileira de Recursos e Reservas; Бразилска комисија за ресурсе и резерве
CIM.....	Canadian Institute of Mining, Metallurgy and Petroleum; Канадски институт за рударство, металургију и нафту
CMMI.....	Council of Mining and Metallurgical Institutes; Савет рударских и металуршких института
CQP	CWB query processor, CWB процесор упита
CWB.....	IMS Open Corpus Workbench
DAML-S	DARPA Agent Markup Language for Services
GCMD DIF ...	Global Change Master Directory; Directory Interchange Format for Datasets
GeoSciML	Geoscience Markup Language; Језик за обележавање геолошких термина
GIF	Graphic Interchange Format
GPL.....	General Public License
HLT	Human language technology; Језичке технологије
IASC	International Accounting Standards Committee; Одбор за међународне рачуноводствене стандарде
IE.....	Information extraction; Екстракције информација
IETF	Internet Engineering Task Force; Интернет инжењерска оперативна група
IMVC.....	International Mineral Valuation Committee; Међународни комитет за процену минералних сировина
IR	Information retrieval; Проналажење информација
IT	Information technology; Информационе технологије
IVSC	International Valuation Standards Committee; Међународно удружење за стандарде процене

JORC	Joint Ore Reserves Committee; Заједнички одбор за руде резерве
JPEG.....	Joint Photographic Experts Group
KWIC.....	Keyword in context; Кључна реч у контексту
KWICn	Keyword in center; Кључна реч у центру
LADL	Laboratoire d'Automatique Documentaire et Linguistique
LOM	Learning Object Metadata; Метаподаци за описивање наставна грађа
LR	Индекс лексичке разноврсности
LSA.....	Latent Semantic Analysis; Семантичка анализа скривених значења
METS.....	Metadata Encoding and Transmission Standard; Стандард за кодирање и пренос метаподатака
MOV.....	QuickTimeMovie
MOV	Методе остварене вредности
MPGE	Moving Picture Experts Group
MPIGM.....	Mongolian Professional Institute of Geosciences and Mining; Стручни институт за геолошке науке и рударство Монголије
NAEN	Национална асоцијација по експертизи недр; Национална асоцијација за експертизу рудника
NASA	National Aeronautics and Space Administration; Национална управа за аеронаутику и свемир
NCSA	The National Center for Supercomputing Applications
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting; Протокол отворених архива за прикупљање метаподатака
OASIS	Organization for the Advancement of Structured Information Standards; Организација за унапређење стандарда за структуриране информације
OBDA	Ontology Based Data Access; Приступ подацима заснованим на онтологијама
OCLC.....	Online Computer Library Center; Онлајн рачунарски библиотечки центар
OCR.....	optical character recognition; оптичко препознавање текста
OGC.....	Open Geospatial Consortium
OLAP	Online Analytical Processing; Online аналитичка обрада

OWL.....	Web Ontology Language; Веб онтолошки језик
PDF	Portable Document Format
PDF/A.....	Portable Document Format Archival
PERC.....	Pan-European Reserves & Resources Reporting Committee; Паневропски комитет за извештавање о минералним резервама и ресурсима
PNG.....	Portable Network Graphics
PoS.....	Part of Speech; Врста речи
PPT	Microsoft Power Point
RDF	Resource Description Framework; Оквир за описивање веб ресурса
RDFS.....	Resource Description Framework Schema
SAMREC.....	South African Code for the Reporting of Exploration Results, Mineral Resources and Mineral Reserves
SME.....	Society for Mining, Metallurgy, and Exploration; Удружење за рударство, металургију и истраживање
SOAP.....	Simple Object Access Protocol
SPARQL.....	Simple Protocol and RDF Query Language
SWEET.....	Semantic Web for Earth and Environmental Terminology; Семанти- чки веб за терминологију из области науке о земљи и екологију
SWWS	Semantic Web Enabled Web Services
TEI.....	Text Encoding Initiative; Иницијатива за кодирање текста
TIFF	Tagged Image File Format
URI	Uniform Resource Identifier; Јединствени идентификатор ресурса
W3C.....	World Wide Web Consortium
WAV	Waveform Audio File Format
WCDL.....	Web Capability Description Language
WMA	Windows Media Audio
WWW	World Wide Web
XML.....	eXtensible Markup Language; прошириви (мета)језик за означавање докумената
XTM	XML Topic Maps

1. УВОД

1.1. ПРЕДМЕТ ИСТРАЖИВАЊА

Документи у папирној форми се све више замењују електронским еквивалентима у савременом пословању, па и у рударству, тако да сваки похрањени електронски документ може бити преузет и обрађен по потреби касније. Рударски инжењери су већ преоптерећени документацијом и подацима, тако да ефикасан систем за управљање рударском пројектном документацијом постаје фактор који утиче на продуктивност и квалитет одлука у рударској привреди.

У светској литератури се на рударство традиционално гледа као на велику и сложenu индустријску грану, ипак, рударска индустрија данас тражи већу флексибилност, робусност уз једноставност и лакоћу руковања пројектном документацијом. Управљање токовима података и докумената је нарочито комплексно у системима чији се пословни процеси одвијају у дистрибуираном окружењу: на руднику, у радионицама и канцеларијама.

Рударство, као мултидисциплинарна грана индустрије користи и генерише документацију у различитим видовима: текстуалну, нумеричку, графичку и картографску, понекада и за рад у скоро реалном времену.

Имајући у виду потребу за ефикасним системом за проналажење информација у обимној и често неструктурираној документацији, потребно је развити систем који би на одговарајуће начине припремио пројектну документацију, описао је метаподацима и похранио у базу података, како би се касније омогућила брза претрага, екстракција информација и повлачење било делова документа било документа у целини, у зависности од потреба и захтева корисника.

Специфичност пројектне рударске документације, осим мултимедијалног садржаја, представља и архивирање текстуалне документације у два писма: ћирилици и латиници, што класични рачунарски системи за руковање

текстуалним документима не подржавају. Комплексна граматика српског језика и различити облици речи представљају још један изазов са којим се свакодневно сусрећемо при претрази докумената.

1.2. ЦИЉ ИСТРАЖИВАЊА

Циљ истраживања описаног у докторској дисертацији "Развој модела за управљање рударском пројектном документацијом" био је развој модела за ефикасно претраживање пројектне рударске документације независно од писма, подржавајући граматичке облике и синониме речи и претрагу по концептима. Сврха овог модела је да омогући рударском сектору успешан процес контроле пројектне документације по фазама развоја, који је не само брз, флексибилан и једноставан за употребу и имплементацију, већ покрива цео животни циклус рударских активности, од анализе, преко истраживања, инжењеринга, набавке, одржавања, сигурности на раду, експлоатације, па све до фазе затварање рудника, рекултивације, ремедијације итд.

Систематизација, структурирање и складиштење законске регулативе уз пројектну документацију треба да омогући унакрсно поређење и проверу сагласности сегмената пројектне документације са актуелним прописима.

Сврха овог модела је да се постигне потребни ниво сложености који предвиђа приступ који прати животни циклус рудника, тако да буде могуће лоцирање ризика и контролисање пројектног циклуса, координирање рада на комплексним пројектима са више уговорних пакета и више заинтересованих страна.

С тим у вези, истражена су светска искуства у овој области, специфичности сличних система за српски језик и могућност примене у успостављеном моделу управљања пројектом рударском документацијом.

У циљу реализације постављеног задатка у оквиру дефинисане теме:

- проучени су постојећи модели за управљање текстуалним документима,
- проучене су методе анализе текста, проналажења и екстракције информација,

- успостављен је систем индексирања пројектне документације,
- проучене су врсте и структуре пројектне документације, стандарда и законске регулативе,
- успостављен је систем описивања пројектне документације метаподацима,
- развијен је јединствени модел за проверу сагласности пројектне документације са законском регулативом и стандардима,
- евалуирани су резултате на реалним подацима.

Овај модел пројектну рударску документацију анализира на примеру пројеката и истраживања спроведених на Рударско-геолошком факултету Универзитета у Београду за потребе различитих рударских компанија.

1.3. ПОЛАЗНЕ ХИПОТЕЗЕ

Основна полазна хипотеза за истраживање била је та да је могуће развити модел за управљање рударском пројектном документацијом који ће обезбедити проверу конзистентности и сагласности са нормативним актима, пратити степен завршености, уз омогућавање ефикасног претраживања по концептима, независно од граматичких и ортографских варијација. Кроз пројектовани модел, развијена је формална онтологија, рударски корпус и имплементиран је алгоритам управљања пројектном документацијом. Основне претпоставке биле су:

- коришћењем семантичке анализе могуће је извршити формалну проверу сагласности рударског пројектног документа са стандардима и законском регулативом (да ли технички документ садржи све потребне елементе - анализа структуре документа, да ли се повезује са прилозима, да ли садржи табеле, слике итд.);
- употреба онтолошки засноване анализе пројектне документације применом система аутоматског резоновања повећава квалитет и обезбеђује адекватнију применљивост пројектованог решења;
- онтологија и систем базиран на знању служи и као семантички алат за претраживање и екстракцију информација;
- могуће је софтверски аутоматизовати онтолошки засновану анализу коректности пројектне документације применом система аутоматског

резоновања, што омогућује ефикасност и поузданост рада институција које се баве рударском проблематиком.

1.4. НАУЧНЕ МЕТОДЕ ИСТРАЖИВАЊА

За решавање проблематике описане у овој дисертацији коришћене су следеће научне методе истраживања:

- дескриптивне методе која укључују описивање, прикупљање и сређивање података,
- компаративне и аналитичке истраживачке методе која подразумевају упоређивање, вредновање и интерпретацију добијених резултата и анализу података из истраживања других аутора,
- статистичка обрада података,
- квалитативне и квантитативне методе за истраживање текста пројектне документације, управљање терминолошким базама и базама знања,
- адекватни ресурси и алати за моделирање знања и екстракцију структурираних података из неструктурираног текста пројектне документације уз обезбеђивање високог квалитета излазних информација.

За анализу текста коришћен је Unitex/GramLab (Unitex, 2017) скуп програма и електронски морфолошки речници за српски језик, а допуњени рударским терминима. За управљање документима су коришћене MongoDB (MongoDB, 2017) и MySQL (MySQL, 2017) базе отвореног кода, као и Microsoft SQL база. За индексирање и претрагу пуног текста докумената (пројектне документације, стандарда и законске регулативе) коришћени су системи отвореног кода (NoSketch Engine (NoSketch, 2017) и CQPweb (CQP, 2017)).

Онтологија пројектне рударске документације, као потпуно нова онтологија, базирана је на постојећим референтним онтологијама као што је Даблинско језгро (енг. *Dublin core*) (DCMI, 2017) за метаподатке, постојећим онтологијама за рударство, стандарде и законску регулативу

1.5. СТРУКТУРА РАДА

Докторску дисертацију, поред општих елемената, које чине апстракт на српском и енглеском језику, кључне речи, предговор, садржај, списак скраћеница, слика, табела и литературе, чине и следеће структурне целине:

1. Увод
2. Преглед досадашњих истраживања
3. Модели за управљање документима
4. Рударска инвестиционо-техничка документација
5. Модел управљања рударском пројектном документацијом
6. Евалуација модела развијеног система
7. Закључак и препоруке за даљи рад

Прво поглавље представља увод и у њему је дат осврт на предмет истраживања, циљ истраживања, описане су полазне хипотезе, као и научне методе које су коришћене у истраживању.

Друго поглавље даје преглед досадашњих истраживања и литературе која је коришћена у овом раду.

Треће поглавље описује неке од модела за управљање документима. Дате су основе теорије управљања знањем и управљања документима. Такође, описане су дигиталне библиотеке, термилошки ресурси, ресурски знања и онтологије, као саставни елементи описаног система.

Четврто поглавље даје преглед и опис рударске пројектне документације према важећој законској регулативи Републике Србије. Дат је и преглед међународних стандарда везаних за процену минералних ресурса и минералних резерви, односно за евалуацију рударских пројекта. Такође, описана је и пратећа рударска документација, неопходна за рад и праћење успешности пословања неког рудника, као што су на пример: прегледни и детаљни извештаји, елаборати о обављеним радовима, елаборати о уоченом (забележеном) реалном стању, решења, жалбе и приговори на решења, итд.

Пето поглавље описује систем управљања рударском пројектном документацијом заснован на језичким технологијама. Детаљно је описана дигитална библиотека ROneka@RGF, као основни репозиторијум пројектне документације, законске регулативе и научне и стручне литературе. Сви објекти дигиталне библиотеке послужили су као основа за креирање првог корпуса текстова, а потом и скупа термина из домена рударства. По први пут су креирани рударски доменски, поддоменски и семантички маркера чији је главни задатак обележавање рударских термина ради каснијег ефикаснијег проналажења информација.

Шесто поглавље се бави евалуацијом модела развијеног система. На неколико примера демонстрирано је претраживање рударске пројектне документације. Приказано је проналажење докумената у целини или само екстракција неких делова, њихово претраживање преко кључних речи, претраживање рударског корпуса локалним граматикама и/или регуларним изразима, као и екстракција информација. Такође, приказано је тестирање упита над изграђеном онтологијом, као и структурно обележавање докумената и анотација именованих ентитета.

Седмо поглавље даје закључке и наглашава научни и практични допринос примењене методологије и развијеног система. Такође у овом поглављу су дате и препоруке за будући научноистраживачки рад у правцу даљег унапређења приказаног модела.

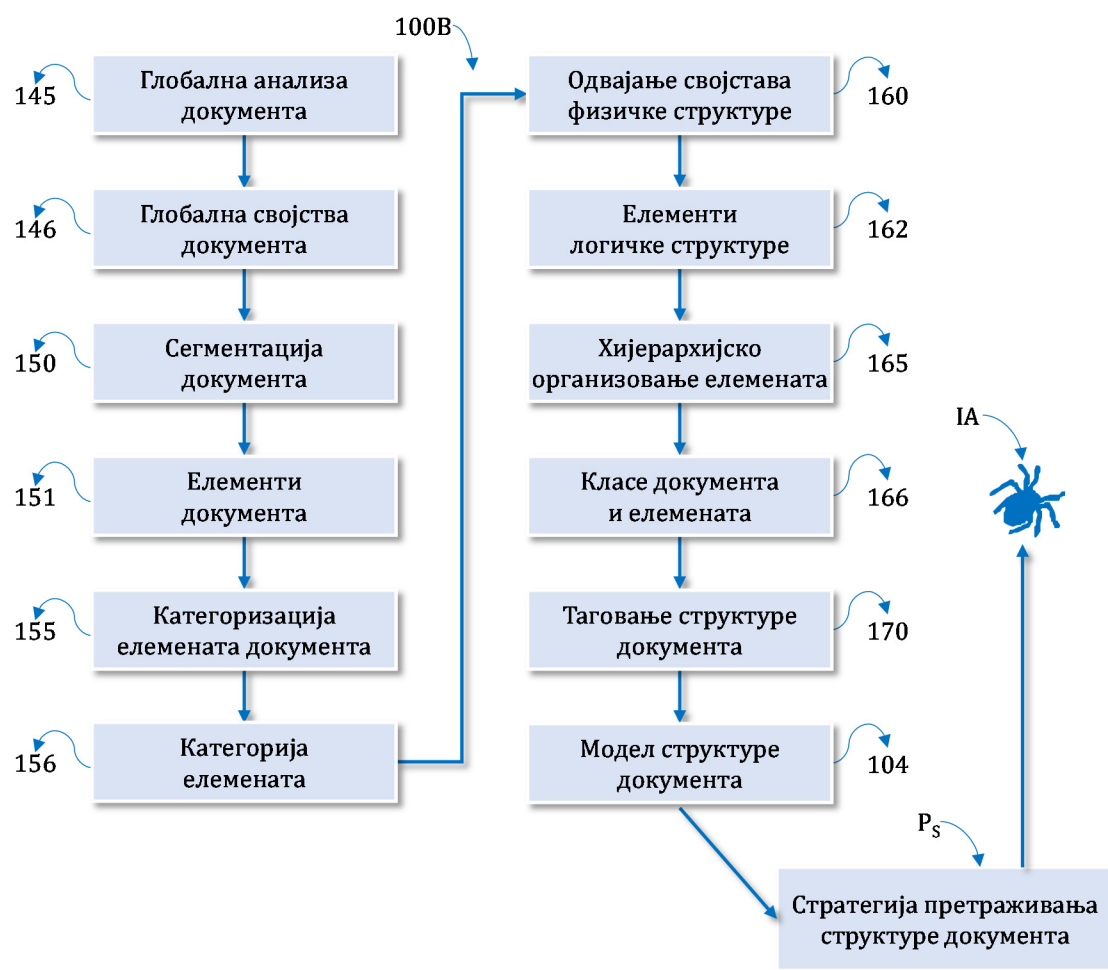
2. ПРЕГЛЕД ДОСАДАШЊИХ ИСТРАЖИВАЊА

Gregory Shreve са универзитета Kent State из Охаја, УСА, 2002. године је патентирао процес за управљање документима и превођење докумената, заснован на корпусима документације, које креирају, обрађују и претражују интелигентни агенти (Shreve, 2002). Метод управљања документима користи корпус докумената, и то почевши од прикупљања изворних докумената у електронској форми, до моделирања корпуса у смислу домена и структуре информација који се екстрахују из докумената, како би се идентификовали параметри којима треба обогатити корпус и снабдети га додатним информацијама. Коришћењем мета-језика за електронско обележавање корпуса и програмирање интелигентних агената, који подржавају параметре проширења корпуса, се омогућава даље унапређење коришћењем интелигентних агената за претраживање и других репозиторијума за проналажење сличних ресурса, чиме се континуирано врши унапређење полазног корпуса. Систем се не ограничава само на један језик, већ има за циљ да подржи вишејезично коришћење докумената уз могућност аутоматског превођења, како упита тако и самих докумената. Континуирано прикупљање нових ресурса и обучавање статистичких модела, треба да обезбеди стално унапређење система у смислу квалитета и квантитета, односно да уз повећање одзива задржи прецизност понуђених одговора.

Метод управљања документима, који користи корпус докумената, обухвата:

- прикупљање изворног корпуса докумената у електронској форми,
- моделирање изворног корпуса у смислу информација о документу и доменима како би се идентификовали параметри побољшања корпуса,
- употребу мета-језика за електронско аотирање изворног корпуса,
- програмирање параметара побољшања корпуса у интелигентном агенту и
- употребу интелигентног агента за претраживање различитих извора сличних појмова и структура, и њихово враћање изворном корпусу што и њега додатно побољшава.

Дијаграм који приказује кораке у изградњу модела структуре документа, приказан је на слици 2.1 и биће описан у наредном тексту.



Слика 2.1. Модел структуре документа (Shreve, 2002)

Глобална анализа (позиција 145) документа укључује утврђивање дужине документа, читљивост, терминолошку густину, језик и друга глобална својства документа (позиција 146). Сегментација документа (позиција 150) подразумева поделу у сегменте или елементе (блокови слика, табела, пасуса), при чему се број сегмената складишти као један од глобалних параметара документа. Категоризација докумената (позиција 155) се врши на основу заједничких карактеристика, на пример броја и величине сегмената, релативног положаја у документу, релативних веза изнад и испод, присуство препознатљивих лексема, властитих имена, колокација, семантички значајних стилистичких информација. Одвајање (позиција 160) својстава

физичке структуре од логичке структуре је праћено очувањем физичког изгледа сваког сегмента и груписањем логичких елемената у класе, где год је то могуће. Хијерархијско организовање елемената (позиција 165) се врши хеуристички на основу различитих својстава: величине слова, набрајања, нумерација, дужине параграфа и сл. Тагирање, односно морфолошко-синтаксичка анотација, документа и елемената документа (позиција 170) користи шеме метаподатака за логичко представљање структуре документа. Модел структуре докумената у корпусу (позиција 104) има различиту грануларност, од микроструктурног нивоа (речи, колокација, фраза, реченица) до макроструктурног нивоа (параграфи, групе параграфа,...) до супер-структурног нивоа (наслов, назив поглавља, одељка и сл.). Ове структуре се могу аутоматски одредити и описати шемом метаподатака мета-језика (на пример XML (XML, 2017)). У случају да већ постоји обележавање у документу, обично га треба очувати и ново обележавање интегрисати као додатни слој.

Моделом управљања електронским документима бавио се и (Ђокић, 2012). Кључна идеја овог решења јесте развој и имплементација модела портала за интелигентно управљање електронским документима, заснованог на концептима персонализације и адаптивности, који унапређује и трансформише пословне процесе. Оригиналноост овог модела огледа се у дефинисању методолошког оквира интеграције различитих извора података и њихове презентације у великим пословним системима. Имплементирано решење модела портала за интелигентно управљање електронским документима (слика 2.2) реализовано је на *cloud computing* инфраструктури, која омогућава флексибилност, проширивост капацитета, поузданост и безбедност система. Развијени модел портала за интелигентно управљање електронским документима примењен је у реалном пословном систему. Постигнут је висок степен размене и доступности информација и електронских докумената, заједнички рад корисника, персонализације и адаптивности, као и поузданости и сигурности самог система. Основна идеја овог модела заснива се на адаптацији постојећег система за управљање електронским документима, без потребе за развојем нових софтверских модула. MS SharePoint (SharePoint, 2018)

је један од најраспрострањенијих система за сарадњу, дељење информација и за управљање документима, који прати комплетан животно циклус документа, од тренутка креирања, преко публикувања, до тренутка архивирања.



Слика 2.2. Реализације модела портала за интелигентно управљање е-документима

MS SharePoint је послужио као платформа за реализацију портала за интелигентно управљање електронским документима. Адаптација је извршена тако да њен процес буде економски исплатив, ефикасан и погодан за примену на коришћеној платформи. Персонализација која је предложена у овом моделу реализује се на нивоу групе према следећим критеријумима: припадности организационој целини, радном месту, *ad hoc* послу, преференцијама или на основу неке од комбинација претходних критеријума. Структура предложеног модела за интелигентно управљање електронским документима заснована је на примени пословне интелигенције и обухвата следеће компоненте: архитектура портала за интелигентно управљање електронским документима, модел пословних процеса, активности развоја

адаптивног система за управљање е-документима, адаптација и мерење перформанси решења. Имплементација и примена развијеног модела портала за интелигентно управљање електронским документима реализована је применом фаза приказаних на слици 2.2.

Приказани модел омогућава претраживање докумената само до нивоа метаподатака и кључних речи, а резултати претраживања су документи у целини. Оно што овим моделом није обухваћено, јесте подршка за српски језик, имајући у виду његову морфолошку комплексност и постојање два равноправна алфабета (ћириличног и латиничног). Ово је нарочито значајно у погледу побољшања начина претраживања, не задржавајући се на нивоу проналажења самог документа, већ претраживања његовог садржаја.

Пример развоја модела управљање пројектима отварања површинских копова угља, који се односи на планирање, праћење и контролу тока реализације пројектних активности и процеса предложио је (Stojanović, 2015). Модел је рађен по Методи остварене вредности (MOV) на примеру површинског копа Угљевик Исток, Република Српска. Ова метода се у пракси показала погодном за праћење реализације већине пројеката, интегришући трајање, трошкове, и временски распоред, чиме је постала стандардан алат за управљање пројектима у многим организацијама. Применљива је за готово све врсте пројеката, без обзира на њихову величину или подручје примене, а посебно је погодна за управљање комплексним пројектима попут инвестиционих пројеката у области површинске експлоатације угља. Модел представља синтезу теоријских основа као и практичних аспеката реализације пројеката отварања површинских копова угља. Управо практични аспекти инкорпорирани у теоријске основе омогућили су развој модела управљања пројектима отварања површинских копова који нуди механизам за планирање и реализацију пројеката као и разноврстан скуп метода и алата за оптимизацију, а све у функцији поуздане реализације оваквих пројеката. Дефинисани процесни модел и развијена структура модела представља процедуру стручњацима који практично реализују пројекте отварања површинских копова. Развијени модел управљања пројектима отварања површинских копова, као методологија и процедура може се

применити на било које рударске пројекте и омогућава њихову ефективнију и ефикаснију реализацију у односу на досадашњи приступ. Аутор истиче да је коришћењем овог модела могуће, још у фази планирања рударског пројекта, избећи неке од потенцијалних ризика, како са техничко-технолошког, тако и са социјалног, еколошког и економског аспекта. Предвиђена итеративност која се постиже симулацијом будућих стања у фази планирања оваквих пројекта омогућује стално преиспитивање резултата и учења што је свакако додатни бенефит реализованог модела. Дефинисани модел у је верификован на примеру отварања површинског копа угља Угљевик Исток.

Међутим, и поред тога што се аутор бави кључним областима управљања пројектом отварања површинског копа (као што су: обим пројекта, динамика и рокови, трошкови, квалитет на пројекту, људски ресурси, комуникација, ризици, набавка опреме, уговори), приказано решење, не нуди решење за прикупљање, складиштење, поновно коришћење, размену података и докумената међу корисницима, као и трајну заштиту и архивирање пројектне документације. Такође, није решено ни питање прелиставања и претраживања пројектне документације.

Један од ефикасних начина за решење ових проблема свакако су дигиталне библиотеке. Дигиталним библиотекама детаљно се бавила (Trtovac, 2016). Посебна пажња у истраживању посвећена је методама проналажења информација у дигиталним библиотекама, са аспекта повећања одзива и побољшања прецизности постављеног упита. Разлог за то је што дигиталне библиотеке треба да понуде боље и шире могућности претраге – преко пуног текста докумената који су у њима садржани, а и преко метаподатака којима су документи описани. Тиме дигиталне библиотеке постају много више од колекција докумената која су у дигиталном облику постављене на мрежу. Аутор се осврнуо на различите методе проналажења информација, испитујући њихову ефикасност. Акцент је стављен на дескрипторе метаподатака и дескрипторе садржаја, као потенцијалне методе бољег и ефикаснијег проналажења информација. Постављајући најразличитије упите и претражујући записе за дигиталне објекте, као и садржај самих дигиталних објеката утврђено

је да, без обзира на све функционалности које нуде различити формати за унос метаподатака, као и везе ка семантичком вебу, претрага преко метаподатака не даје добре резултате у смислу прецизности. Одзив је углавном висок, међутим број релевантних докумената је углавном изузетно низак. С друге стране, претрага преко комплетног текста коришћењем дескриптора садржаја дала је неупоредиво боље резултате нарочито у оним дигиталним колекцијама које су опремљене квалитетним лексичким ресурсима, у првом реду морфолошким речницима. Рад на овом истраживању резултирао је и допуном општег електронског речника српског језика, као и допуном терминолошког речника из области библиотекарства и информатике.

Приликом израде рударског корпуса коришћене су методе и алати за конструкцију корпуса савременог српског језика као референтног језичког ресурса. Овом проблематиком бавио се (Utvić, 2014) у својој докторској дисертацији. У раду су описане могуће класификације корпуса и посебно су издвојени национални корпуси као општи, референтни корпуси који претендују да репрезентују језик једне земље. Детаљно су анализирани национални корпуси словенских језика. Као параметри корпуса, посебно су анализирани носач, домен и намена, обим (величина), период, извор/медијум, анотација и вишејезичност. Приказане су радње неопходне за обраду корпуса: прикупљање, дигитализација и класификација текстова за корпус, конверзија корпусних текстова у одговарајући формат електронског текста, лингвистичка обрада и анотација електронских текстова за корпус, као и индексирање и компресија текстова корпуса. Када је у питању анализа корпуса, детаљно су описани механизми претраге корпуса, посебно формализам регуларних израза, потом конкорданце као метод за визуелизацију података из корпуса који одговарају корисниковом упиту и на крају основе статистичке анализе корпуса.

У истраживању (Vasiljević, 2015) разматрана су посебна језичка правила у текстовима законских прописа на српском језику која се могу изразити методама рачунарске лингвистике. Законски 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. Ово је веома значајно и за рударску инвестиционо техничку документацију, која такође има јасно дефинисану логичку структуру, те су ова истраживања била од посебног значаја. Циљ истраживања био тај да се на теоријском нивоу створи основа за развој софтверских алата за обраду текстова прописа и да се испитају конкретне могућности имплементације. Том приликом коришћене су методе засноване на правилима, на супрот 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.

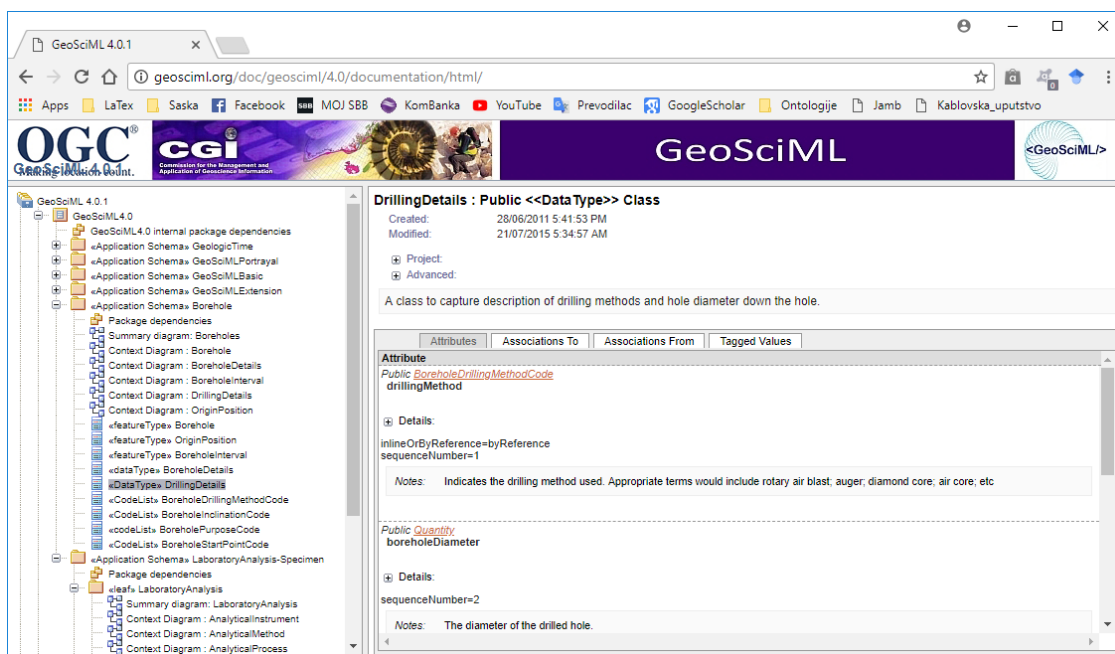
Као технолошка основа за подршку наведеној методологији је коришћен систем Unitech који је развијен у Laboratoire d'Automatique Documentaire et Linguistique (LADL) у оквиру Institut Gaspard Monge, Université de Marne-la-Vallée, као и електронски речници развијени у оквиру Групе за језичке технологије Универзитета у Београду (Krstev, 2008).

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.

Пример онтологија из домена геонаука су: GeoSciML (енг. *Geoscience Markup Language*) (GeoSciML, 2018) за домен геологије и EarthResourceML (EarthResourceML, 2018) за домен рударства, коришћене као онтологије вишег нивоа општости, које је развила међународна комисија за управљање и примену информација у геонаукама IUGS CGI (енг. *Commission for the Management and Application of Geoscience Information*) (CGI, 2018).

GeoSciML је језик за обележавање (и моделовање) геолошких карактеристика. Најновија верзија 4.1 званично је потврђена као OGC (OGC, 2017) (енг. *Open Geospatial Consortium*) стандард. GeoSciML је стандард за размену дигиталних података из домена геонаука, заснован на XML-у. Прилагођен је за представљање и описивање геолошких карактеристика (геолошке јединице, структура, материјали) и елемената геолошких истраживања (на пример: бушотине, узорци, мерења) (слика 2.3). Објекти, као што су речници или геолошка временска скала, као повезани ресурси, користе се као класификатори примарних објеката.



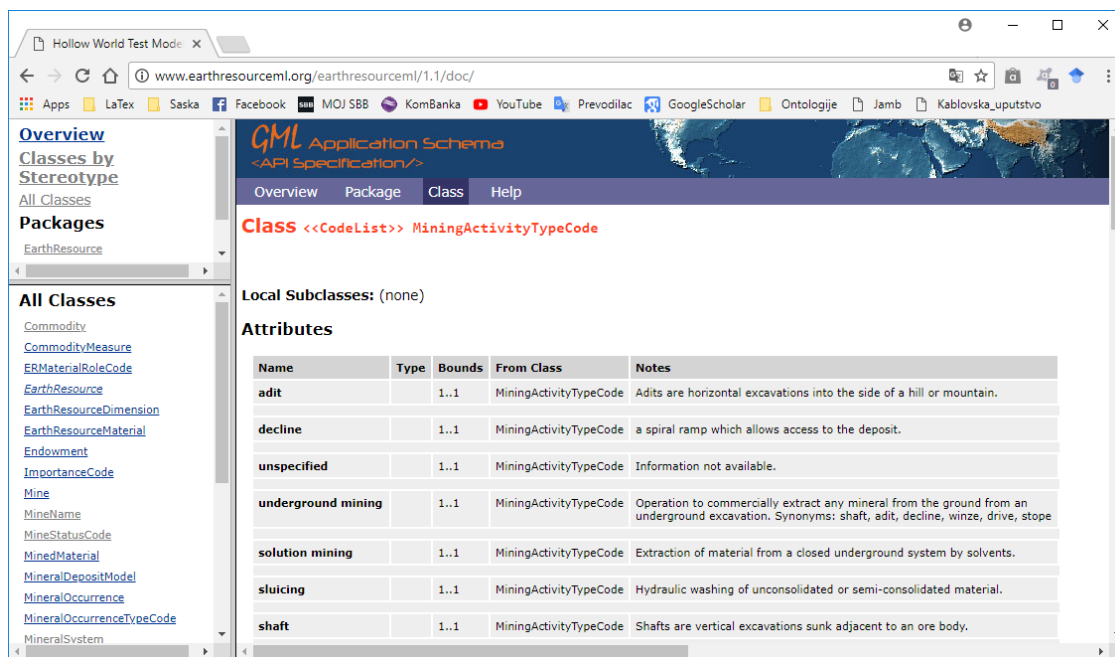
Слика 2.3. Портал GeoSciML¹

EarthResourceML је XML стандард за размену дигиталних података о минералним ресурсима, минералним резервама, рудницима и рударској активности (слика 2.4).

Основне класе ове онтологије, коришћене у изради онтологије *RuDokOnto* су: *ClassificationMethodUsed* (класификације минералних ресурса и минералних резерви засноване на светским стандардима: CIM, CRIRSCO Code, JORC CODE, NI 43-101, PERC Code, Russian Code, SAMREC Code, SME Guide, UNFC Code и др.),

¹ <http://geosciml.org/doc/geosciml/4.0/documentation/html/> (приступљено 15.01.2018.)

MineStatus (статус рудника: активан, неактиван или у развоју, *MiningActivity* (активност рудника: површинска, подземна или подводна), *ReserveCategory* (класификација минералних ресурса) и *ResourceCategory* (класификација минералних резерви).



Слика 2.4. Портал EarthResourceML

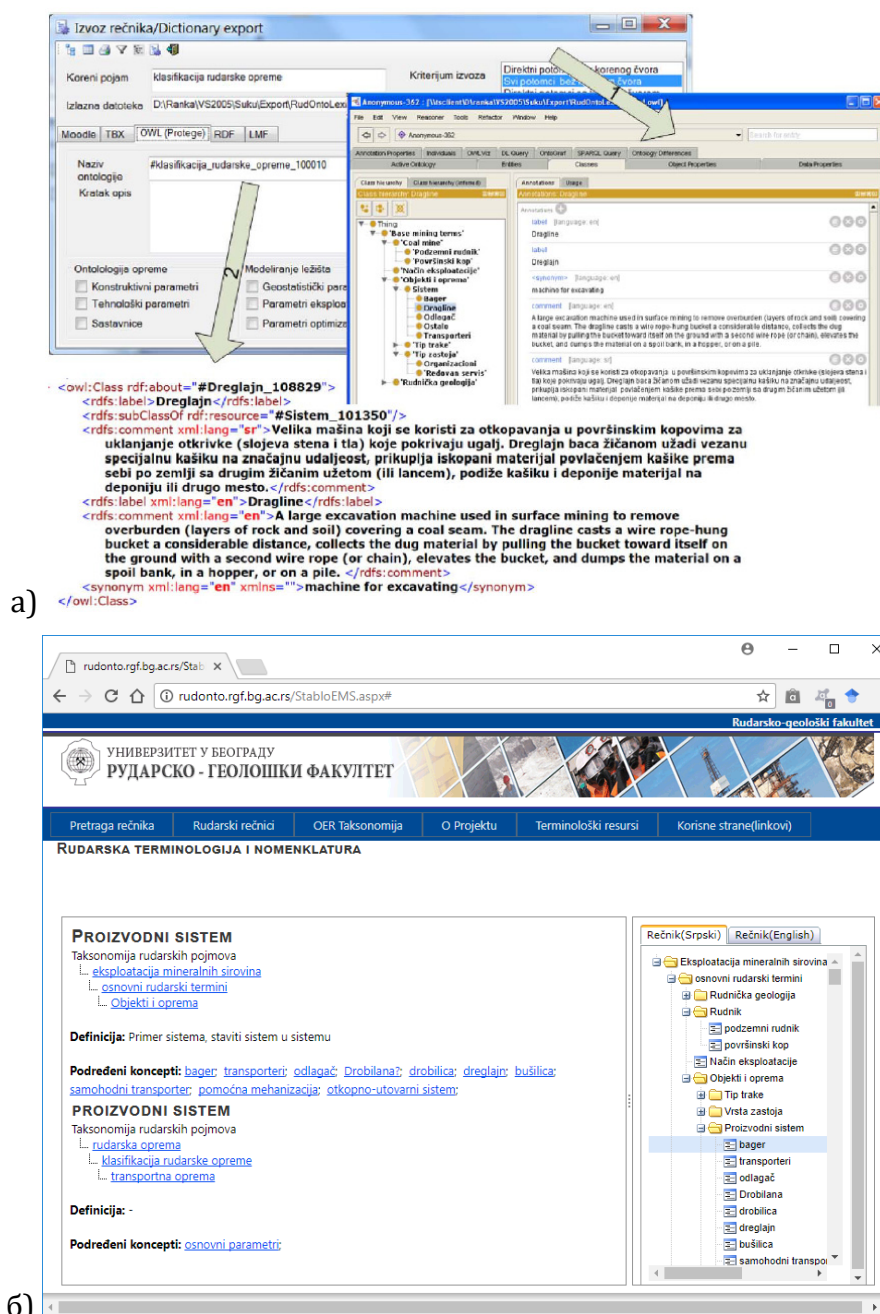
Значајно је напоменути истраживање Hilal Sevindik Mentès, чији је циљ био да се покаже да онтологије засноване на OWL језику (OWL, 2017) могу бити искоришћене за управљање и представљање геолошког знања, за унапређење коришћења и побољшање интеграције хетерогених геолошких података (Mentès, 2012), што је резултирало креирањем онтологије *MinExOnt* (слика 2.5).

Доменска онтологија *MinExOnt* описује истраживања минералних сировина, пружајући на знању засноване, семантички међусобно повезане речнике. Развијени речници представљају знање о истраживању минералних сировина и експлицитно формализују семантички однос између појмова, на пример, истраживања лежишта и природних ресурса. У раду је описана употреба софтвера Protégé и Knoodl-OntVis, који се примењују за развој онтологије. Онтологија приказана у овом истраживању је јавно доступна. Комплетан код је публикован тако да се може користити за даља истраживања.



Слика 2.5. Owl Viz визуелизација онтологије *MinExOnt*

Применом онтологија у рударству бавила се (Kolonja, 2016). Она је дефинисала методологију и развила модел система управљања заштитом на раду RudOnto (слика 2.6) како би се унапредили послови управљања заштитом на раду и омогућило правовремено и ефикасно планирање и спровођење потребних мера заштите чиме би се обезбедило унапређење стања заштите на раду.



Слика 2.6. Онтологија *RudOnto*: панел за експортовање одредница речника са хијерархијом и садржајем OWL структуре (*горе*) и прелиставање концепата на српском или енглеском језику (*доле*)

Како постојећи ниво организације заштите на раду на рудницима у већини случајева не обезбеђује благовремено откривање узрока нарушавања безбедности, то се указала потреба за развојем система заснованог на аналитичким алатима пословне интелигенције, који би омогућио рано откривање ризика од повреда на раду, пратећи потенцијалне узроке, лоцирајући места догађаја, времена дешавања и тежину повреда. Модел интегрише онтологију *RudOnto* са стандардним OLAP (енг. *Online Analytical Processing*) технологијама које се користе за организацију пословних база података и за подршку пословној интелигенцији. Своје утемељење систем је нашао у трећој фази Деминговог циклуса² – Провера и корективне мере. Прва фаза развоја система обухватила је спецификацију корисничких захтева и пословног модела који специфицира архитектуру информационог система кроз концептуални модел. У следећој фази развоја, моделирање је обухватило: концептуално моделирање података (даје целокупан поглед на систем), структурно моделирање (формализује организацију пословног система), процесно моделирање са спецификацијом пословних активности у систему и моделирање понашања и интеракције између ресурса.

Евиденција и анализа индикатора стања заштите на раду, као што су незгоде, повреде на раду и професионална обољења, треба да омогуће комплексну и свеобухватну анализу прикупљених података и оцену успешности примењених мера заштите као и унапређење процеса доношења одлука везаних за побољшање примењених и избор нових мера при управљању процесима заштите на раду у рудницима.

Значај система огледа се у повећања доступности, проналажењу информација и приближавању знања, као и могућност контролисаног одржавања података ажурним.

Могућност практичне примене овог система пословне интелигенције за управљање заштитом на раду у рудницима илустрован је на подацима евиденције података из области безбедности и здравља на раду служби заштите на површинским коповима Електропривреде Србије.

² Демингов PDCA циклус: планирај-уради-провери-побољшај (енг. *Plan-Do-Check-Act*)

У раду "Развој геолошког термилошког речника GeolISSTerm" (Stanković *et al.*, 2011b) приказана је структура и имплементација електронског геолошког речника GeolISSTerm (GeolISS, 2017), као таксономије основних геолошких појмова и термина. Иницијална верзија геолошке терминологије је настала крајем 2005. и током 2006. године у оквиру развоја геолошког информационог система, GeolISS-а, као ресурс за грађење домена, односно за контролу и валидацију уноса података у базу. У развоју су учествовали стручњаци са Рударско-геолошког факултета Универзитета у Београду и институција које се баве геологијом. Геолошки речник развијан је у оквиру издвојеног подпројекта GeolISS -а под називом "Развој геолошке терминологије и номенклатуре за геолошку базу података Републике Србије". Циљ овог пројекта био је да се постојећи фонд геолошких појмова и термина допуни и прошири у првом реду појмовима из области специјалистичких геолошких дисциплина који у иницијалној верзији нису били заступљени у жељеном обиму. У оквиру GeolISSTerm-а су развијени речник геолошких метаподатака и основни геолошки термини, петролошка, структурно-геолошка, стратиграфска, палеонтолошка, инжењерско-геолошка, хидрогеолошка геохемијска и економско-геолошка терминологија и класификација. Овај пројекат је финализиран 2009. године у виду функционалног геолошког појмовника који оптимално опслужује Геолошки информациони систем Републике Србије (GeolISS, 2017). Ово је први пут да се у један геолошки информациони систем уграђује конзистентан геолошки речник који обухвата основне геолошке класификације у области фундаменталне и апликативне геологије. Посебно је објашњен концепт контролисања домена у геолошком информационом систему Србије.

3. МОДЕЛИ ЗА УПРАВЉАЊЕ ДОКУМЕНТИМА

3.1. ОСНОВЕ ТЕОРИЈЕ УПРАВЉАЊА ЗНАЊЕМ

Знање је флуидна мешавина оквирног искуства, вредности, контекстуалних информација и експертских мишљења, која осигурава оквир за процену и укључивање нових искустава и информација – изворно долази из и примењује се у главама зналаца. Знање је у организацијама похрањено не само у документима или репозиторијума већ исто тако и у организацијским рутинама, процесима, праксама и нормама (Davenport & Prusak, 1998).

Знање се може дефинисати и као нематеријални ресурс, као слика стварности исказана човековим схватањем док посматра простор, објекте, односе и догађаје који га окружују. Оно се састоји од интуиције, скупа идеја, искустава, вештина и учења, који имају потенцијал стварања нове вредности (Brzaković & Straživuk, 2016).

У општем смислу, знање може бити имплицитно и експлицитно (Nonaka & Takeuchi, 1995). Имплицитно или индивидуално знање (енг. *tacit knowledge*) је неписано, неизговорено, скривено знање које поседује практично свако нормално људско биће, а стекло га је на основу својих емоција, искустава, запажања, интуиције и у комуникацији са другим људима. Представља основу на којој почива експлицитно знање. Експлицитно или изричито знање (енг. *explicit knowledge*) је артикулисано знање, које се изражава речима, бројевима, кодовима, математичким и научним формулама или музичким нотама. Ово знање се лако дистрибуира визуелним или вербалним путем, помоћу књига, филмова, интернета, итд.

Према Zack-у, знање се дели на: декларативно – знати нешто о некоме или нечему (енг. *know-what*), процедурално – знати како (енг. *know-how*), каузално – знати зашто (енг. *know-why*), кондиционално – знати када (енг. *know-when*) и релационо – знати ко/шта с ким/чим (енг. *know-who, know-where*) (Zack, 1999b). Он, такође, знање дели, према начину управљања на: тактичко и стратешко.

Тактичко знање се односи на краткорочно (енг. *short-term*) позиционирање предузећа у односу на тржиште, конкуренцију и добављаче и др. Стратешко знање односи се на дугорочно (енг. *long-term*) позиционирање организације у односу на неку визију и стратегију за остварење те визије (Zack, 1999a).

Према Jean-Francois Lyotard-у знање се дели на: наративно (пресудно у обликовању традиционалног знања, а прича је првенствени облик овог знања), научно (континуирано преплитање истраживања и учења), позитивистичко (темељи се на актуелном осећајном искуству) и критичко (темељи се на критичком погледу и мишљењу) (Lyotard, 1993).

У наредном тексту биће дат кратак преглед основних појмова управљања знањем (Blodgett *et al.*, 2005):

Група корисника (енг. *Community of Practice*) - група појединаца који деле знање, уче једни од других у директном контакту или виртуелно. Они се удружују захваљујући заједничком интересу или домену и вођени су жељом да деле проблеме, искуства, сазнања и добру праксу. То су, на пример, групе за иновацију, стратешке, групе за узајамну помоћ, заједнице за прикупљање најбољих примера из праксе.

Едитор, редактор или уредник (енг. *Knowledge Worker*) - неко ко се првенствено фокусира на производњу, обраду и коришћење знања.

Инвентар знања (енг. *Knowledge Inventory*) - системска идентификација знања једне организације. Будући да такво знање често може бити прећутно (енг. *tacit knowledge*), инвентар често може бити "индикатор" људима, а не само знање.

Интелектуални капитал (енг. *Intellectual Capital*) - способност организације за управљање знањем и генерисање вредности на основу: *људског капитала* – стручности, вештине и способности запослених, *структуре капитала* - организациони процеси, процедуре, инфраструктура, патенти и ауторска права; *друштвеног капитала* - способност запослених да комуницирају и сарађују кроз време и простор.

Интересна група (енг. *Community of Interest*) - група појединаца са заједничким интересом. Интересовање не мора да се односи на свакодневни

рада те особе или тренутни задатак. Интересне заједнице могу да деле идеје и комуницирају или сарађују.

Људски капитал (енг. *Human Capital*) - све особине које појединци у организацији поседују, као што су стручност, искуство, могућношћу, способност, креативност, прилагодљивост.

Репозиторијуми знања (енг. *Knowledge Repositories*) - колекције знања засноване на примерима добре праксе. Корисници прегледају њихов квалитет и исправност, и организовати их на начин који обезбеђује лакши приступ.

Систем лоцирања стручности (енг. *Expertise Locator System*) - систем који обухвата и категорише вештине и способности запослених од стране организације, региона, предмет експертизе или области интереса. Систем је користан за идентификацију и проналажење особе са стручношћу или знањем у датој области ради консултација.

Управљање знањем (енг. *Knowledge Management*) - промишљен процес за ефикасно управљање и примену интелектуалног капитала како би се омогућиле брже и боље организационе одлуке. Овај процес осигурава најбоље расположиве информације за доношење исправне одлуке.

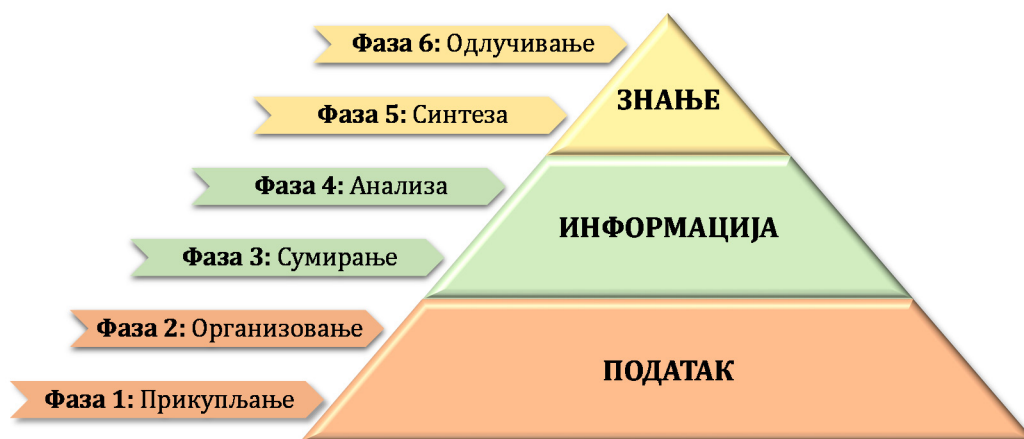
Шампион знања (енг. *Knowledge Champion*) - лице које развија заједничку визију, обезбеђује координацију, олакшавањем и охрабривањем оних који су укључени у напоран процес управљања знањем.

3.1.1. Хијерархија знања

Хијерархију знања дефинишу подаци, информације и знање. Подаци представљају запажања или чињенице извучене из контекста и стога су без директног смисла. Када се подаци ставе у неким смислени контекст, настају информације. Знање је оно у шта човек почиње да верује и вредност заснована на смислено организованој акумулацији информација (порука) искуством, комуникацијом или извођењем (Zack, 1999a).

Хијерархија знања се најчешће приказује помоћу тзв. пирамиде знања (енг. *Knowledge Pyramid*), дате на слици 3.1, која дефинише структурне и/или функционалне везе између података, информација, знања.

Rajeev Verma дели трансформацију података у знање у шест фаза које, када се изврше, применом комбинације техничких, организационих и људских вештина, податке претварају у информације, а информације у знање. Ова шема (слици 3.1), је универзална за све кориснике, али ресурси који се користе, могу бити јединствени за свако корисничко окружењу (Verma, 2015).



Слика 3.1. Пирамида знања³

Фаза 1: Прикупљање података - најважнији корак у процесу управљања знањем, чији је циљ да обезбеди поуздане и релевантне податке. Ако се прикупе неважни или нетачни подаци, добијено знање може бити непрецизно или погрешно. Последично, одлуке донете на основу таквог знања не могу бити ваљане. Прикупљање података треба да прати адекватна процедура која мора бити добро документована и стриктно поштована од стране људи који су укључени у процес прикупљања података. У овој фази дефинишу се и механизми екстракције података, као и њихово складиштење. За ту сврху користе се софтвери за обраду података и системи за управљање базама података.

Фаза 2: Организовање података - прикупљени подаци треба да буду организовани на основу дефинисаних правила. На пример, сви подаци везани за исту активност чувају се у истој табели базе података. Ако постоји много података у бази, техника нормализације може да се користи за организовање

³ Извор: <http://www.forwardmotion.eu/consulting/knowledge-management.php>

и смањење редунадансе, односно непотребног понављања података. На овај начин, подаци се логички распоређују и стварају се релације које омогућавају лакше претраживање. Када податак прође ову фазу, постаје информација.

Фаза 3: Сумирање информација - информације се сумирају на начин да се из њих извуче суштина. Информације се приказују табеларно или графички и чувају на одговарајући начин.

Фаза 4: Анализирање информација - информације се анализирају како би се пронашле релације, обрасци и редунадантност. Експертски тимови или искусни стручњаци у овој фази играју важну улогу. Након анализе, добијени резултати се презентују у облику извештаја.

Фаза 5: Синтеза знања – преломна тачка у којој информација постаје знање и дају се резултати анализе (обично извештаји). Образац или понашање једног ентитета може се применити да би се објаснио други ентитет. Знање се чува у бази знања и може му се приступити са било ког места путем интернета имплементацијом одговарајућих софтвера, који могу бити комерцијални или у отвореном приступу (енг. *open-source*).

Корак 6: Доношење одлука - знање се користи за одлучивање. Као пример, када се врши процена неког пројекта или задатка, може се користити знање стечено у раду на ранијим проценама. Ово убрзава процес процене и даје процени високу поузданост. Тако организовано управљање знањем побољшава пословање и дугорочно гледано, доноси корист у финансијском и културолошком погледу, као и у развоју људских ресурса.

3.1.2. Модел управљање знањем

Управљање знањем (енг. *Knowledge Management*) је активност која крајем XX века почиње да добија све већи замах. Широм света, научници се све више баве проучавањима у овој области, а стручњаци у пракси примењују стечена знања. Томе је допринео нагли развој информационих технологија. У прошлости се процес управљања знањем сводио на ручно складиштење и анализу информација. Са развојем информационих технологија тај процес је аутоматизован. Због

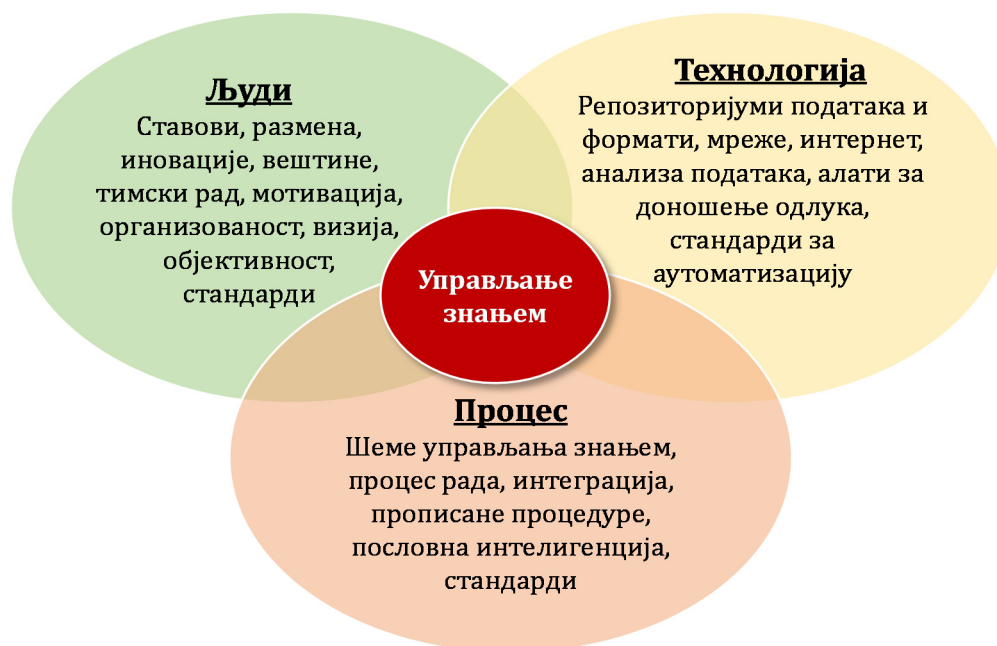
тога су складиштење, проналажење и дељење информација постали неупореди-во једноставнији и комфорнији. Систем управљања знањем је експлицитне информације учинио доступним кориснику коме су потребне, у тачно утврђеној форми и у време када је то кориснику потребно (McKenna, 2008). Са друге стране, овај систем је омогућио смањење грешака или њихово често понављање, смањење трошкова истраживања и развоја понављање истих, смањење трошкова истраживања и развоја, доношење квалитетнијих одлука, побољшање производног процеса, повећање ефикасности и самосталности радника као и њихову међусобну сарадњу.

Према једној од дефиниција, управљање знањем је промишљена и систематска координација људи, технологије, процеса и организационе структуре, како би се кроз поновну употребу знања и иновације, стварањем, разменом и применом знања, створила нова вредност (Dalkir, 2005, Evans *et al.*, 2014). Однос ових компонената управљања знањем приказан је на слици 3.2. То је процес којим предузеће систематски прикупља, организује, анализира и размењује знања која су релевантна за то пословно окружење и оперативну дисциплину. Многа предузећа се ангажују у области управљање знањем како би се осигурала од "губитка памћења" оперативних дисциплина које су од суштинског значаја за пословни опстанак, као и дисциплина које се протоком времена и флукуацијом људи лако могу заборавити. Управљање знањем је често у делокругу управљања квалитетом и унапређења пословних процеса. (Chambers, 2018)

Управљање знањем је планирање, организовање, мотивисање и контрола људи, процеса и система у организацији како би се осигурало да се њена интелектуална имовина (знања) побољша и ефикасно користи (King, 2009).

Према *British Standards Institution* управљање знањем је стварање и накнадно управљање окружењем које подстиче стварање, дељење, учење, унапређење, организовање и коришћење знања за добробит организације и њених клијената. Ова дефиниција претпоставља да се знањем не може управљати у традиционалном смислу, али да организација може оптимизовати вредност свог знања

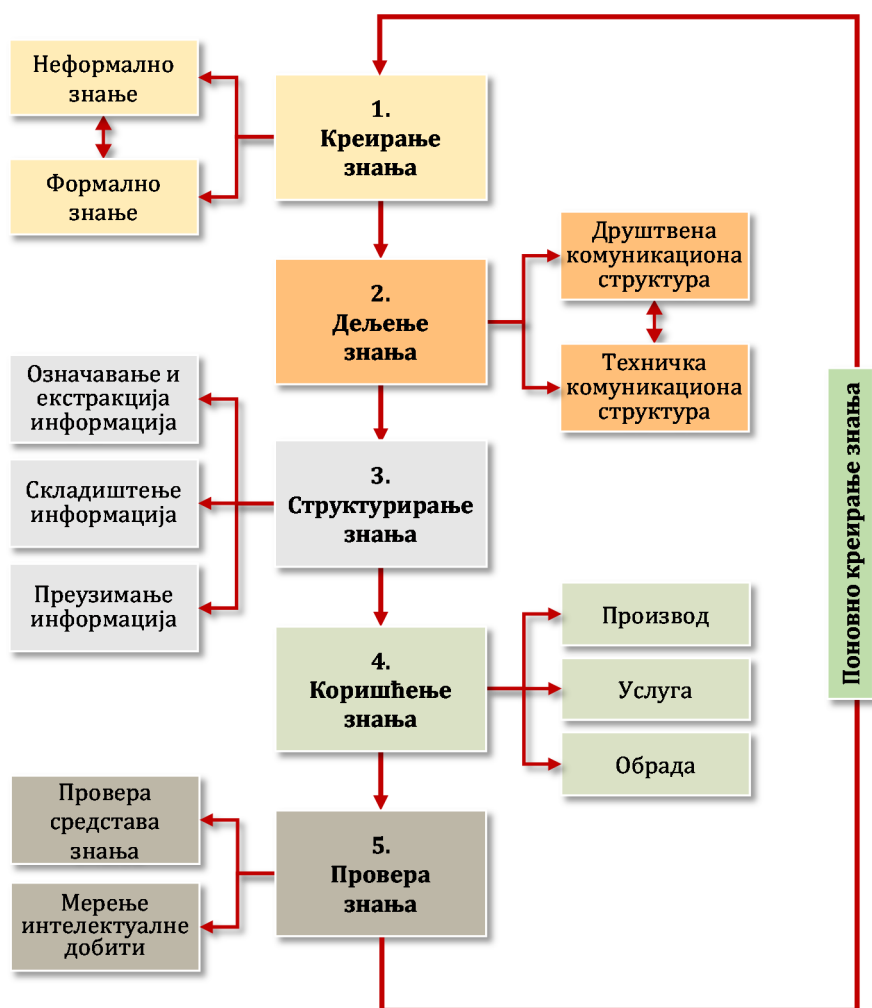
кроз одговарајући спој лидерства, вредности, културе, процеса, алата и вештине за подршку приступу и коришћењу знања (Girard & Girard, 2015). Начин на који се акумулирани интелектуални капитал мења и еволуира током времена је тада зависан од стратегије за управљање знањем у стварању, приступу и коришћењу знања (Young *et al.*, 2003).



Слика 3.2. Компоненте управљања знањем (Bhatt, 2003)

Управљање знањем је интегрисани системски приступ који, када се примењује у организацијама, омогућава оптимално коришћење благовремених, тачних и релевантних информација; такође олакшава проналажење знања и иновација, подстиче развој организације која учи и повећава могућност разумевања интеграције свих извора информација, као и појединачних и колективних знања и искустава (Blodgett *et al.*, 2005).

У процесу управљања знањем, користе се различите методе и алати за прикупљање информација, њихово организовање, складиштење, дељење и анализу. Анализирајући више модела, дошло се до закључка да су пет фаза карактеристичне за већину модела. То су: креирање, дељење, структурирање, коришћење и провера знања. Пример модела за управљање знањем приказан је на слици 3.3 (Sağsan, 2006).



Слика 3.3. Модел управљања знањем

Правилно анализирани информације се чувају као "знање" организације које ће се касније користити за активности као што су доношења одлука и обука нових чланова колектива. Данас, већина организација имају своје развојна окружења за управљање знањем. Они дефинишу начин прикупљања знања, технике и алате који се користе за прикупљање и складиштење информација, и алате, технике и механизме њихове анализе.

3.2. УПРАВЉАЊЕ ДОКУМЕНТИМА

Резултат пословних активности једног предузећа представља скуп докумената различитих типова и формата. То могу бити: пословна документација, пословна преписка, уговори, дозволе, пројектна документација и др. У рударству, текстуални документи увек су повезани са графичким документима (на пример: ситуационе карте, попречни профили и др., рађени у програмским пакетима AutoCAD или ArcGIS). У већини случајева документација је у папирном облику, неадекватно систематизована и сортирана, најчешће ускладиштена на начин који не обезбеђује трајну заштиту документа на начин који би обезбедио њихово безбедно чување и тражење потребних информација итд.

Чак и уколико је документација оригинално настала у дигиталном облику, она најчешће није повезана са постојећом недигитализованом документацијом. Све ово су разлози да тражење потребних информација постаје спор и често узалудан, посао који не даје жељене резултате.

Да би елиминисало наведене проблеме, савремено пословно окружење се окреће имплементацији система за праћење и управљање пословном документацијом. Ово је нарочито значајно за рударство, као веома сложену индустријски грану, која захтева заједнички рад великог броја стручњака различитих профила, који раде на различитим, често веома удаљеним локацијама. Ефикасан систем управљања документацијом омогућио би свим учесницима на пројекту нови приступ тимском раду који би олакшао приступ документима и њихово праћење у свим фазама, од анализе пројекта, преко истраживања лежишта, евалуације пројекта, пројектовања и производње, па до затварања рудника.

Документи у електронском облику, анотирани на адекватан начин, омогућавају брзо проналажење потребних информација, читавање, преузимање, измену која подразумева чување изворних верзија докумената али и свих његових верзија, уз одговарајућу размену између учесника на пројекту. Повезивање ових докумената са приказом на мапи може побољшати и убрзати процес планирања

и пројектовања и значајно допринети доношењу добрих пословних одлука. Њихова систематизација и имплементација система за претраживање представљају основу за увођење концепта електронског пословања.

3.2.1. Предности система за управљање пројектном документацијом

Предности система за управљање пројектном документацијом су бројне. Овде ће бити поменуће неке о значајнијих (Chaouni, 2015, Fenton, 2014, Poursaba, 2015, Rosson, 2017).

Архивирање докумената у папирној форми изискује складиштени простор. Такво чување докумената може бити скупо и небезбедно. Системи за управљање документима обезбеђују централизовано архивирање докумената у електронској форми, чиме се обезбеђује да сви документи буду на једном месту. Такви документи су лако доступна, једноставна за прегледање, мењање и размењивање. Њиховом употребом, трошкови архивирања се свде на минимум.

Одржавање докумената на једној локацији осигурава и њихово симултано ажурирање, а спречава и рад на више беспотребних копија докумената, чиме се чува интегритет и тачност података.

Праћење самог садржаја докумената дистрибуцијом шаблона, осигурава да сви садржаји буду објављени на јединствен начин у складу са наменом и врстом документа.

Осим форме, могуће је и усклађивање докумената са прописима, који могу бити прилично сложени, а неусаглашеност може довести до проблема у пословању: новчане казне, одузимање дозволе, а у неким случајевима може доћи и до кривичне одговорности.

Системи управљања документима обезбеђују висок степен заштите докумената. Она се постиже кроз ригорозне безбедносне мере и контролу приступа. Приступ документима се може контролисати на различитим нивоима, на пример: ко и када приступа документу, ко и када врши измене и

др. Све измене остављају траг, те се могу пратити, што онемогућава злоупотребе. У случају елементарних непогода или других непредвиђених околности, обезбеђује се да подаци, који су од виталног значаја за пословање, буду сачувани.

Приступ документима могућ је са било које локације у било ком тренутку, без обзира на уређај који се користи, што је посебно корисно када се чланови тима налазе на различитим локацијама, што је чест случај у рударству. Чување осетљивих информација у систему управљања документима, такође помаже да се повећа безбедност. Заштита докумената постиже се давањем различитих степена овлашћења корисницима за приступ одређеним документима.

Најзначајнија предност ових система је лако проналажење и преузимање докумената. Према неким анализама, запослени потроше 1,8 сати дневно, односно 9,3 сата недељно за тражење информација⁴. Основни задатак система за управљање документацијом јесте управо да спречи велики губитак радног времена. Могућност лако и брзог проналажења жељеног документа омогућава доношење добрих пословних одлука и смањује губитак времена за тражење информације.

Ефикасност претраживања обезбеђује се обележавањем докумената метаподацима и етикетама, као и њихово разврставање у колекције и подколекције. У зависности од примењеног решења, резултати претраживања могу бити: документи у целини или њихови делови. Системи могу бити прилагођени да помогну у филтрирању информација како би се олакшало организовање и лоцирање докумената.

Побројане предности система за управљање документима из области рударства значајне су за целокупан животни век једног пројекта, односно рудника с једне стране, а са друге стране значајне су због обезбеђивања подршке релевантних институција као што су министарство, Влада и сл.

⁴ <https://www.articlecube.com/research-shows-searching-information-work-wastes-time-and-money> (приступљено 1.03.2018.)

3.2.2. Општи модел система за управљање документима

Општи модел система за управљање документима прати животни циклус једног документа. Модел, приказан на слици 3.4, се састоји из више фаза (Saffady, 1996, Sprague, 1995), и то:

- *креирање* – израда аналогних или дигиталних докумената,
- *скенирање* – конверзија аналогних докумената у дигиталне превођењем у форму погодну за даљу дигиталну обраду,
- *индексирање* – ручно или аутоматско додељивања ознака документима на основу којих се могу једнозначно идентификовати, и каталогизација,
- *управљање* – складиштење докумената на начин који ће омогућити брз приступ,



Слика 3.4. Општи модел система за управљање документима

- *претраживање* – проналажење докумената или делова докумената, који одговарају постављеним захтевима израженим кроз упите, коришћењем класификација, метаподатака, претраге пуног текста, претраге по концептима, претраге по обрасцима, вишејезично, ...,
- *прегледање* – прегледање резултата претраживања,

- *администрација* – управљање документима, њиховом структуром, управљање корисничким налозима,
- *поновна обрада* – поновно коришћење докумената и управљање њиховим верзијама,
- *заједнички рад* – групни рад корисника на документу,
- *дељење* – извоз и дистрибуција докумената,
- *задржавање* – дефинисање приступа документима према унапред договореним правилима,
- *уништавање* – уништавање докумената након истека периода предвиђеног за његово чување по унапред дефинисаној процедури и
- *одржавање* – обезбеђивање дугорочног континуираног приступа документима.

Даљи поступак са документима зависи од конкретних захтева пословног процеса.

3.3. ДИГИТАЛНЕ БИБЛИОТЕКЕ

Зачетак идеје брзог и једноставног приступа информацијама, онога што данас знамо као "дигиталне библиотеке", везује се за период после Другог светског рата, када је др Ваневар Буш (*Vannevar Bush*) објавио чланак под насловом "*As We May Think*" (Bush, 1945). У раду је описан аутоматизовани сто *Memex*⁵ (слика 3.5), који је користио фотографију како би сачувао информације. Ово је био уређај са фантастичним, до тада никада виђеним и незамисливим перформансама (Cleveland, 1998). Због ово решења, Буш је често сматран за прву особу која је артикулисала нову визију дигиталне библиотеке, али то није било тачно јер је рад на *Memex*-у био заснован на ранијем раду објављеном у Немачкој пре Другог светског рата. Значај овог чланка био је у опису међусобних односа између информација и научних истраживања, и у опису потенцијала технологије да научницима омогући прикупљање, складиштење, проналажење и преузимање информација (Arms, 2000).



Слика 3.5. *Memex*⁶

Идеја Буша била је инспирација и подстрек научницима и проналазачима да започну развој информационих технологија. Са доласком рачунара, концепт развоја је био усмерен на велике базе података, системе претраживања и јавног приступа информацијама. Када су рачунари, повезани у велике мреже, формирали Интернет, концепт је наставио развој, а истраживања су се окренула ка стварању дигиталних библиотека којима је могао да приступи било ко из било ког дела света.

⁵ *Memex* – кованица настала од две речи: *memory* и *index*

⁶ <http://trevor.smith.name/memex/>, приступљено: 15.03.2018.

3.3.1. Дефинисање концепта дигиталне библиотеке

У литератури је познато мноштво дефиниција чија је разноликост последица неколико недоумица везаних за дигиталне библиотеке.

Прва је везана за сам назив. У употреби их је било неколико, као што су: електронска библиотека, виртуелна библиотека, библиотека без зидова (енг. *library without walls*), дигитална библиотека. Неки ове појмове сматрају синонимима, док други покушавају да дефинишу јасне разлике међу њима. Филип Бејкер (*Philip Barker*) каже да електронска библиотека фаворизује приступ документима путем рачунара; дигитална библиотека се састоји само од дигитализованих докумената и у њој нема више традиционалних књига, периодике, новина; виртуелна библиотека обухвата укупност чулног искуства (Barker, 1994). Никада, међутим, ни један од ових назива није прецизно дефинисан, али је данас најшире прихваћен и готово искључиво у употреби назив "дигитална библиотека".

Друга недоумица лежи у чињеници да су дигиталне библиотеке предмет интересовања многих области истраживања, те се и оне разликују у зависности од тога која истраживачка заједница их описује. На пример (Nürnberg *et al.*, 1995):

- са аспекта проналажење информације, дигитална библиотека је велика база података;
- са аспекта стручњака који се баве хипертекстом, дигитална библиотека је апликација за примену хипертекстуалних метода;
- са аспекта оних који раде у области публикавања информација, дигитална библиотека је веб апликација;
- за библиотекарство, то је још један корак у сталном процесу аутоматизације библиотеке.

А заправо, дигиталне библиотеке су све наведено и ови различити истраживачки приступи само додатно доприносе њиховој разноврсности и развоју.

Трећа недоумица произилази из чињенице да постоји различите колекције докумената на интернету које се називају дигиталним библиотекама, а које, са становишта класичног библиотекарства, то нису. На пример:

- за компјутерске научнике и програмере, збирке компјутерских алгоритама и софтверских програма су дигиталне библиотеке;
- за оне који раде у области публиковања информација, веб апликација су дигиталне библиотеке;
- за велике институције, дигиталне библиотеке су системи за управљање документацијом, који контролишу њихове пословне документе у електронској форми;
- за издаваче, то могу бити online каталози.

Најинтересантнији пример онога што многи људи сматрају дигиталном библиотеком данас је WWW. Веб је скуп великог броја докумената и многи га сматрају огромном дигиталном библиотеком јер могу пронаћи жељене информације, на исти начин као што могу обавити банкарске трансакције у "дигиталним банкама" или купити неки производ у "дигиталним продавницама." Па на дилему да ли је или не веб дигитална библиотека, Клифорд Линч (*Clifford Lynch*) каже: "Понекад се чује да је Интернет светска библиотека дигиталног доба. Интернет, а посебно његова колекција мултимедијалних ресурса, позната као WWW, није направљен да подржи организовано објављивање и проналажење информација, што је својствено библиотекама. Он је еволуирао у нешто што би се могло сматрати хаотичним колективним складиштем за производе светских дигиталних "штампарија".... Укратко, Нет није дигитална библиотека." (Lynch, 1997).

Најчешће цитирана дефиниција дигиталних библиотека је она коју је дао Вилијам Армс (*William Y. Arms*). Он сматра да су дигиталне библиотеке контролисане, системски организоване колекције информација са придруженим сервисима, ускладиштене у дигиталном формату, којима се приступа преко мреже. Заједничко свим дигиталним библиотекама је да су информације организоване на рачунарима и да су доступне корисницима преко

интернет мреже, са процедурама за избор информација, њихово организовање ради повећања доступности корисницима, и архивирање (Arms, 2000).

Федерација дигиталних библиотека (енг. *Digital Library Federation*) дефинише дигиталне библиотеке као "...организације које обезбеђују ресурсе и специјализовано особље које бира, структурира, интерпретира, дистрибуира, чува интегритет и обезбеђује приступ и трајност дигиталних библиотека, тако да су увек доступне корисницима" (Shiri, 2003).

Александра Тртовац каже да су дигиталне библиотеке велике колекције дигиталних објеката, који су у виду различитих дигиталних података (текст, слика, звук, видео, анимација) или њихових комбинација (мултимедија) похрањени на мрежи, описани различитим метаподацима и повезани са другим информационим сервисима. Крајњи корисници им могу приступати и користити их без временског и просторног ограничења, са могућношћу да сами креирају нове или ажурирају постојеће дигиталне објекте (Trtovac, 2016).

3.3.2. Архитектура дигиталних библиотека

Први захтев који лежи у основи сваке дигиталне библиотеке је развој и унапређење техничке архитектуре ради смештања дигиталних докумената. Архитектура укључује компоненте као што су:

- брзе локалне мреже и брзе везе са Интернетом,
- релационе базе података које подржавају различите дигиталне формате,
- машине за претрагу пуног текста које индексирају ресурсе и обезбеђују им приступ,
- различити сервери, комуникациони, веб и фајл сервери,
- функције електронског управљања документима које ће помоћи у укупном управљању дигиталним изворима.

Архитектура дигиталне библиотеке обухвата ресурсе:

- библиографске базе часописа и дигиталних ресурса
- индекси и ресурси за претрагу

- колекције показивача на ресурсе на вебу, директоријуме
- материјале у различитим дигиталним форматима: текст, фотографије, цртежи, нумерички подаци, формуле, звук, видео,...

3.3.3. Дигитални објекти

Основни елементи дигиталних библиотека су дигитални објекти који могу бити: аналогни (дигитализовани документи настали конвертовањем папирних или других облика докумената у дигиталну форму) и дигитални (документи који су настали као дигитални и само у том облику постоје).

Према врсти садржаја дигитални објекти могу бити различити: текстови, слике, аудио и видео записи, анимација или мултимедијални објекти настали комбинацијом различитих формата.

Најчешћи формати дигиталних објекта су:

- за текстове и слике: PDF (Portable Document Format), PDF/A (Portable Document Format Archival), TIFF (Tagged Image File Format), JPEG (Joint Photographic Experts Group), GIF (Graphic Interchange Format), PNG (Portable Network Graphics), PPT (Microsoft Power Point);
- за аудио записе: WAV (Waveform Audio File Format), WMA (Windows Media Audio), MPGE (Moving Picture Experts Group);
- видео записи: AVI (Audio Video Interleave), MOV (QuickTimeMovie).

Неке од важнијих карактеристика дигиталних објеката су: доступност, ажурирање, вишејезичност, стандардизација и технологија.

Доступност дигиталним објектима, с обзиром на то да се налазе на мрежи, под одређеним условима, омогућена је свима и у сваком тренутку. За крајње кориснике то значи могућност приступа документима без временског и просторног ограничења.

Ажурирање дигиталних објеката подразумева могућност једноставног исправљање техничких грешака или нетачно унетих података, допуњавање

докумената новим подацима, замена постојећих и додавање нових докумената. С обзиром на laku доступност документима и ажурирање је једноставније.

Вишејезичност – коришћење вишејезичних могућности дигиталних библиотека, а пре свега паралелног, или здруженог, претраживања, требало би да се повећа и приступ дигиталним документима, јер корисници желе да добију информације о одређеној теми из свих расположивих извора, независно од језика (Krstev, 2002). Претраживање информација код вишејезичких објеката захтева коришћење посебних електронских речника, тезауруса, али и семантичких мрежа.

Стандардизација - приликом успостављања стандарда морају се узети у обзир различите структуре података похрањених у дигиталним библиотекама, информатички гледано, а свакако и њихово повезивање са другим базама података, мрежама хипертекстуалних докумената и семантичким вебом. Коришћење стандарда је неопходно да би се омогућила и размена података између дигиталних библиотека што се нарочито односи на преузимање одређених дигиталних објеката и података о њима јер су саставни део двеју или више дигиталних библиотека. Преузимање скраћује време потребно за креирање дигиталне библиотеке, а тиме се избегава и поновна дигитализација истоветног аналогног садржаја.

Технологија - технолошка разноврсност докумената у дигиталним библиотекама је велика, а та особина дигиталних објеката омогућава и већу дистрибутивност њиховог садржаја. Дигитални објекти су описани у стандардним форматима, за њихову дигитализацију користе се стандардни формати. Честа је ситуација да су формати за опис компатибилни у различитим системима, па то омогућава и размену података међу системима. У оваквим случајевима, дистрибутивност не изостаје јер се дигитални објекти настали коришћењем различитих технологија могу користити у различитим системима и досегнути најшири круг корисника. Такође се на овај начин избегава вишеструка дигитализација истог аналогног објекта који је саставни део неколико дигиталних библиотека.

3.3.4. Метаподаци

Метаподаци су информације од централног значаја за развој дигиталних библиотека. Структурирани су тако да опишу, објасне, идентификују, лоцирају или на неки други начин олакшају преузимање, коришћење или управљање извором информација (Hodge, 2001). Веома често се за метаподатке каже да су то подаци о подацима или информације о информацијама. Гејл Хоџ посебно наглашава да су метаподаци кључ сигурности да ће описани извори остати доступни и у будућности.

3.3.4.1. Категорије метаподатака

Метаподаци могу бити (Trtovac, 2016):

- **дескриптивни** - описују ресурсе за потребе проналажења и идентификације и садрже основне елементе као што су: наслов, аутор, издавач, место, година, језик, јединствени идентификатор, опис, кључне речи, предметне одреднице, апстракт и сл.;
- **структурални** – описују структуру сложених ресурса (типове, верзије, везе између дигиталних објеката и друге особине, а омогућавају и описивање веза ка метаподацима надређених дигиталних објеката чији су саставни део. Ово значи да структурални метаподаци омогућавају да се повежу, на пример оригинална документа и све његове верзије, при чему се укључују информације о верзијама, последњим променама и сл.;
- **административни** - дају информације о употреби и управљању ресурсима везано за интелектуално право, очување, итд. Могу бити:
 - ♦ **метаподаци о правима коришћења** (енг. *rights metadata*) - дефинишу управљање правима приступа дигиталном објекту у складу са ауторским правима и заштитом интелектуалне својине;
 - ♦ **технички метаподаци** (енг. *technical metadata*) - доносе податке о томе када је извор настао (датум креирања), о техничким детаљима о извору, величини и врсти датотеке, приступу извору, податке о свим изменама, о опсегу, формату приказа;
 - ♦ **метаподаци о очувању дигиталног објекта** (енг. *preservation metadata*);

- ♦ **метаподаци о коришћењу** (енг. *use metadata*) односе се на активно праћење броја корисника који посећују и користе одређени садржај, као и праћење употребе садржаја дигиталног објекта у новом контексту или у новој верзији путем преузимања метаподатака и дигиталног објекта за другу дигиталну библиотеку.

3.3.4.2. Формати метаподатака

Постоји више формата, односно шема, за унос метаподатака за дигиталне објекте који чине неку дигиталну библиотеку. Неки од најчешће коришћених формата су:

- **Даблинско језгро** (енг. *Dublin Core*) - стандард за метаподатке, односно шема за унос метаподатака, а састоји се од сета елемената за описивање широког спектра;
- **TEI** (енг. *Text Encoding Initiative*) (TEI, 2017) - иницијатива за кодирање текста, софтверски независна метода која развија и одржава податке из хуманистичких наука (нарочито у литерарним и лингвистичким истраживањима) који се депонују у електронском облику. Овај формат биће детаљније обрађен у поглављу 6.5;
- **METS** (енг. *Metadata Encoding and Transmission Standard*) (METS, 2017) - намењен кодирању метаподатака дигиталних објеката и дигиталних библиотека, заснован је на XML платформи;
- **LOM** (енг. *Learning Object Metadata*) - стандард за метаподатке којим се описују наставна грађа, учила и слични дигитални објекти чија је основна намена да кориснику (ученику, студенту, полазнику курса) пружи подршку у процесу учења;
- **GILS** (енг. *Global Information Locator Service*) (GILS, 2018) - глобални сервис за лоцирање информација

3.3.4.3. Даблинско језгро

Назив потиче од места Даблин у Охају, где се 1995. године одржала иницијативна радионица под називом Metadata Workshop спонзорисана од стране OCLC (*Online Computer Library Center*) (OCLC, 2017) и NCSA (*The National Center for Supercomputing Applications*) (NCSA, 2017). Развојем Даблинског језгра руководи Даблинска иницијатива (енг. *Dublin Core Metadata Initiative*) (DCMI, 2017).

Даблинско језгро је стандард за метаподатке који се састоји од сета елемената за описивање широког спектра извора на мрежи и има за циља (Milenković, 2003):

- једноставност у креирању и одржавању како би сваки корисник могао да направи скуп описних исказа;
- разумљиву семантику како би се олакшала претрага преко глобалне мреже свима који имају потребу за информацијом;
- локализацију: оригинално је имплементиран у Енглеској, али постоје верзије које су написане за многе друге језике (српски, руски, кинески, фински, норвешки, јапански...).

Даблинско језгро чини 15 основних елемената, и то: наслов (dc:title), предмет (dc:subject), опис (dc:description), врста (dc:type), извор (dc:source), однос (dc:relation), покривеност (dc:coverage), аутор (dc:creator), издавач (dc:publisher), сарадник (dc:contributor), права (dc:rights), датум (dc:date), формат (dc:format), идентификатор (dc:identifier), језик (dc:language). Опис ових елемената дат је у табели 3.1.

Пример употребе Даблинског језгра за описивање дигиталног објекта дат је на слици 3.6.

Табела 3.1. Елементи Даблинског језгра (ISO 15836-1:2017, Milenković, 2003)

Елемент	Назив	Напомена
dc:title	наслов	наслов по коме је ресурс познат
dc:creator	аутор	ентитет (име особе, организације или сервиса) одговоран за прављење садржаја ресурса
dc:subject	предмет	кључне речи или фразе које описују садржај ресурса
dc:description	опис	опис садржаја ресурса може да садржи апстракт, садржај, репрезентацију графичког објекта, ...
dc:publisher	издавач	ентитет (име особе, организације или сервиса) заслужан за објављивање ресурса
dc:contributor	сарадник	ентитет (име особе, организације или сервиса) заслужан за давање доприноса садржају ресурса
dc:date	датум	датум је од значаја за неки догађај из животног циклуса ресурса, обично је то датум креирања ресурса
dc:type	тип	природа садржаја ресурса; препоручује се коришћење контролисаног речника
dc:format	формат	физичка или дигитална манифестација објекта, садржи врсту медија или величину ресурса
dc:identifier	идентификатор	недвосмислена референца на ресурс, вредност додељена према договореним правилима, формални идентификациони систем садржи URI, DOI и/или ISBN
dc:source	извор	реферише на објекат из којег је садашњи објекат настао
dc:language	језик	језик садржаја ресурса, препоручује се коришћење контролисаног речника, као што је RFC 4646
dc:relation	веза	реч или број, преко којих се може приступити сродним ресурсима, у вези је са идентификатором
dc:coverage	покривеност	просторне референце дефинисана координатама могућа је употреба контролисаног речника као што је, на пример Getty Thesaurus of Geographic Names [TGN]
dc:rights	права	информације о интелектуалној својини ресурса, на пример изјава о имовинским правима укључујући права интелектуалне својине


```

<?xml version="1.0"?><!DOCTYPE rdf:RDF PUBLIC "-//DUBLIN CORE//DCMES DTD 2002/07/31//EN"
"http://dublincore.org/documents/2002/07/31/dcmes-xml/dcmes-xml-dtd.dtd">
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/">

<rdf:Description rdf:about="http://saska.rgf.rs/admin/items/show/284">
<dc:title>Управљање квалитетом угља</dc:title>
<dc:subject>угаљ</dc:subject>
<dc:subject>квалитет</dc:subject>
<dc:subject>хомогенизација</dc:subject>
<dc:subject>површински коп</dc:subject>
<dc:subject>Тамнава</dc:subject>
<dc:description>Ова монографија је настала као резултат истраживања по пројекту &quot;
Управљање процесом хомогенизације угља у циљу повећања искоришћења нискоквалитетних угљева
и уштеде мазута у термоелектранама&quot; који је финансирало Министарство за науку и
заштиту животне средине Републике Србије. Истраживања по овом пројекту су обављена током
2005. године, па се сва сазнања везана за практичну примену система управљања на коповима
Тамнава односе на стање у том периоду. Монографија је штампана уз финансијску помоћ
Министарства за науку и заштиту животне средине Србије, Електропривреде Србије и
Термоелектрана &quot;Никола Тесла&quot;.</dc:description>
<dc:creator>Игњатовић др Драган</dc:creator>
<dc:creator>Кнежевић др Динко</dc:creator>
<dc:creator>Колова др Божо</dc:creator>
<dc:creator>Лилић др Никола</dc:creator>
<dc:creator>Станковић др Ранка</dc:creator>
<dc:publisher>Рударско-геолошки факултет Универзитета у Београду</dc:publisher>
<dc:date>2017-02-12</dc:date>

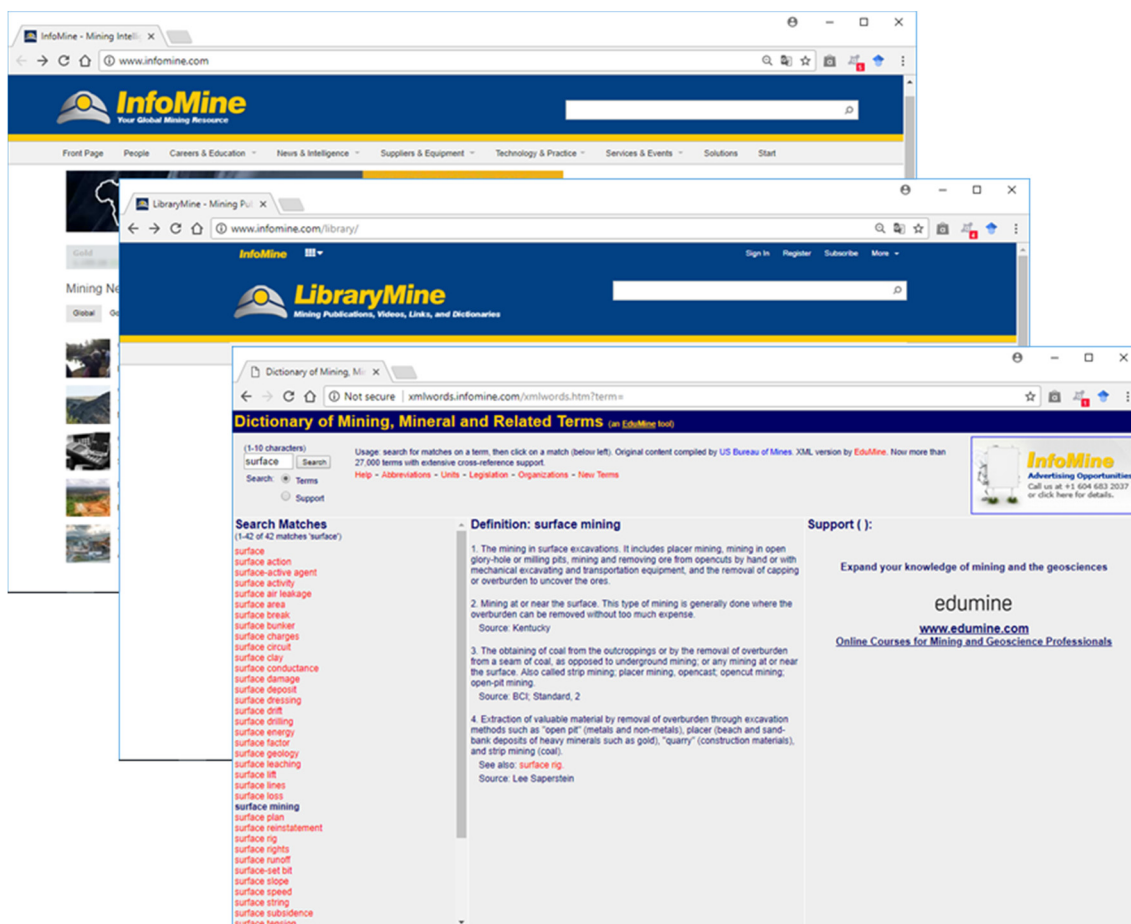
```

Слика 3.6. Пример употребе Даблинског језгра за описивање дигиталних објеката

3.3.5. Примери дигиталних библиотека из домена рударства

Дигиталних библиотека LibraryMine (LibraryMine, 2017) је део InfoMine, једног најзначајнијих сајтова који пружа информације из области рударства. Слободно се може рећи да је она рудник информација о глобалној рударској индустрији. Намењена је како рударским компанијама, тако и финансијерима, инжењерима, наставницима. Ова дигитална библиотека прикупља, складишти и размењује информација. Све информације доступне су путем веб страна, електронске поште, на друштвеним мрежама, на конференцијама и скуповима. Садржи каталоге публикација, корисних линкова, аудио-визуелних садржаја и речника. Рударски речник садржи више од 27.000 појмова и њихових дефиниција и описа са екстензивним унакрсним референцирањем. Изглед веб стране дат је на слици 3.7.

Дигитална библиотека са садржајем из домена рударства је и библиотека универзитета Penn State, САД, приказана на слици 3.8. Колекција *Geography and Earth Sciences* садржи подколекцију *Mining and Mineral Resources* (PSUL, 2017) са књигама, енциклопедијама, приручницима, стандардима, техничким извештајима и др.

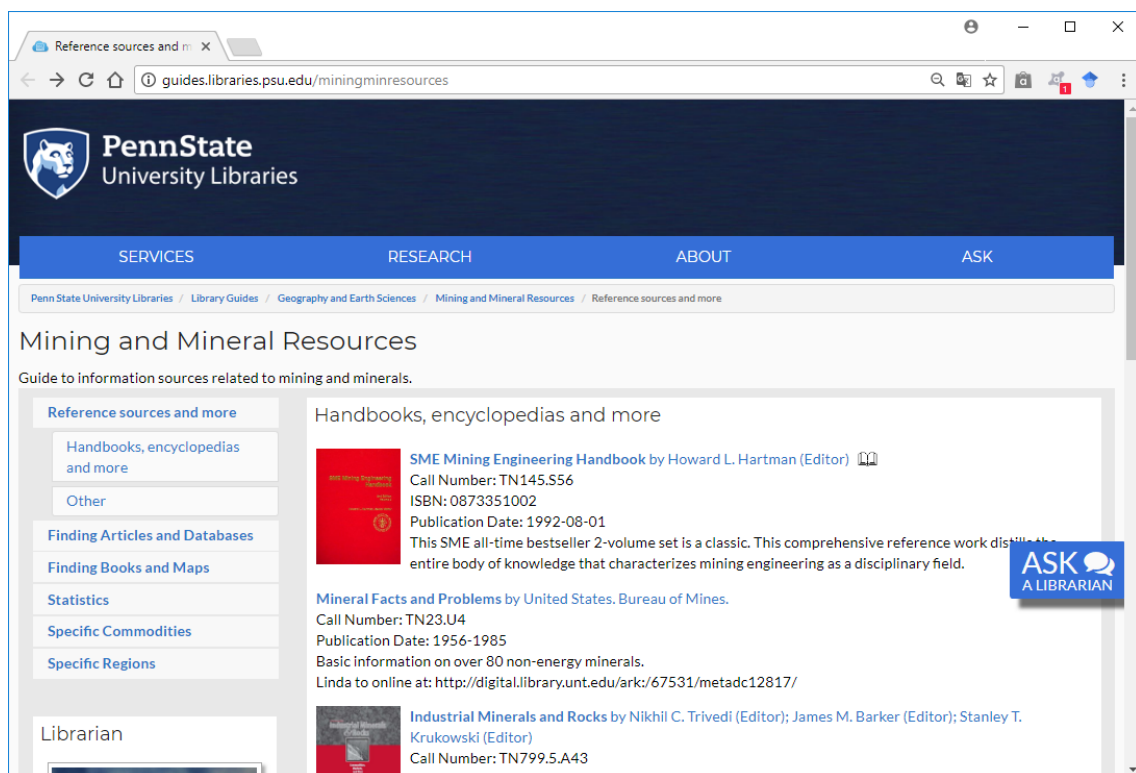


Слика 3.7. Дигиталних библиотека LibraryMine

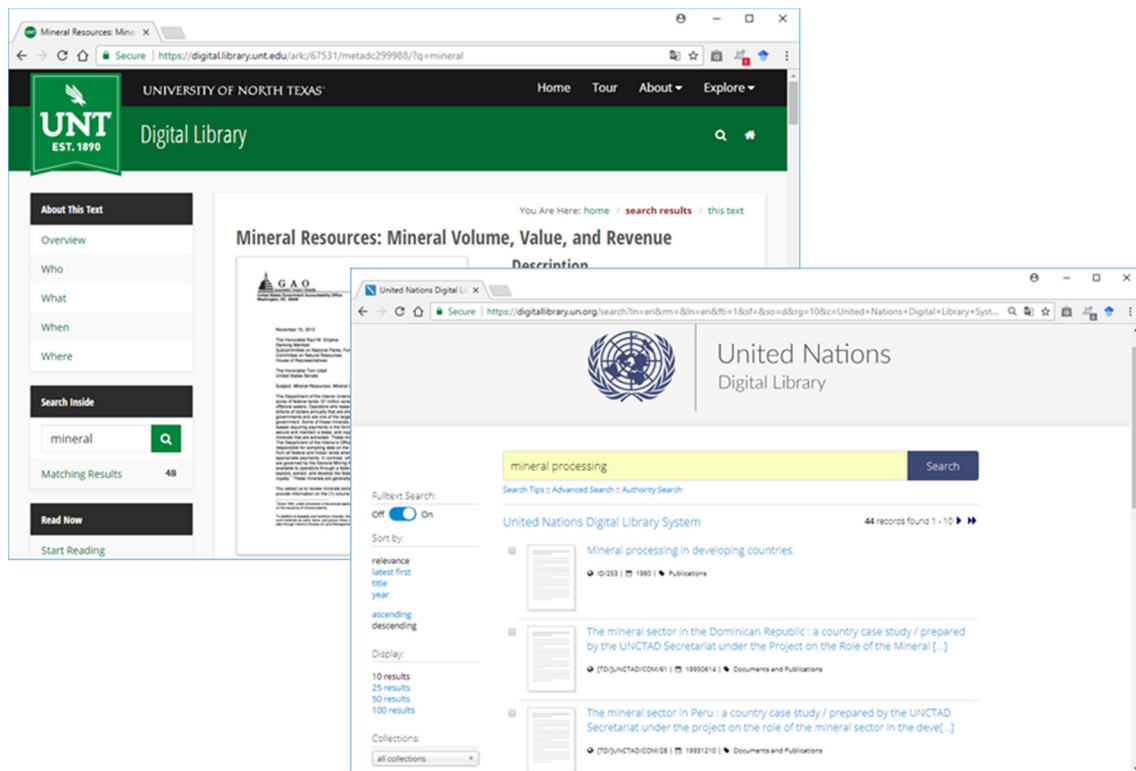
Поред поменутих дигиталних библиотека са колекцијама докумената из домена рударства, могу се споменути и: *United Nations Digital Library* (UNDL, 2017) и *University of North Texas Digital Library* (OCLC, 2017). Приказ ових веб страна дата је на слици 3.9.

Дугорочни циљ ове докторске дисертације је да се направи једна оваква дигитална библиотека на српском језику, која би обухватила што већи корпус знања из рударства (Томашевић *et al.*, 2017). Тренутно постоји дигитална библиотека "Подземни радови" (УМЕ, 2018), која прикупља научне и стручне радове из рударства и саставни је део двојезичке дигиталне библиотеке "Библиша"⁷.

⁷ <http://jerteh.rs/bibliша/ListaDokumenata.aspx?CID=2&lng=en>



Слика 3.8. Дигитална библиотека универзитета *Penn State*



Слика 3.9. Дигитална библиотека *United Nations* и *University of North Texas*

3.3.6. Дигитализација

Једна од основних метода креирања дигиталних објеката је дигитализација. То је конверзија аналогних медија, као што су књиге, чланци, фотографије или слике, у дигитални, електронски облик.

Процесом дигитализације могу се обухватити (Cleveland, 1998):

- сви дигитални објекти унутар дигиталне колекције,
- одређени делови колекције,
- карактеристична докумената којима се репрезентује специфичност колекције,
- најчешће тражени документи чиме се повећава њихова доступност или
- насумично одабрани документ.

Ови приступи се могу применити појединачно или у комбинацији, зависно од циљева дигитализације, при чему се у обзир узимају многи фактори, као на пример:

- потенцијал документа за дуготрајну употребу,
- његова интелектуална и културолошка вредност,
- ограничења везана за ауторска права и др.

Кључни кораци у процесу дигитализације реализују се кроз активности које укључују, али се не ограничавају на (Ministarstvo kulture i informisanja, 2017):

- процедуре у процесу стварања дигиталних објеката;
- одређивање приоритета и опсега дигитализације;
- селекцију и припрему грађе за дигитализацију;
- успостављање система метаподатака на нивоу установе;
- стварање дигиталних докумената;
- стварање, архивирање и управљање дигиталним објектима;
- проверу квалитета дигитализоване грађе;
- видљивост и доступност дигиталних објеката, збирки и библиотека.

3.3.7. Омека као платформа за креирање дигиталне библиотеке

За израду дигиталне библиотеке RОmeка@RGF коришћена је Омека, веб платформа за приказивање дигиталних колекција и систем за управљање њиховим садржајем⁸. За њен развој заслужан је истраживачки тим Центра за историју и нове медије Рој Розенцвајг (енг. *Roy Rosenzweig Center for History and New Media* - RRCHNM) са Универзитета Џорџ Мејсон (*George Mason*) у Вирџинији. Припада групи софтвера отвореног кода (енг. *Open Source Software*), под лиценцом GPL v3.0 (енг. *General Public License*), што значи да је изворни код јавно доступан, те се платформа може побољшавати и прилагођавати потребама корисника. Изворно је осмишљена за научне институције које изучавају културно наслеђе, али данас је користе истраживачи многих других научних области. Налази се на самом врху листе сродних софтвера због чињенице да је флексибилна и једноставна за коришћење. Карактерише је атрактиван и прилагодљив визуелни дизајн, једноставна инсталација, могућност проширивања која дозвољава измену постојећих и додавање нових функционалности, флексибилан приступ метаподацима, подршка за веб стандарде (CSS, HTML, RSS), увоз и извоз података у стандардизованим форматима (RDF (RDF, 2017), CSV, XML, JSON) (Kucsma *et al.*, 2010). Није осмишљена за ИТ-стручњаке и не захтева специфична информатичка знања, што корисницима омогућава да се фокусирају на садржај дигиталне библиотеке и њено описивање и тумачење, а не на програмирање. Има уграђене функције за каталогизацију и представљање дигиталних објеката, засноване на Даблинском језгру, којим се обезбеђује стандардизовано описивање и организовање дигиталних објеката.

3.3.7.1. Развојно окружење Омеке

Постоје две верзије веб платформе Омека, и то (Томашевић *et al.*, 2017):

- Скраћена (енг. *lite*) верзија (Омека.net), која не захтева сопствени сервер, капацитет простора за складиштење датотека ограничен је на 500 MB,

⁸ Content Management System (CMS)

величина датотеке ограничена је на 64 MB, број расположивих додатака (енг. *plugins*) ограничен је на 15, нема могућност функционалног прилагођавања и могућност измене изгледа је ограничена. Проширивање платформе је могуће, али захтева финансијска средства која зависе од врсте корисничког пакета.

- Пуна верзија (Omeka.org) – може се инсталирати на локални диск или као виртуелна машина и има могућност потпуног функционалног прилагођавања.

Пре покретања инсталације Омеке, извршене су одговарајуће припреме на серверу, које подразумевају инсталирање: веб (HTTP) сервера Apache, система за управљање базама података MySQL (верзија 5.0 или новија) и интерпретатора програмског језика PHP (верзија 5.3.2 или новија). Дистрибуције оперативног система Linux, на којима Омека стабилно ради, су: Fedora, OpenSuse и Ubuntu. ROmeka@RGF је инсталирана на виртуелној машини са оперативним системом Ubuntu верзија 15.10.

Инсталирање платформе је започело креирањем MySQL базе са додељеним администраторским привилегијама. Најновија верзије Омеке, преузета са званичног сајта, распакована је, и добијени директоријум (у даљем тексту: *omeka-root*) смештен је у одговарајући директоријум веб сервера. Коришћење система за управљање базама података MySQL је омогућено изменама у датотеци *omeka-root/db.ini* (вредности поља *database host*, *username*, *password* и *database name*, односно адреса рачунара, корисничко име, лозинка и име базе података). Сви директоријуми, у којима се налази Омека, су добили дозволу за уписивање. После завршетка инсталације кориговане су привилегије. Инсталација је покренута путем веб читача уношењем IP адресе или домена на коме се налази дигитална библиотека, где се захтева дефинисање администраторског налога и назива сајта.

Радно окружење веб платформе је вишејезично и преведено је, потпуно или делимично, на 50 језика. Српска верзија је једна од десет потпуно преведених.

3.3.7.2. Програмски додаци Омеке

Изглед веб платформе може се мењати и прилагођавати избором неке од петнаестак понуђених тема или креирањем сопствене теме, а проширивање функционалности омогућено је применом додатака (*plugins*). Развијено је 90 додатака, прилагођених различитим верзијама Омеке. Корисници Омеке развили су преко 300 додатака, али како су углавном рађени за старије верзије Омеке, захтевају додатна прилагођавања најновијој инсталираној верзији платформе. У зависности од намене, додаци се могу сврстати у неколико група и то: за масовно креирање колекција и објеката, за организовање садржаја, за преглед датотека, за описивање доприноса заједнице, за геопросторну обраду и приказ на мапи. Овде ћемо приказати неколико додатака који су коришћени приликом израде дигиталне библиотеке RОmeка@RGF или који су корисни и важни за сваког корисника Омеке (Томашевић *et al.*, 2017).

Archive Repertory омогућава да се увезене датотеке на серверу чувају са изворним именима, као и да се групишу у хијерархијску структуру, најпре по објектима, а потом по колекцијама у којима се објекти налазе. Задржавање изворних имена чини URL-адресе датотека читљивијим, а уједно олакшава руковање датотекама.

Bulk Metadata Editor омогућава брзо и једноставно претраживање и измену метаподатака за велики број објеката истовремено. У првом кораку врши се избор објеката према различитим критеријумима (према тим критеријумима, могу се бирати: сви објекти дигиталне библиотеке, објекти у одабраној колекцији или објекти чији метаподаци задовољавају један или више задатих критеријума). У другом кораку бира се метаподатак који ће бити измењен. У трећем, завршном кораку, врши се сама измена метаподатка (та измена може подразумевати: претраживање и замену текста, додавање новог метаподатка у одабрано поље, додавање текста у постојеће метаподатке, уклањање дупликата и празних поља у одабраном опису објекта, уклањање дуплираних датотека у одабраним објектима или брисање свих постојећих метаподатака у одабраним пољима).

Catalog Search омогућава претраживање других каталога коришћењем поља dc: subject. Понуђени каталози су: Archive Grid, Digital Public Library of America, Google Books, Google Scholar, Nathi Trust, JSTOR, Library of Congress, WorldCat. Дата је и могућност додавања линкове ка каталозима других институција.

COinS уграђује метаподатке о цитирању за сваки објекат (*item*) у веб странице дигиталне библиотеке. Када је активиран, *COinS* омогућава видљивост објеката на *on-line* платформама, као што је Zotero⁹, аутоматским уграђивањем метаподатака о цитатима у стране других веб сајтова. Такође, у библиотеку платформе Zotero могу се додати појединачни објекти било ког Омека сајта, док се додавање више објеката истовремено имплементира скриптама које раде у позадини. Додатак *COinS* олакшава истраживање и поспешује компатибилност са другим системима.

Collection Tree обезбеђује визуални приказ хијерархијске структуре колекција у дигиталној библиотеци, што олакшава њихово прелиставање.

CSV Import омогућава масован увоз метаподатака, ознака и датотека, приказаних табеларно у формату CSV. Уколико су називи колоне усаглашени са Даблинским језгром, мапирање података врши се аутоматски; у супротном је неопходно податке мапирати ручно.

Drop Box омогућава администратору једноставнији увоз датотека које се налазе на серверу, у каталогу */plugins/Dropbox/files*. Преко административног интерфејса датотеке се увозе појединачно или масовно одабиром са понуђеног списка.

Dublin Core Extended проширује листу метаподатака Даблинског језгра, чиме се обезбеђује потпуна анотација објеката. На 15 основних елемената (*title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage, rights*) додаје скуп од 40 додатних елемената и то: *abstract, access rights, accrual method, accrual periodicity, accrual policy,*

⁹ Zotero је некомерцијални софтвер намењен за прикупљање, уређивање и управљање библиографским белешкама и аутоматизовано форматирање библиографских референци. Као и Омека, развијен је у „Центру за историју и нове медије Рој Розенцвајг“, <https://www.zotero.org/>

alternative title, audience, date available, bibliographic citation, conforms to, date created, date accepted, date copyrighted, date submitted, audience education level, extent, has format, has part, has version, instructional method, is format of, is part of, is referenced by, is replaced by, is required by, date issued, is version of, license, mediator, medium, date modified, provenance, references, replaces, requires, rights holder, spatial coverage, table of contents, temporal coverage, date valid.

Geolocation омогућава да се информације о локацијама у вези са дигиталним објектима додају на географску карту, уз могућност њиховог претраживања.

Item Relations омогућава креирање релације између дигиталних објеката. Додатак следи RDF модел дефинисања односа између објеката у виду RDF графа сачињеног од RDF тројки: субјекат-предикат-објекат. На пример, уколико је један дигитални објекат (RDF субјекат) део неког другог дигиталног објекта (RDF објекат), између њих се успоставља веза *isPartOf* (RDF предикат). Слично, ако је један дигитални објекат верзија неког документа, између тих дигиталних објеката се успоставља веза *isVersionOf*. На овај начин формирају се RDF тројке, које омогућавају касније истраживање текста коришћењем техника семантичког веба.

Hide Elements омогућава избор метаподатака који ће бити прикривени на форми за унос, на веб страни администратора или на јавној доступној веб страни, као и на форми за претраживање метаподатака.

METS Export омогућава извоз дигиталних објеката као METS XML датотека (METS, 2017), појединачних докумената, колекција или дигиталне библиотеке у целини. Подржава га Иницијатива Федерације дигиталних библиотека (енг. *Digital Library Federation*) која предлаже XML шему метаподатака за управљање објектима дигиталне библиотеке и за размену тих објеката међу репозиторијумима или између репозиторијума и корисника. Има нарочиту улогу у окупљању и одржавању докумената који чине један дигитални објекат, с обзиром на њихов број и разноврсност. *METS Export* повезује више дигиталних докумената и омогућава навигацију кроз њих. Такође, укључује техничке информације неопходне за управљање дигиталним објектима: формат документа, технолошке карактеристике, начин скенирања, дигиталне

трансформације. *METS Export* не прописује обавезни скуп описних метаподатака који ће се унети за дигитални документ, већ креатору метаподатака оставља одлуку о томе које ће описне метаподатке у ту сврху унети. Описни метаподаци за *METS* се могу на једноставан начин преузети из записа у Даблинском језгру (Trtovac, 2016).

Neatline, *NeatlineFeatures*, *NeatlineSmile*, *NeatlineText*, *NeatlineTime* и *Neatline Waypoints* представљају серију додатака који омогућавају повезивање просторних и временских тачака на мапи са објектима у дигиталној библиотеци, као и повезивање докумената са *Neatline* изложбом. Ови додаци нису активирани у дигиталној библиотеци *ROMeka@RGF* због некомпатибилности са додатком *Geolocation*, који је за инжењерске потребе далеко значајнији.

OAI-PMH Harvester прикупља метаподатке од *OAI-PMH* (енг. *Open Archives Initiative Protocol for Metadata Harvesting*) добављача података, мапира их у локалну базу и увози. Може се позивати једнократно или перманентно за ажурирање и синхронизацију дистрибуираних система. Тренутно може да увезе формате Даблинско језгро и *METS*.

OAI-PMH Repository припрема објекте за размену, представљајући инверзну функцију претходном додатку. Подржава Даблинско језгро, *MODS* и *METS*.

PDF Text омогућава оптичко препознавање карактера текста (енг. *Optical character recognition - OCR*), његову екстракцију из *PDF* формата и претраживање. Уколико текст није прочитан на задовољавајући начин, дата је могућност његове корекције или увоз новог текстуалног документа у предвиђено поље.

Reference додаје стране са абecedним, односно азбучним, индексом унапред дефинисаних елемената на којима је могуће прелиставање по унапред задатим метаподацима.

Search By Metadata омогућава администратору да дефинише метаподатке за напредно претраживање на *HTML* страни.

SimplePages омогућава администратору да креира динамичке *PHP* странице не захтевајући притом специфична информатичка знања.

SimpleVocab и *SimpleVocabPlus* омогућавају креирање контролисаних речника и њихово синхронизовање на облаку. У дигиталној библиотеци ROMEKA@RGF креиран је контролисани речник аутора, чиме су обезбеђени доследан унос и лакше претраживање.

TEI Display претвара (енг. *render*) постављену TEI датотеку у визуелно јасан облик. Подразумевана XSLT трансформација омогућава два типа приказивања: приказује се или цео документ или његове појединачне целине. Први подразумева да се цео документ трансформише у HTML док други приказује садржај документа (*div1* или *div2*), што је погодно за веће документе. Начин приказивања и XSLT трансформације се могу прилагодити, док се метаподаци из TEI заглавља могу аутоматски мапирати у поља Даблинског језгра за објекте и датотеке.

3.3.7.3. Креирање дигиталне библиотеке

Основне елементе дигиталне библиотеке у Омеки чине:

- објекти (енг. *items*¹⁰) са придруженим документима и ознакама,
- колекције (енг. *collections*) и
- веб стране.

Дигитална библиотека може садржати неограничен број објеката, докумената, ознака и колекција. Ограничења у броју елемената нема. Једино ограничење је да један објекат може бити придружен искључиво једној колекцији. Дијаграм елемената дигиталне библиотеке у Омеки приказан је на слици 3.10. За сваки од поменутих елемената дигиталне библиотеке је могућа потпуна контрола видљивости на вебу, почевши од појединачних метаподатака, па до сваког елемента у целини.

Веб платформа Омека дизајнирана је за приказивање објеката (Lee, 2014), који су основни елемент сваке дигиталне библиотеке. Због тога, креирање дигиталне библиотеке управо и започиње креирањем објеката.

¹⁰ Израз се често преводи као: ставка, архива, извор или ресурс.



Слика 3.10. Дијаграм елемената дигиталне библиотеке у Омеки¹¹

Према врсти, објекти могу бити различити. Кориснику је понуђена листа од 15 основних врста објеката, уз могућност додавања нових врста, у складу са потребама. Доступне врсте објеката су:

- покретне слике (*Moving Image*), све форме видео записа: анимације, филмови, телевизијски програми;
- звучни записи (*Sound*), све форме аудио записа: аудио компакт дискови, снимљени говор или звукови;
- усмена предања (*Oral History*), информације добијене у интервјуима са особама које поседују знање из прве руке;
- статичне слике (*Still Image*), визуелни приказ текстова, слика, цртежа, графичког дизајна, планова и карата);
- веб стране (*Website*), HTML странице и повезане слике, аудио и видео датотеке, итд.;
- догађаји (*Event*), временски ограничене појаве, на пример: изложба, веб конференција, радионица, чајанка, пожар, битка, суђење, свадба);
- електронска пошта (*Email*), текстуалне поруке са необавезним прилогом, које једна особа шаље другој особи или другим особама;
- план наставе (*Lesson Plan*), детаљан опис тока наставног процеса на курсу;
- особа (*Person*);
- интерактивни ресурси (*Interactive Resource*), веб странице, мултимедијални наставни објекти, сервиси за ћаскање;

¹¹ https://omeka.org/codex/Managing_Items, приступљено 19. маја 2017. године

- скупови података (*Dataset*), кодирани подаци у дефинисаној структури: листе, табеле и базе података;
- физички објекти (*Physical Object*), неживи, тродимензионални чврсти објекти, које у дигиталним библиотекама репрезентују типови као што су: покретне слике или статичне слике и др.;
- сервиси (*Service*), на пример: сервис за фотокопирање, банкарски сервис, међубиблиотечка размена или веб сервери;
- софтвери (*Software*);
- хипервезе (*Hyperlink*), веза или референца ка другом ресурсу на Интернету.

Објекти су скуп:

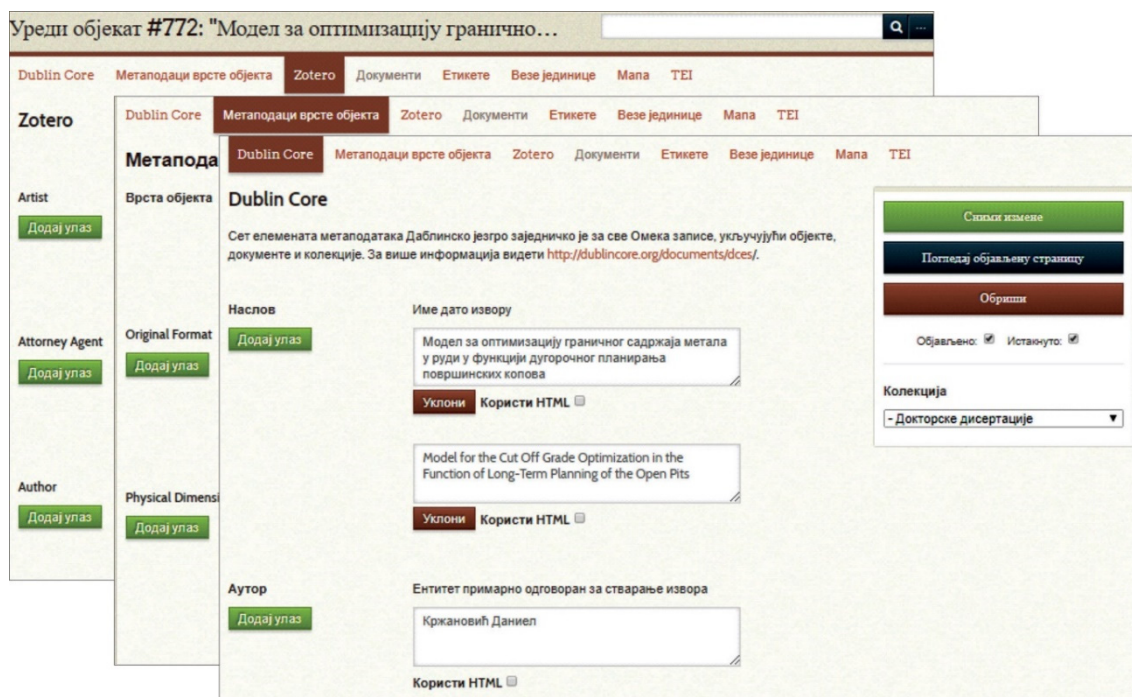
- метаподатака из Даблинског језгра који описују сам дигитални објекат,
- метаподатака који описују врсту објекта (*item type metadata*),
- етикета или ознака (*tags*) и
- докумената.

Објекти се најпре описује метаподацима из проширеног Даблинског језгра. Врста објекта се, такође, описује метаподацима, који се разликују у зависности од врсте. На пример, уколико је објекат "усмено предање", тада су метаподаци који га описују: особа која врши интервју, интервјуисана особа, локација, транскрипција, трајање интервјуа итд. Уколико је објекат "веб страница", тада се уноси само URL, а уколико је објекат "особа", одговарајући метаподаци су: датум рођења, место рођења, датум смрти, занимање, биографија, библиографија итд. Објекат је могуће описати и метаподацима који ће га учинити видљивим у оквиру платформе Зотеро (слика 3.11).

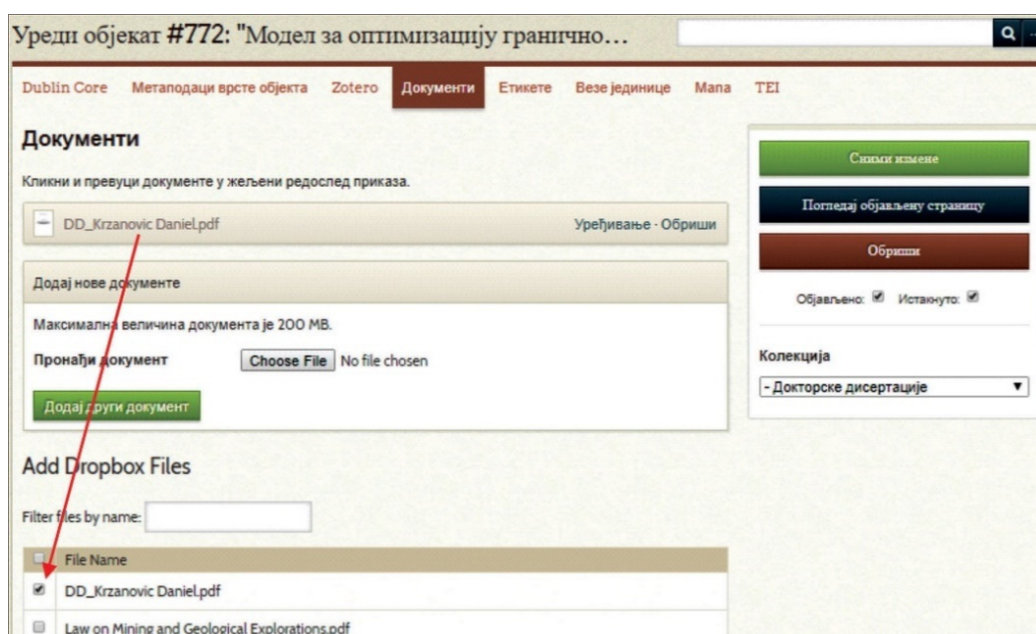
Документи који се придружују објектима могу се увозити појединачно и масовно, било помоћу додатка *Drop Box* (слика 3.12) било увозом из датотеке у формату CSV. Документи могу бити различитих формата, а неки од најчешће коришћених су:

- за текст: txt, css, csv, rtf, rtx, doc, docx, pdf, pps, ppt, pptx;
- за табеле: xls,xlsx;
- за базе података: mdb;

- за слике: bmp, gif, jpeg, jpg, tiff, png;
- за видео записе: avi, divx, mpeg, mov, mp4;
- за аудио записе: mp3, mid, midi, wav, wma;
- за извршне датотеке: exe, zip.



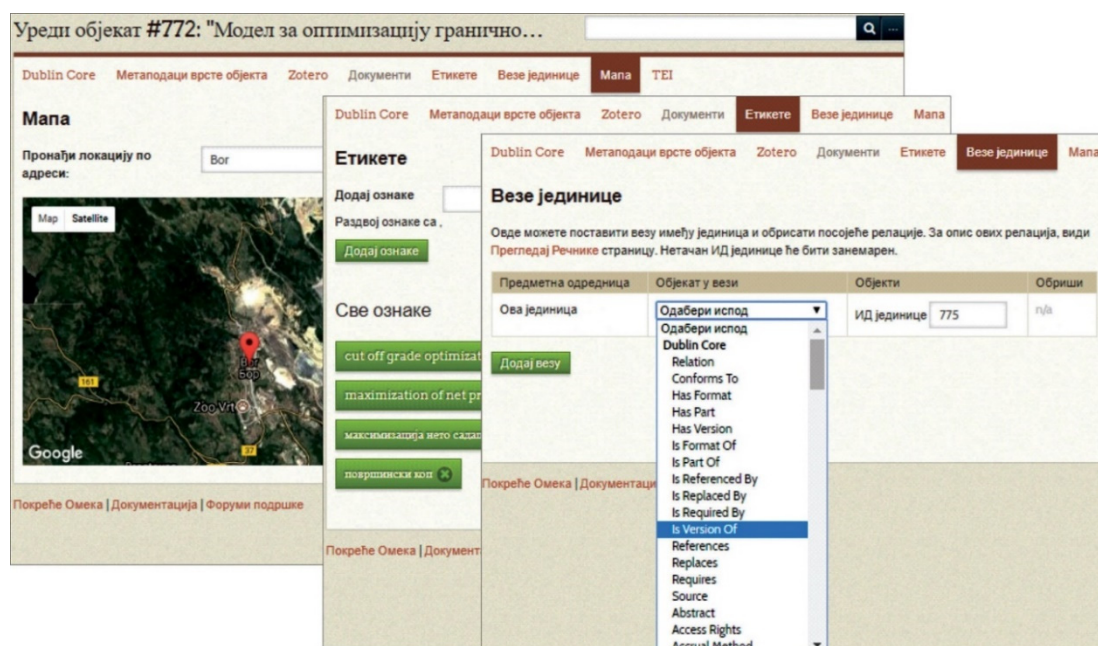
Слика 3.11. Описивање објекта метаподацима



Слика 3.12. Панел за увоз докумената

Сваком објекту додељене су етикете или ознаке (*tags*)¹² хијерархијски неструтуриране кључне речи или фразе које класификују садржај тако да се он касније може лако пронаћи. Слика 3.13 приказује панеле за унос геопросторних одредница, додељивање ознака и успостављања релација између објеката.

На панелу *Мапа* додају се информације о локацији која је у вези са објектом, уз могућност претраживања објеката према локацији.



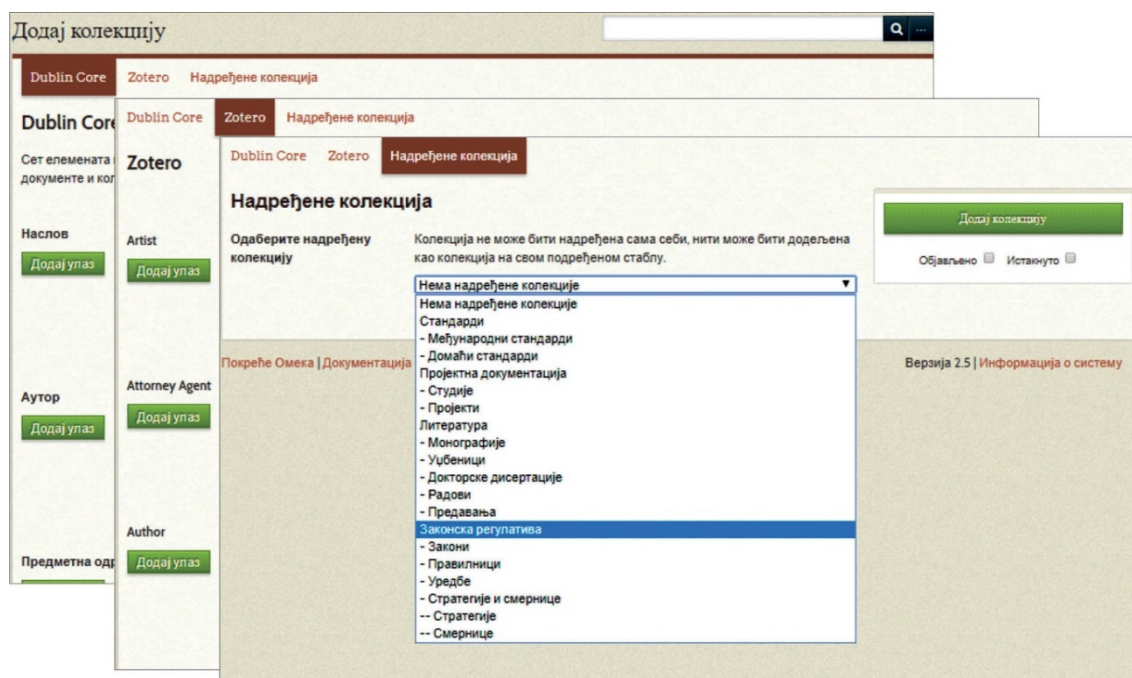
Слика 3.13. Унос геопросторних одредница, додељивање ознака и успостављања релација између објеката

Између два или више објекта могуће је успоставити релације. На пример, Закон о изменама и допунама Закона о безбедности и здрављу на раду је у релацији *isPartOf* са Законом о безбедности и здрављу на раду.

Колекције представљају скупове објеката груписаних тако да се могу лакше претраживати. Детаљно се описује метаподацима из проширеног Даблинског језгра и са платформе Zotero (слика 3.14). Могу имати хијерархијску структуру, што значи да могу имати дефинисане надређене и подређене колекције. Једна колекција не мора имати подређене колекције или их може

¹² https://omeka.org/codex/Managing_Tags_2.0

имати неограничен број, али може имати само једну надређену колекцију. За сваку колекцију могуће је дефинисати видљивост на веб страници.

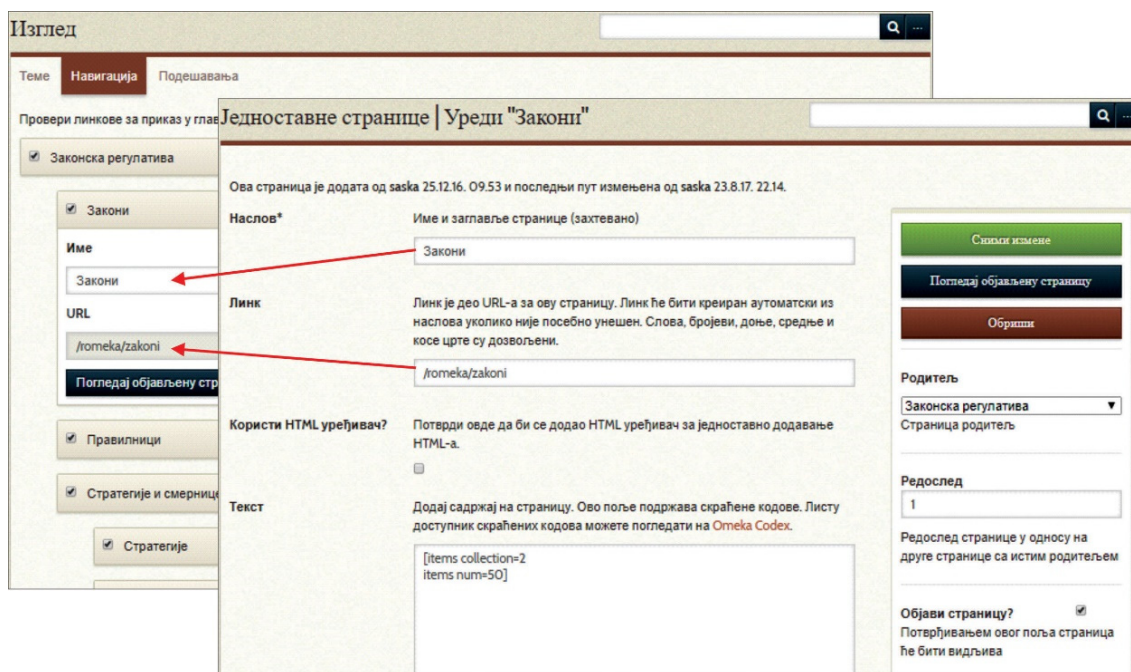


Слика 3.14. Панели за креирање колекција

Како би објекти и колекције били видљиви на вебу, неопходно је креирати веб странице. Осим назива веб стране неопходно је дефинисати и релативну путању до ње, као део URL-адресе те веб стране (слика 3.15). У делу *Текст* дозвољен је унос кодова са листе скраћених кодова, којима се прилагођава изглед саме стране. У примеру, приказаном на слици 3.15, дат је кôд за приказ колекције која се идентификује са ID 2 (*Закони*), при чему се број приказаних објеката на једној страни ограничава на 50.

```
[items collection=2  
items num=50]
```

После креирања веб стране, на панелу за уређивање изгледа сајта, у делу *Навигација* унети су основни подаци (назив и URL) за сваку појединачну веб страну, а њихов распоред уређује се једноставним превлачењем поља на жељену позицију. На овај начин веб стране се хијерархијски структурирају и на сајту се приказују у виду падајућих менија.



Слика 3.15. Креирање веб странице

3.4. ТЕРМИНОЛОШКИ РЕСУРСИ КАО РЕСУРСИ ЗНАЊА

Све језичке заједнице морају имати дефинисану термилошку политику и управљање терминологијом, што укључује развој термилошких стандарда, речника специјализованих области, термилошких база података и др.

Разлог за то је да се омогући директно и што прецизније преношење знања, при чему тежиште свакако треба ставити на проналажење и разврставање постојећих фондова знања помоћу база података.

Основу за формирање термилошких ресурса чине енциклопедије и речници. Њихов садржај је разноврстан и адекватно структуриран, те стога и представљају чврст темељ за надоградњу новим знањем, чији извори могу бити: ауторска дела, научни и стручни радови, пројектна документација, законска регулатива.

У зависности од садржаја, структуре и начина организације, термилошки ресурси могу бити: индекси, глосари, таксономије, тезауруси и онтологије (Stanković *et al.*, 2011a). Иако границе међу овим типовима ресурса нису увек јасне, на слици 3.16 приказан је један поглед који представља сваки од ових ресурса као проширење претходног, полазећи од слабе ка јакој семантици.

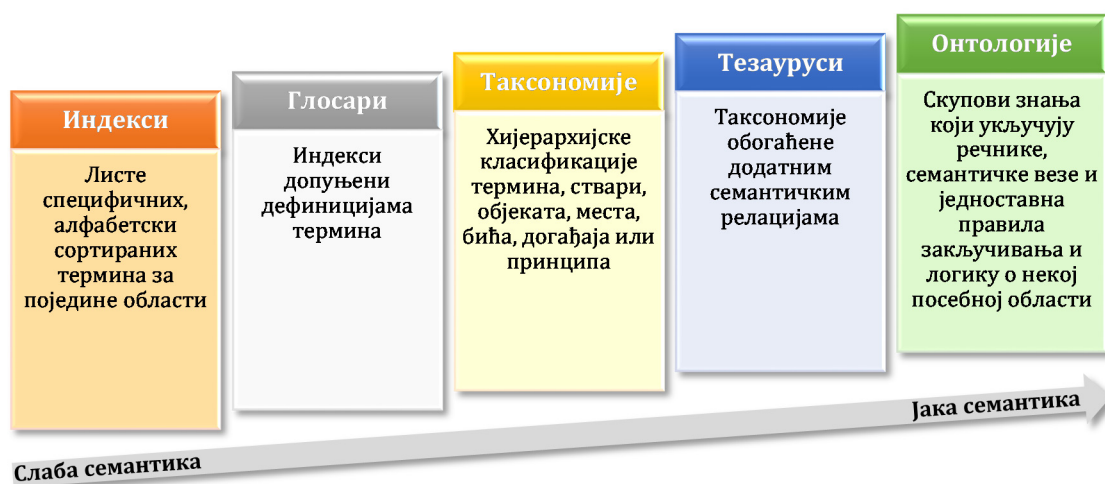
У даљем тесту они ће бити детаљније описани.

Индекси су једноставне листе, које садрже само алфаветски сортиране термина за поједине области.

Глосари представљају индексе допуњене дефиницијама термина на природном језику¹³. То су конзистентне колекцију термина за неки специфичан домен. Формирају их стручњаци из области за коју је контролисани речник намењен. Могу бити дво- или вишејезични.

¹³ Природни језик подразумева сваки људски језик настао природним процесом. Насупрот њему стоји термин вештачки језик, који може бити међународни или помоћни језик (есперанто) или неки програмски језик (Java или C++...) који се користи за комуникацију између рачунара и човека (Ђорђевић, 2017).

Таксономије представљају хијерархијску класификацију термина, ствари, објеката, места, бића, догађаја или принципа који се класификују. Имају структуру стабла за дати скуп објеката. У информатици представљају класификацију информатичких ентитета у облику хијерархије у складу са претпостављеним везама које постоје између објеката у реалном свету које ови информатички објекти представљају.



Слика 3.16. Семантика у терминолошким ресурсима

Тезауруси се могу дефинисати као контролисани речници уређени на јасан начин и структурирани тако да релације еквиваленције, хомографије¹⁴, хијерархије и асоцијативности између термина буду јасно изложене и да се лако могу препознати преко стандардних индикатора. Првенствена улога тезауруса је да олакша проналажење докумената и да оствари конзистентност у индексирању забележених докумената. Да би задовољили ову функцију, тезауруси обично повезују грубо значење једног термина са грубим значењем другог термина. Они су најчешће доменски ограничени – на пример: рударство, геологију, војну технику, итд. Контролисани речник тезауруса је тако замишљен да подржи проналажење информација пружајући

¹⁴ Хомографија (према грч. ὁμόγραφος: једнако написан) је тип хомонимије, тј. појава да се две или више речи различитог значења (хомографи) исто пишу, на пример: лук (поврће или аркада или део оружја); град (насеље или временска непогода); коса (на глави или алат за кошење траве или стрма падина) и сл.

Извор: <http://www.enciklopedija.hr/natuknica.aspx?id=26021>, приступљено 9. јануар 2018.

истовремено помоћ човеку који додељује термине, односно индексира документ, и човеку који жели да пронађе информације да претражује користећи исте термине који су коришћени при индексирању. Из свега овога се закључује да су тезауруси превасходно намењени човеку, чак иако се користе у информатичком окружењу.

Онтологије се могу посматрати као скупови знања, који укључују речнике, семантичке везе и једноставна правила закључивања, и логику о некој посебној области. Такође, могу се посматрати и као модели података који представљају одређене области, односно домене и користи се за закључивање међу објектима у том домену и везама међу њима. Онтологије се користе у вештачкој интелигенцији, семантичком вебу, информационим системима, и сл. као облик представљања знања о свету или неком његовом делу. Уопштено, онтологије описују: објекте на основном нивоу, затим класе као скупове, колекције или типове објеката, потом атрибуте као својства, обележја или параметре које објекти могу да деле и релације којима су објекти међусобно повезани (Nirenburg & Raskin, 2004).

Иако границе међу овим типовима ресурса нису увек јасне, на слици 3.16 је приказан један поглед који представља сваки од ових ресурса као проширење претходног, полазећи од слабе ка јакој семантици.

3.5. ОНТОЛОГИЈЕ

3.5.1. Дефиниција и основни појмови

Онтологија, је филозофска дисциплина која у свом изворном значењу представља науку о бићу, о томе који све концепти (типови ствари) постоје и какви су њихови међусобни односи. Означава теорију о систематском објашњавању постојања (Corcho *et al.*, 2003).

Појам онтологије се етимолошки изводи из грчких речи: онтос (ὄντος) што значи "биће", "бивствујуће", "бити", и речи логос (λόγος) што значи "учење", "реч", "говор", "наука". Дефинисан је тек у 17. веку, али се темељна разрада онтологије као филозофске дисциплине везује за античку Грчку. Према Аристотелу, наука о бићу као бићу (онтологија), заједно са науком о бићу у односу на људску душу (психологијом), науком о бићу у односу на свет (космологијом) и науком о бићу у односу на божанско (теологијом), чини метафизику, филозофску дисциплину коју је он називао "прва филозофија" (Uzelac, 1998).

Аристотелово дело "Категорије" представља први покушај човечанства да одреди системску формалну онтологију. Управо је он први формално дефинисао родове, врсте и разлике (лат. *genus/species/differentia*) као оквир за категоризацију скупа објеката, где је врста дефинисана помоћу рода којем припада и разлике једне врсте у односу на другу врсту унутар рода којем припадају (Smith, 2000). Тако је рођена наука о класификацији: први пут је знање могло бити организовано у облику таксономије тј. хијерархијски, где се знање "наслеђује" кроз произвољан број односа између родова и врста (Legg, 2007).

Тек током деведесетих година прошлог века појам онтологије, позајмљен из филозофије, добија потпуно нови смисао и значење, и постаје релевантан за истраживаче на подручју инжењеринга знања (енг. *Knowledge Engineering*) и вештачке интелигенције (енг. *Artificial Intelligence*), који су препознали потребу за складиштењем знања о реалном свету у информационе системе и његовим приказивањем (Sugumaran & Storey, 2002, Vickery, 1997). За

истраживаче, главно питање на које онтологија треба да одговори није шта је информациони систем, већ шта информациони систем мора бити у стању да закључи како би могао обавити корисне задатке (Borst, 1997).

Постоји велики број дефиниција онтологије у информатичком свету, али ни једна није општеприхваћена.

Једну од првих дефиниција онтологије су дали (Neches *et al.*, 1991). Они кажу да "онтологија дефинише основне појмове и односе које обухвата речник једне тематске области, као и правила за комбиновање термина и односа како би се речник прошири".

Uschold и Gruninger кажу да је онтологија манифестација дељеног разумевања домена која је договорена између више агената¹⁵. Такав договор омогућава прецизну и ефикасну комуникацију значења, које даље доноси нове предности као што су интероперабилност, поновно коришћење и дељење (Kolonja, 2016, Uschold & Gruninger, 1996).

Navigli и Velardi сматрају да је циљ онтологије домена да се смањи (или елиминишу) концептуална и терминолошка конфузија међу члановима једне виртуелне заједнице корисника који треба да поделе електронска докумената и информација о разних врста (Navigli & Velardi, 2004).

Најчешће цитиране су дефиниције Gruber-а и Borst-а. Gruber каже да је "онтологија експлицитна спецификација концептуализације" (Gruber, 1993). Borst, међутим, сматра да је за већину истраживача ова дефиниција сувише широка и даје своју, прецизнију, која каже да је "онтологија формална спецификација дељене концептуализације" (Borst, 1997).

"Концептуализација" се односи на апстрактан модел неког феномена тако што се идентификују релевантни појмови тог феномена (Studer *et al.*, 1998); "експлицитна спецификација" подразумева представљање концептуализације у одређеном облику употребом неког од језика за имплементацију онтологија и

¹⁵ Агент у рачунарству је софтвер, или део софтвера, оспособљен за аутономно, флексибилно, наменско деловање и резонување, с циљем испуњавања једног или више задатака.

основних елемената онтологије – класа, инстанци, релација, функција и аксиома (Gruber, 1993); "формална" се односи на чињеницу да онтологија треба да буде машински читљива, што искључује природне језике; "дељена" значи да унутар заједнице која уводи или користи ту онтологију постоји договор о томе која је намена онтологије и које заједничко знање (Studer *et al.*, 1998).

Онтологија дефинише речи и концепте, који се користе за представљање неког подручја знања, и на тај начин стандардизује значења. Онтологије користе људи, базе података и апликације који деле информације из истог домена. Оне кодирају знање из неког домена као и оно знање које покрива више домена. На тај начин оне чине могућим поновно коришћење кодираног знања. Једна онтологија садржи: класе (или опште обрасце ентитета) из различитих домена од интереса, примерке (појединачне ентитете или инстанце), везе између ових ентитета, особине и вредности особина ових ентитета, функције и процесе у које су ови ентитети укључени и ограничења и правила која укључују ове ентитете. Дакле, онтологија се може посматрати као скуп знања, који укључује речник, семантичке везе и једноставна правила закључивања и логику о некој посебној области. Такође, може се посматрати и као модел података који представља одређену област, односно домен и користи се за закључивање међу објектима у том домену и везама међу њима. Онтологије се користе у вештачкој интелигенцији, семантичком вебу, информационим системима, и сл. као облик представљања знања о свету или неком његовом делу. Уопштено, онтологије описују: објекте на основном нивоу, затим класе као скупове, колекције или типове објеката, потом атрибуте као својства, обележја или параметре које објекти могу да деле и релације којима су објекти међусобно повезани (Nirenburg & Raskin, 2004).

Неки основних појмова везаних за онтологије биће описани у наредном прегледу (Horridge *et al.*, 2004):

Класа (енг. *class*) је концепт (термин или ствар) организован у таксономију. Све класе у онтологији су подкласе највише класе owl:Thing. На пример, у рударству то може бити *Basen*.

Подкласа (енг. *subclass*) је класа изведене из надређене класе. На пример, подкласе класе *Basen* могу бити *Leziste* и *Rudnik*.

Примерак (инстанца) класе (енг. *individual*) је основна јединица класе. На пример, примерци класе *AktivanRudnik* могу бити: Тамнава-Западно поље, Соко, Лубница итд.

Атрибут је својство, особина или карактеристика класе, које пружа додатне информације о класи. На пример, за класу се може дефинисати: кључне речи, алтернативни назив, коментар, URL итд.

Својство (енг. *Property*): атрибут примерка класе и описује концепте у онтологији. Разликују се:

- **својство објекта** (енг. *Object Property*) успоставља релацију између примерака различитих класа. На пример, релација између примерка класе *PoslovniSubjekt* (ЕПС) и класе *Rudnik* (Тамнава-Западно поље);
- **својство типа податка** (енг. *Datatype Property*) - успоставља релацију између примерка и неког податка. На пример, релација између примерка класе *Rudnik* (Тамнава-Западно поље) и атрибута *godisnjaProizvodnja*.

Релација (енг. *Relation*) је тип асоцијације између концепата домена. На пример, ЕПС *imaRudnik* Тамнава-Западно поље.

Релација "is-a" приказује везу између класе и подкласе. На пример, *Rudnik* је подкласа класе *Basen* и везан је релацијом "is-a".

Домен (енг. *Domain*) је прва компонента релације, односно субјекат у RDF тројци.

Опсег (енг. *Range*) може бити концепт, тип података објекта или атрибут.

Субјекат (енг. *Subject*), **објекат** (енг. *Object*) и **предикат** (енг. *Predicate*) су делови RDF тројке (реченице) Субјекат-Предикат-Објекат. У релацији "ЕПС *imaRudnik* Тамнава-Западно поље", ЕПС је субјекат, Тамнава-Западно поље је објекат, а релација *imaRudnik* је предикат.

3.5.2. Врсте онтологија

Онтологије се разликују по томе колико детаљно описују знање из одређене области (класе, њихова ограничења и релације). Онтологије су једноставни лексикони са малим бројем релација, али и онтологије које садрже знања о великом броју појмова и њихових релација везаних за неку област. Поред онтологија које су везане за одређени домен или прилагођене одређеним врстама апликација, постоје и веома широке онтологије којима је обухваћен велики број појмова из различитих области. Онтологије се разликују и према начину представљања онтолошког знања, чиме се уједно представља и степен формализације (изражајности) тог знања (McGuinness, 2005, Vujičić Stanković, 2016).

Према нивоу знања које обухватају, онтологије могу бити: онтологије највишег нивоа (енг. *top-level ontologies*) које описују опште концепте, независне од било којег домена или задатка, доменске онтологије (енг. *domain ontologies*) које описују концепте везане за одређени домен и апликацијске онтологије (енг. *application ontologies*) које описују концепте везане за одређене задатке (Guarino, 1998).

Према (McGuinness, 2005), у смислу практичне примене, онтологије могу бити једноставне и структуриране. Под једноставним онтологијама се подразумевају хијерархијски речници тј. таксономије. Структуриране онтологије, осим машински читљивих, кодираних хијерархијских односа, садрже информације о појединим својствима и њиховим вредносним ограничењима чиме се одређени појам директно повезује са конкретним примером на кога се односи.

Једна од првих класификација онтологија презентирана је на међународном онтолошком скупу "*Ontology Summit*" одржаном 2007. године. По овом моделу, приказаном на слици 3.17 класификација онтологија представљена је у две димензије: семантичкој и прагматичној. Семантичка димензије укључују изражајност, структуру и формализацију детаља, док прагматична димензија укључује намену, управљање, аутоматско резонување, методологију дизајна и домене (знања, апликација и функција) (Gruninger *et al.*, 2008, Petrušić, 2016).



Слика 3.17. Класификација онтологија према (Gruninger *et al.*, 2008)

3.5.3. Принципи развоја онтологије

На основу искуства се може издвојити пет основних принципа развоја онтологија:

Проширивост: Онтологија треба да буде лако проширива да омогући изградњу онтологија специјализованих домена над општијим, претходно развијеним онтологијама.

Независност од апликација: Структура и садржај онтологије треба да се заснива на суштинском знању одређене дисциплине, више него на начину како се доменско знање користи.

Независност од природног језика: Структура треба да обезбеди представљање концепата, а не термина. Концепти остају исти без обзира на дијалекат, технички жаргон, стране језике, и слично. Синоними (на пример крост стене - ломљивост марине, одрон - урвина) се могу мапирати одвојено у једном елементу онтологије.

Ортогоналност: Сложени концепти треба да се рашчлане на своје саставне делове, да олакшају поновно удруживање концепата на нов начин.

Учешће заједнице: Учешће заједнице треба да води развој сваке онтологије, јер она има смисла само ако је користи више апликација (и тимова).

3.5.4. Језици за моделирање онтологија

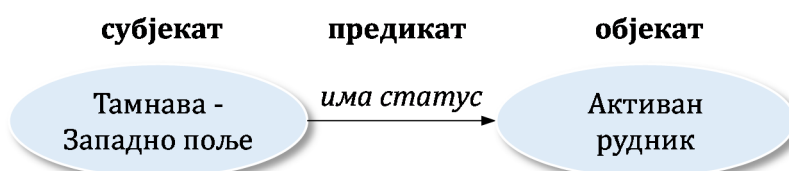
Стандардни језици за размену података на вебу и моделирање онтологија, према W3C (енг. *World Wide Web Consortium*) стандарду, су: XML, RDF/RDFS и OWL. У даљем тексту ови језици ће бити детаљније описани.

XML (енг. *Extensible Markup Language*) је настао у покушају дефинисања стандарда којим ће се размењивати подаци на вебу, али тако да се сва пажња усмери на садржај, а не на приказ. XML је језик који нема предефинисан скуп кључних речи (елемената и атрибута) већ је то језик за дефинисање других језика, па се XML може посматрати као метајезик¹⁶. Насупрот HTML-у којим се може описати искључиво изглед документа без његове структуре, XML је намењен за представљање произвољне структуре докумената. XML документ се састоји из етикета које су отворене и затворене, где свака етикета може имати одређен број парова атрибут-вредност. Имена етикета дају се произвољно. Основно обележје XML -а јесте то да број етикета није ограничен, већ се може дефинисати помоћу XML Scheme. XML Scheme дефинише који су то елементи који ће се наћи у једном XML документу, затим који су то атрибути који припадају појединим елементима и какав садржај се очекује унутар сваког елемента. XML шемом се дефинише каква ће бити структура XML документа.

RDF (енг. *Resource Description Framework*) је предложен од стране W3C конзорцијума (енг. *World Wide Web Consortium*) за моделирање метаподатака о ресурсима на вебу. Развијен је како би се њиме изразило значење структура докумената. RDF није језик, већ је модел којим се подаци могу представљати на различите начине. Како је XML постао стандард за серијску репрезентацију и размену информација, најшире је прихваћен и за представљање и размену RDF описа.

¹⁶ Језик помоћу којег је могуће изградити друге језике.

RDF описује семантичке везе између електронских извора. Информације у RDF-у су тзв. искази у облику тројки: *субјекат*, *предикат* и *објекат*. Ова три саставна дела исказа се називају *Resource* а идентификују се помоћу URI (енг. *Uniform Resource Identifier*). Један ресурс (*субјект*) је повезан с другим ресурсом (*објектом*) преко везе која је одређена трећим ресурсом (*предикатом*), слика 3.18.



```
<owl:NamedIndividual rdf:about="http://www.semanticweb.org/RuDokOnto#TamnavaZapad"/>  
<rdf:Description rdf:about="http://www.semanticweb.org/RuDokOnto#imaStatus"/>  
<owl:NamedIndividual rdf:about="http://www.semanticweb.org/RuDokOnto#AktivanRudnik"/>
```

Слика 3.18. RDF искази у облику тројка: субјекат, предикат и објекат

RDFS (енг. *Resource Description Framework Schema*) (RDFS, 2017) - како RDF ипак не може у потпуности да представи значење садржаја који описује уводи се RDFS који је намењена дефинисању речника за RDF податке, и који дефинише, на пример, врсте објеката на које се одређено својство може применити.

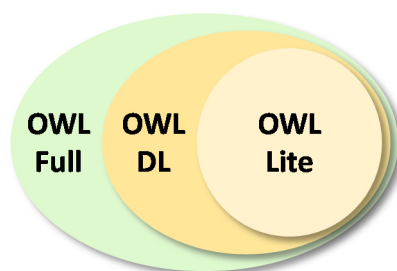
OWL језик (енг. *Web Ontology Language*) је W3C је прогласио за стандардни језик којим се могу представљати и размењивати онтологије на вебу. Настао је као потреба за проширењем RDF и RDFS, а изведен је из DAML+OIL онтолошког језика.

Пошто је OWL заснован на RDF-у и RDFS, OWL онтологија се може представити у виду графа на ком су представљене RDF тројке. То значи да се онтологија написана у OWL-у представља коришћењем синтаксе RDF-а. Оно што издваја OWL изнад RDF-а и RDFS јесте нови ниво закључивања који је на знатно већем нивоу у односу на исти у RDF-у. OWL карактерише већи степен изражавања у односу на RDFS, у смислу да се OWL-ом могу прецизније описати класе и особине информационих ресурса.

Један од основних захтева у току развоја OWL-а био је да се омогући његова проширивост у смислу да се онтологије написане у OWL језику могу једноставно комбиновати са другим онтологијама написаним у OWL-у.

W3C је за различите потребе писања онтологија увео три међусобно зависне врсте језика (слика 3.19):

- OWL Full,
- OWL DL и
- OWL Lite.



Слика 3.19. Међусобни однос онтолошких језика:
OWL Full, OWL DL и OWL Lite

OWL Full има максималну изражајност и синтаксно је независан од RDF-а, али не даје никакве гаранције по питању времена рачунарске обраде. Главна карактеристика OWL Full језика, у поређењу са OWL DL и OWL Lite, је да једна класа, која је по дефиницији колекција индивидуа, може бити и сама индивидуа. Ова особеност OWL Full-а чини време потребно за обраду на њему заснованих модела бесконачним.

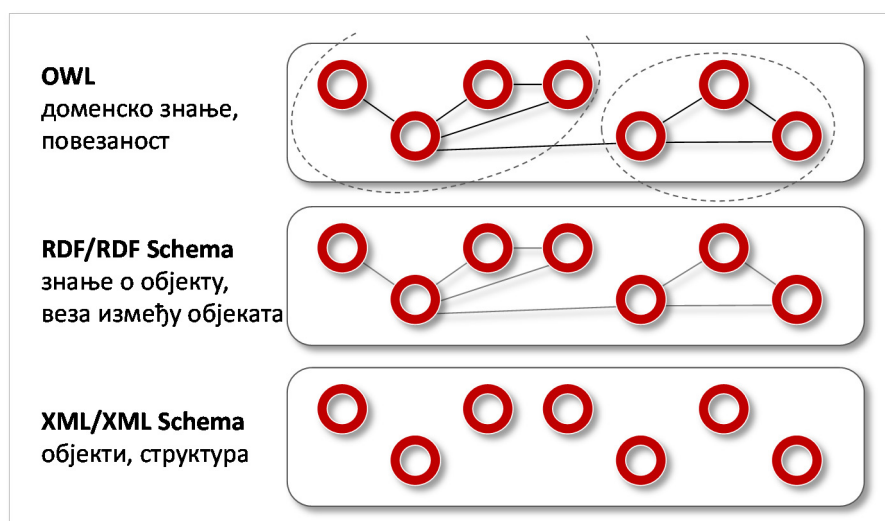
OWL DL (енг. *Descriptive Logic*) карактеришу: висок степен изражајности, гарантована комплетност рачунарске обраде (за све тврдње се гарантује да ће се израчунати) и одлучивост (све обраде ће се завршити у коначном времену). OWL DL укључује све конструкције из OWL Full и додаје извесна ограничења. Најзначајније ограничење је да класа не може бити индивидуа или особина, и да особина не може бити индивидуа или класа. OWL DL има доста добру формалну основу пошто је заснован на дескриптивној логици.

OWL Lite је првенствено намењен да подржи класификациону хијерархију и једноставна ограничења. Овај језик је добра полазна основа за развој онтолошких алата. Користан је за миграцију постојећих таксономија у OWL.

OWL Full проширује OWL DL, који проширује OWL Lite, што значи да је свака OWL Lite онтологија истовремено и OWL DL и OWL Full онтологија и да је свака

OWL DL онтологија истовремено и OWL Full онтологија. Обрнуте релације не важе. Међусобни однос онтолошких језика приказан је на слици 3.19.

Ниво представљања знања и семантика дати су на слици 3.20 (Petrušić, 2016). На слици се уочава да се на првом нивоу, XML-ом се само описују објекти и дефинише њихова структура. На следећем нивоу, језиком RDFS се се подиже ниво знања успостављањем веза између објеката. На највишем нивоу налази се OWL, језик којим се објекти повезују у домене и производи знање.



Слика 3.20. Програмски језици Семантичког веба (Petrušić, 2016)

3.5.5. Семантички веб и веб-сервиси

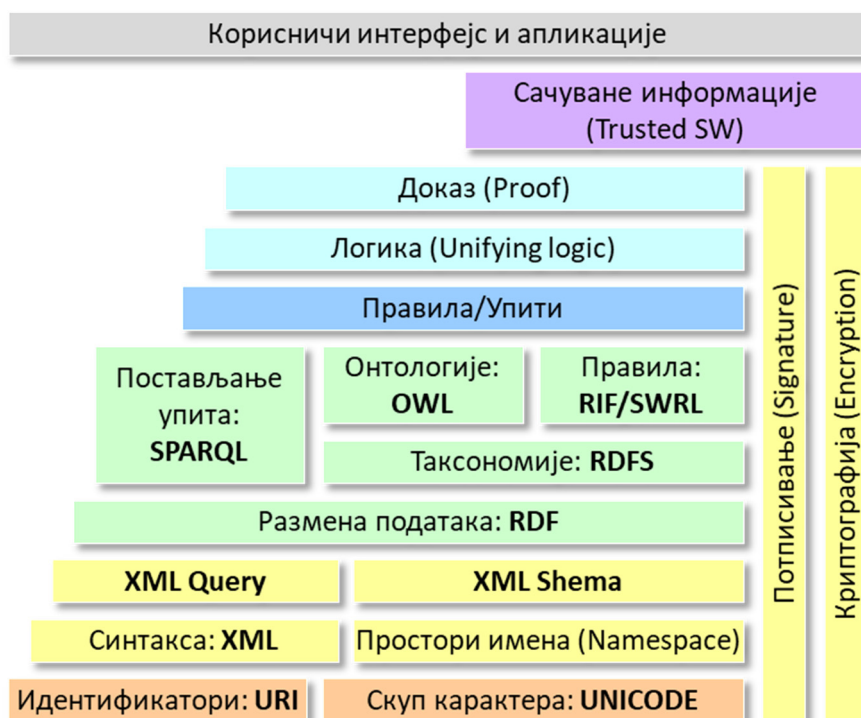
Коришћење електронских извора информација зависи од добрих могућности претраживања, и проналажења а појам семантичког веба је настао у трагању за ефикаснијим решењима за проналажење информација. WWW (енг. *World Wide Web*), уз помоћ машина за претраживање и огромног броја расположивих метаподатака омогућава добијање информација али које тек у неком степену могу задовољити информациону потребу корисника. Један од великих недостатака данашњег веба је немогућност издвајања одговарајућег значења уколико реч која се тражи има више значења. Та вишезначност се може манифестовати у једном језику (*коцка* као геометријско тело и *коцка* као игра на срећу), а може се односити и на различите језике (*ред*, у српском језику значи између осталог, поредак, а у енглеском језику *црвен*). Тако би

једноставна претрага оваквих речи, путем неког од претраживача данашњег веба, дала за корисника углавном неадекватне резултате.

Истовремено је све више истраживачких напора да се на самом вебу повећа ефикасност претраживања до добијања што релевантнијих информација. Tim Berners-Lee, творац данашњег веба, је још 1998. године покренуо нову иницијативу за развој будућег веба, пригодно названог Семантички веб који треба да превазиђе истакнуте проблеме. Он је 1999. године писао: "Први корак је да се подаци на вебу ставе или конвертују у такав облик да их машине могу разумети. То ствара оно што ја зovem Семантички веб – сплет података које машина може директно или индиректно обрађивати" (Berners-Lee, 1999). Као један од најновијих резултата рада W3C конзорцијума, семантички веб представља скуп организованих технолошких стандарда, производа и информација који су повезани на такав начин да се информације могу индексирати и семантички филтрирати кроз процес обраде на глобалном нивоу. Семантички веб доприноси ефикасном претраживању и тиме што омогућава посебан начин представљања информација на Вебу које се могу посматрати и као глобално повезане онлајн базе података.

Према томе, семантички веб је замишљен као проширење садашњег веба, које ће садржати не само документа које човек може да чита уз помоћ прелистача, већ документа која су анотирана мета-информацијама. Ове мета-информације дефинишу о чему се у документу ради на начин који машина може да обради. Експлицитно представљање мета-информација заједно са доменским теоријама, односно онтологијама, омогућиће да семантички веб пружи квалитетно нов ниво услуга. Он повезује невероватно велику мрежу људског знања, коју допуњује могућношћу машинске обраде. Многобројни аутоматизовани сервиси помажу кориснику да постигне свој циљ, приступајући и обезбеђујући информације у машински читљивом облику. Овај процес би на крају могао да доведе до стварања изузетно великог система знања који поседује разне специјализоване системе за резонување који би подржавајући готово све аспекте живота могли бити у будућности за човека незаобилазни.

Архитектуру семантичког веба чине три важна стандарда – XML, RDF и онтологије. Међусобни однос компонената семантичког веба је дат на слици 3.21.



Слика 3.21. Архитектура семантичког веба¹⁷

Бројне организације, као што су W3C, IETF (енг. *Internet Engineering Task Force*) и OASIS (енг. *Organization for the Advancement of Structured Information Standards*) пружају подршку технологијама базираним на XML-у, као што су веб сервиси засновани на SOAP (енг. *Simple Object Access Protocol*) који обезбеђују интероперабилни интерфејс са апликацијама широм Интернета. RDF обезбеђује начин за повезивање информација. Користећи XML као синтаксу за линеаризацију, RDF је постао основа за остале језике засноване на онтологијама за семантички веб. XTM (енг. *XML Topic Maps*) (XTM, 2017) је стандард за технологију мапа тема помоћу којих се семантички карактеришу и категоришу документи и делови докумената на вебу и то у односу на њихов садржај са циљем да истакну којим темама или предметним подручјима се документ бави. Веб сервиси пружају механизам софтверским програмима да комуницирају међусобно. Онтолошки језици (OWL, DAML+OIL) су већ на

¹⁷ <https://gyires.inf.unideb.hu/GyBITT/08/ch02s03.html>

располагању и многе организације их користе да обогатиле своје базе знања семантиком. Семантички веб и веб сервиси се паралелно развијају. XML и језик WSDL (енг. *Web Services Description Language*) који описује SOAP интерфејсе ка веб сервисима, нису довољни за развој семантичког веба. Потребна је аутоматизована подршка за рад са бројним специјализованим форматима података. Може се очекивати да ће у скорој будућности бити могућ семантички опис проблема и пословних процеса из готово свих домена. Онтологије су полако постале кључни концепт семантичког веба у коме се преплићу човеково разумевање симбола са могућностима машине. Велики напори се улажу у онтологије у веб сервисима. DAML-S (енг. *DARPA Agent Markup Language for Services*) је један од таквих праваца који се заснивају на RDF-у и RDF Schema и DAMN+OIL. Једноставно речено, DAML-S је онтологија за веб сервисе. SWWS (енг. *Semantic Web Enabled Web Services*), развијен 2002. године, треба да обезбеди што обухватнији оквир за проналажење и описивање веб сервиса, као и скалабилно посредовање међу веб сервисима. Заједно, ове две технологије треба да допринесу ширем аутоматизованом коришћењу веб сервиса.

3.5.6. SWEET онтологија

Пример више онтологије је SWEET (SWEET, 2017) онтологија, развијена у оквиру пројекта *Semantic Web for Earth and Environmental Terminology*, под покровитељством NASA (енг. *National Aeronautics and Space Administration*), за научне дисциплине које се баве Земљом (енг. *Earth System Science*) и животном средином (слика 3.22).

SWEET онтологије обухватају више хиљада појмова, који покривају различите области науке о Земљи и повезане концепте користећи OWL језик. Онтологије су јавно доступне и могу се преузети са званичне веб стране. Основне онтологије су објашњене у наставку текста. Појединачне онтологије су представљене правоугаонцима, а линија којом су повезане указује где су основна својства искоришћена за дефинисање концепата онтолошких простора (Raskin, 2006).

Земљина кугла (енг. *EarthRealm*) је онтологија која обухвата Земљине сфере и заснована је на физичким својствима планете. Елементи ове онтологије су "атмосфера", "океан" и "чврсто тло", и придружене подобласти (као "дно океана" и "атмосферски гранични слој"). Сваки елемент има подобласти које се разликују од надређених класа по неком својству, на пример "тропосфера" је подкласа од "атмосфере" где је елевација између 0 и 15 km.



Слика 3.22. SWEET онтологије¹⁸

¹⁸ <https://sweet.jpl.nasa.gov/>

Мртва природа (енг. *Non-LivingSubstances*) обухвата неживе чиниоце природе и могу бити, на пример разне врсте зрачења или хемијска једињења итд.

Жива бића (енг. *LivingSubstances*) укључују биљке и животиње. Ова онтологија је импортована из таксономије биосфере GCMD (енг. *NASA's Global Change Master Directory - GCMD*).

Физички процеси (енг. *PhysicalProcesses*) укључује процесе који утичу на жива бића и мртву природу.

Физичка својства (енг. *PhysicalProperties*) је онтологија која обухвата температура, притисак, висину, састав, итд. Ова својства су обично мерљиве физичке величине са придруженим мерним јединицама и односе се на жива бића, мртву природу, физичка својства итд.

Мерене јединице (енг. *Units*) су дефинисане коришћењем Unidata's UDUnits. Онтологија укључује конверзионе факторе између различитих мерних јединица, па је тако, на пример, ознака km дефинисана као специјалан случај од m са одговарајућим конверзионим фактором.

Време (енг. *Time*) је у суштини нумеричка скала са терминологијом специфичном за временски домен, у којој су опсеги и временске релације посебни случајеви нумеричких опсега и релација. Временски опсеги (енг. *extents*) укључују: трајање, сезону, век, 1996, итд. Временске релације би биле: после, пре, итд.

Простор (енг. *Space*) је вишедимензионална нумеричка скала са терминологијом специфичном за просторни домен, у којој су опсеги и временске релације посебни случајеви нумеричких опсега и релација. Просторни опсеги би били: земља, Антарктик, екватор, итд. Просторне релације су: горе, доле, северно од, итд.

Нумерика (енг. *Numerics*) укључује: интервал, тачку, 0, P2 итд. Нумеричке релације су: веће од, мах, итд. Дефинисани су вишедимензиони концепти, пошто они нису у изворним OWL и XML окружењима.

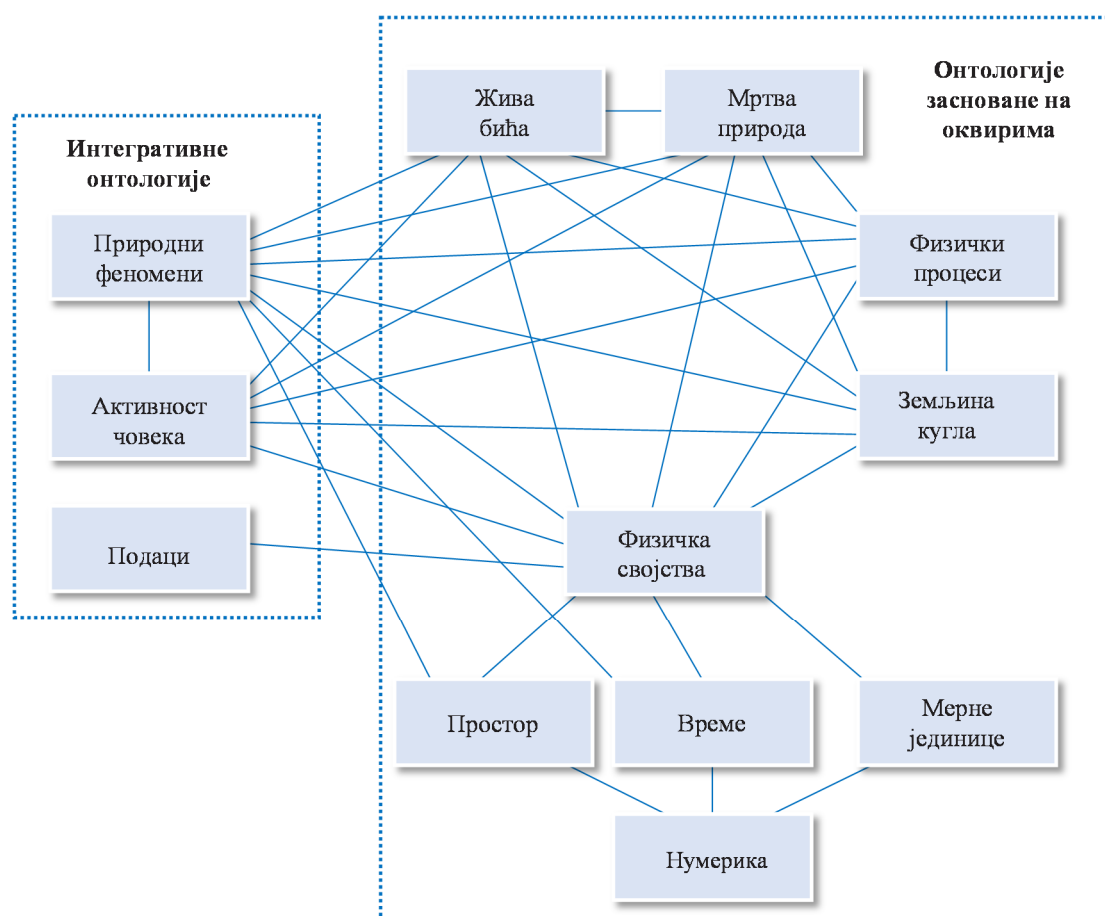
Природни феномени (енг. *NaturalPhenomena*) је онтологија која се користи за дефинисање пролазних догађаја, на примери: ураган, земљотрес, Ел Нињо,

вулкан, терористички догађај, и сваки је повезан са временом, простором, односно са: *Time, Space, EarthRealms, NonLivingElements, LivingElements* итд. Такође су укључени специфични случајеви појава, који обухватају око 50 догађаја.

Активности човека (енг. *HumanActivities*) је онтологија која представља активности којима се људи баве. Потреба за представљањем ових активности произилази из чињенице да људске активности утичу на процесе и појаве.

Подаци (енг. *Data*) је онтологија која пружа подршку за концепте скупова података, укључујући и репрезентацију, складиштење, моделирање, форматирање, ресурсе, сервисе и дистрибуцију.

Међусобне релације између основних онтологија приказане су на слици 3.23.



Слика 3.23. Основне онтологије и њихове међусобне релације
(Raskin, 2006)

Ове онтологије представљају концепт простора за научне дисциплине која се баве Земљом. Исказивање концепта речима може да се разликује од једне особе до друге у зависности од њихове научне заједнице, културног порекла, језика и нивоа стручности. Да би се задовољиле и те потребе SWEET омогућава да исти концепт буде представљен користећи различите термине.

Алат за претрагу претрагу садржи функцију која, уз помоћ онтологије, потенцијално може да пронађе ресурсе без тачног подударања кључних речи. Тако алатка за претрагу користи SWEET онтологију да пронађете синонине и специфичније термине од оних који су дати у захтеву за претрагу. Након проналажења одговарајућих термина, алат прослеђује унију ових термина GCMD алату за претрагу и приказује резултате.

Даље побољшање претраге засноване на онтологији може бити аутоматско откривање веза између термина. На пример, између појмова "угљен-диоксид" и "глобално загревање" постоји јака веза, тако да када је један термин појављује, вероватно се појављује и други. За аутоматску сумаризацију текста се користи GCMD DIF (енг. *Global Change Master Directory, Directory Interchange Format for Datasets - GCMD DIF*), из ког се прави матрица асоцијација (веза). Да би се пронашле скривене релације између појмова користи се метода семантичке анализе скривених значења (*latent semantic analysis, LSA*). Релације који су откривене овом анализом се такође користе у алату за претрагу.

4. РУДАРСКА ПРОЈКТНА ДОКУМЕНТАЦИЈА

Рударство, као једна од економски веома значајних привредних грана, своје пословање организује и изводи по одговарајућим стандардима светске рударске праксе. Сходно томе, циљ ове дисертације је да се омогући кориснику да на ефикасан начин изврши анализу те документације, упореди са светском праксом и на тај начин омогући поравнање, односно усаглашавање документа у циљу сагледавања потреба државе а тиме и заинтересованих инвеститора. На овај начин се стварају реални услови за лакше усаглашавање ових документа на нивоу светске рударске праксе и ствара се повољнија инвестициона клима у овој привредној грани.

Рударски пројекти имају веома дуг животни век. Од почетних истраживања лежишта, преко процене ресурса и резерви, детаљних техно-економских евалуација, производње и прераде па до затварања рудника, може понекад проћи чак и више од сто година. Свака фаза развоја пројекта, без обзира на тренутно стање у коме се он налази, подлеже одређеним процедурама извођења истражних и/или рударских радова и евалуације резултата како би пројекат могао прећи у наредну фазу развоја.

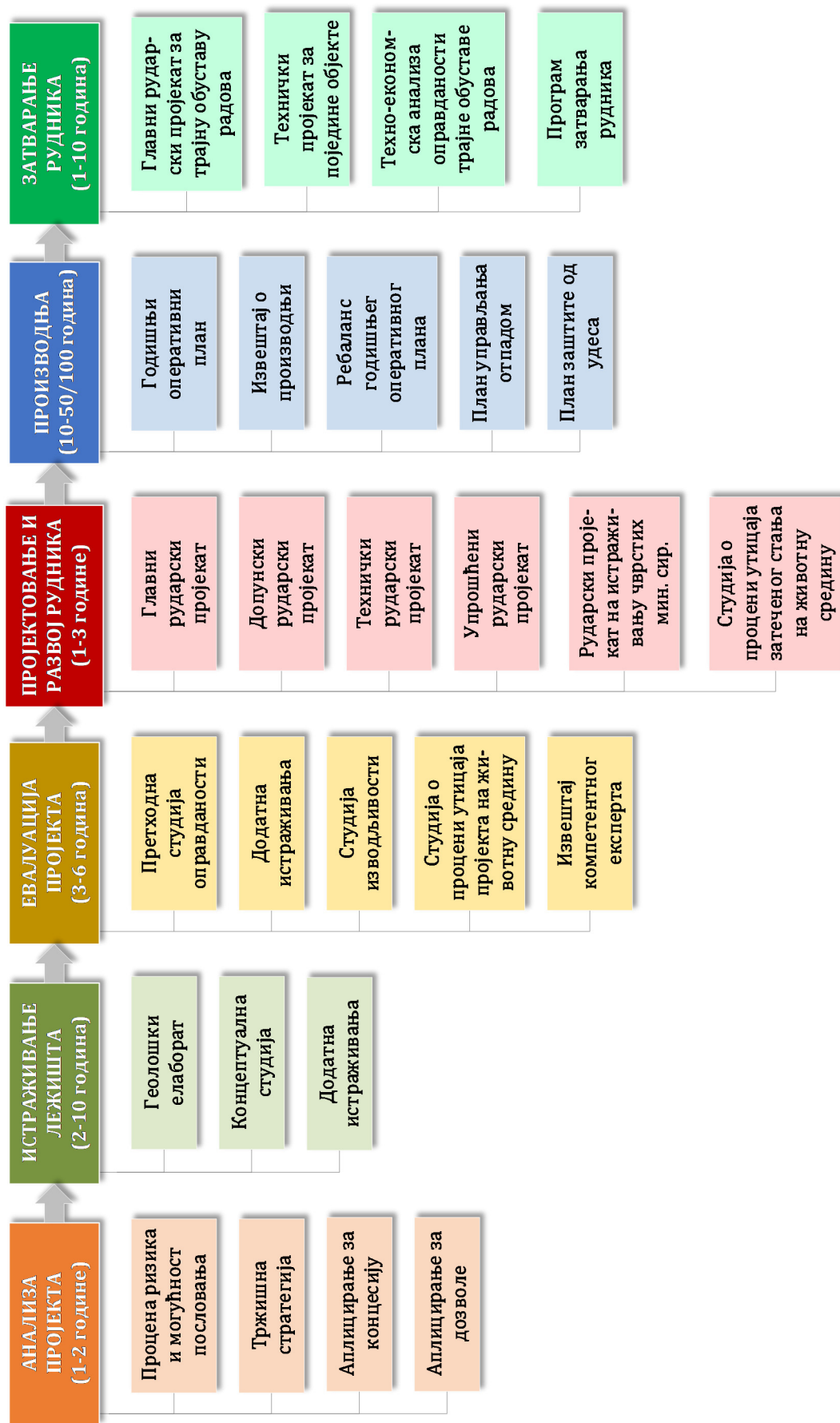
У циљу потпуног разумевања структуре и процедура развоја рударског пројекта, потребно је јасно дефинисати животни циклус рудника, на основу кога проистичу потребе за изградом појединих рударских пројеката на основу којих се доносе инвестиционе одлуке или изводе рударски радови.

Фазе развоја једног рударског пројекта су (слика 4.1):

- **Анализа пројекта**, почетна фаза и којој се процењују ризици и могућност пословања, креира се тржишна стратегија и уколико су анализе повољне, аплицира се за концесију и дозволе неопходне за даљи развој пројекта. Трајање ове фазе је 1 – 2 године.
- **Истраживање лежишта**, фаза у којој се применом основних геолошких метода (геофизичка мерења, геохемијска анализа, узорковање итд.)

дефинише лежиште, израђује геолошки елаборат и концептуална студија, а по потреби се раде и додатна истраживања. Трајање ове фазе може бити од 2 – 10 година.

- **Евалуација пројекта**, фаза која имају за циљ доказивање техничке, економске и еколошке одрживости будућег рудника у оквиру које се раде претходна студија оправданости, студија изводљивости и студија о процени утицаја пројекта на животну средину. Сврха ових студија је да омогуће компанијама доношење одлуке о томе да ли и колико је пројекат сигуран, еколошки оправдан, економски одржив и друштвено одговоран. Трајање ове фазе може бити од 3 – 6 година.
- **Пројектовање и развој рударских радова**, фаза у оквиру које се ради неопходна инвестиционо-техничка документација и која служи за успостављање инфраструктуре на руднику (изградња: путева, постројења за прераду, система управљања заштитом животне средине, објеката за становање запослених и других објеката). Трајање ове фазе може бити од 1 – 3 године.
- **Производња**, фаза експлоатације руде унапред дефинисаном методом, припреме и прераде руде (дробљење, млевење, одвајање руде и јаловине, прерада руде у концентрат, топљење и рафинација - добијање метала из концентрата, и пречишћавање метала. Маркетинг се односи на пласирање производа, или концентрата ако се не ради даља прерада, купцу (топионици или произвођачу). Трајање ове фазе може бити 100 и више године, у зависности од величине лежишта.
- **Затварање рудника**, фаза у којој се ради план управљања затварањем рудника, затим техно-економска анализа трошкова предложене стратегије затварања, решава се питање заштите животне средине и анализирају се социјално-друштвене последице процеса затварања на локалну заједницу. Трајање ове фазе може бити до 10 година.



Слика 4.1. Животни циклус рударских пројеката

Истраживање минералних сировина је истовремено комерцијално одржив и профитабилан посао и наука. Због тога постоје техничко-технолошки, економски и геолошки аспекти ових истраживања. Рударском пројекту претходе процеси истраживања, дефинисања и економске оцене минералног лежишта. Ова истраживања су фронтални део истраживања и рударске индустрије. Компаније и/или појединци истражују земљу да би пронашли и дефинисали нова економски значајна минерална лежишта. Само када се овим истраживањима утврди да је једно лежиште економски вредно, рударски пројекат може да се настави кроз фазу експлоатације.

Ово поглавље дисертације почиње прегледом пословних циклуса рударског пројекта укључујући активности од истраживања до затварања рудника и описом пратеће рударске документације неопходне за реализацију сваке од фаза рударског пројекта.

Нивои истраживања варирају од раних фаза проспекције терена, која има за циљ проналажење нових перспективних подручја, до великих програма бушења који су усмерени на оцену постојећих ресурса, слика 4.1. Овај посао се предузима у зависности од типова истражних подручја која варирају у величини, локацији, степену познавања минерализације и потенцијалу за ширење познате минерализације, а тиме и рударског пројекта.

Према (Stevens, 2010) истражно подручје је површина земље за коју је компанија, група или појединац прибавила легална права истраживања минералних сировина или права експлоатације. Истражни простори варирају у величини, од малих величине испод квадратног километра, до оних чија је величина и преко хиљаду квадратних километара.

Истраживање се спроводи на неком простору, односно делу земље за који су прибављена права на истраживање минералних сировина од стране појединца или компаније. Неко истражно подручје се генерално прибавља било преко прибављене дозволе или преко уговора о заједничком истраживању са другом компанијом или особом која има права за истраживања.

Процес добијања дозволе или права над минералним сировинама је различит у свету. У неким правним системима, захтев за дозволу се тражи за власништво над обележеном парцелом, у другим правним системима захтев за права над минералним сировинама и подноси се владиној агенцији. Процес добијања права у неким регионима може даље да се компликује због вишестепених права власништва. На пример, права над минералним сировинама за неки регион могу делимично да буду у власништву државе, група или појединаца, који могу да имају различита права везана за стицање права и поседовање.

Упркос процесу прибављања дозволе или права над минералним сировинама, од виталног значаја је обезбедити сигурност преко закупа. Инвеститор мора да буде уверен да поседује чиста легална права над минералним ресурсима који подлежу том праву пре него што започне инвестирање у истраживање и развој рудника. Генерално, права над минералним сировинама поседује држава, док власништво над земљом (површинско право) могу да поседују држава, корпорације, групе или појединци. У већини случајева власник површинских права мора да дозволи власнику права над минералним сировинама приступ земљи у циљу истраживања минерала испод површине земље.

Главни циљ већине правних система јесте подстицање активности на истраживањима, како би компаније које инвестирају у истражне радове могле да сачувају минерална права.

Не постоји јединствена дефиниција или класификациони систем за истражни простор. Ипак, термини, као што су: "greenfields", "brown-fields", "advanced" и "pre-feasibility", се данас најчешће користе као ознаке за истражна и/или експлоатациона подручја или већ истражене просторе.

Термин "**greenfields**" се односи на релативно неистражено подручје, за које није познато да ли садржи минерална лежишта. "Greenfields" подручја могу да се добијају на основу доступних геофизичких и геохемијских података који се налазе у државном власништву, пре него на основу познатог присуства минерализације (Stevens, 2010).

Термин "**brown-fields**" се односи на подручје које представља познату рударску област, подручје у околини неког рудника или оно за које постоје сазнања о постојању минералног лежишта. Потенцијал "brown-fields" подручја се заснива на претпоставци да тамо где постоји једно лежиште велика је вероватноћа да има више њих (Stevens, 2010). Ова подручја имају већи потенцијал за откривање нових минералних лежишта од оних означених термином "greenfields", али су зато скупља.

Термин "**advanced**" се односи на подручје које садржи познато минерално лежиште, при чему је неопходно даље дефинисање његове величине и карактера. Минерално лежиште показује потенцијал да експлоатација буде економски исплатива, али су потребна додатна истраживања да би се оценила његова потенцијалност. Истраживање је фокусирано на распрострањење лежишта и на оцену минералних ресурса. Већина радова у "advanced" нивоу укључује бушење бушотина којим се настоји да се утврди величина лежишта ради оцена претходно претпостављених минералних ресурса. Поред бушења могу се користити и друге технике истраживања, посебно геофизичка мерења и геофизички каротаж (Stevens, 2010).

У овој фази се врше припреме за израду претходне студије изводљивости: изводе се опити прераде руде и металуршки тестови, процене величине рудника или друге специфичне студије. Такође, обављају се консултације са локалном заједницом о активностима компаније и могућности развоја рудника. Такође, врше се и основна истраживања животне средине, као почетни корак за даље анализе и процене заштите животне средине.

Термин "**pre-feasibility**" се односи на подручје које садржи минерално лежиште за које је извршена процена минералних ресурса. Примарни циљ истраживања на овом нивоу је да се унапреди процена минералних ресурса у процесу припреме за израду студије изводљивости. Карактеристике "pre-feasibility" подручја су (Stevens, 2010):

- Додатна бушења, чији је циљ побољшање поузданости процене минералних ресурса од категорије индицираних ресурса до измерених

категорија. Сврха додатних бушења је повећање густина до тада израђених бушотина (енг. *infilling drilling*), растојања између бушотина у овој фази су релативно мала, од 25-50 метара или мање у неким случајевима.

- Инжењерске, металуршке, еколошке и економске студије које су, поред додатних бушења, непосредна припрема за израду студије изводљивости.
- По завршетку "pre-feasibility" фазе рударски пројекат може да настави са финалном студијом изводљивости и планирања и пројектовања развоја рудника (слика 4.1), или може да стави пројекат на чекање до побољшања тржишних и развојних услова.

Као што је речено, истраживање се одвија у етапама полазећи од почетног тражења појава минералних сировина до завршног развоја експлоатације и самог рудника, тако да у свакој наредној фази постоји постепено смањење ризика. У раној фази постоје неизвесности да ли минерализација уопште постоји на датом подручју. Једном када се одреди перспективност или открије појава минерализације, ризик је смањен. Следећа фаза утврђује да ли појављивање минералне сировине представља минерално лежиште - ако је одговор потврдан, ризик се још једном смањује. Постепено смањивање ризика наставља се како истраживање напредује све време до развоја рудника и експлоатације. Једном када се рудник отвори, ризик је на најнижем нивоу, мада ни тада није елиминисан. Цене руде, неочекивани изазови у производњи, промене у политици владе и друге непознанице наставиће да представљају ризик у економској оправданости процеса експлоатације.

Наравно, са смањењем ризика значајно се повећава цена истраживања и потенцијалне добити. На путу од проналажења лежишта до развоја рудника постоји неколико кључних, преломних тачака, у којима менаџмент компаније мора да одлучи да ли активности на рударском пројекту треба да се наставе, привремено или трајно обуставе, да ли пројекат може бити предмет заједничког улагања, или пак делимичне или потпуне продаје другој компанији. С обзиром да се цене повећавају на сваком кораку, важно је да се поуздано оцени да ли потенцијал подручја даје гаранције да ће се трошкови

покрити. Ово није лака одлука и захтева комбинацију науке, вештине, интуиције и врло често среће.

Као што је претходно елаборирано и илустровано на слици 4.1, да би један рударски пројекат прешао у фазу планирања и пројектовања рудника потребна је позитивна оцена финалне студије изводљивости (енг. *Bankable Feasibility Study*), која представља основу за израду рударске инвестиционо-техничке документације.

Студијом изводљивости минерални ресурси се превode у рудне резерве (експлоатационе резерве), као што је илустровано на слици 4.2.



Слика 4.2. Веза између минералних ресурса и минералних резерви

Претходном студијом изводљивости се индицирани и/или измерени ресурси превode у веројатне рудне резерве, док се студијом оправданости измерени ресурси превode у доказане експлоатационе резерве.

Сваки од рударских инвестиционих и техничких докумената, зависно од фазе развоја пројекта, се израђује према одређеном стандарду за ту врсту документа. У наставку текста се даје преглед и краћи опис ових стандарда, како код нас тако и у светској рударској пракси.

4.1. ИНВЕСТИЦИОНО-ТЕХНИЧКА ДОКУМЕНТАЦИЈА У РЕПУБЛИЦИ СРБИЈИ

Инвестиционо-техничка документација у Републици Србији, према којој се врши експлоатација минералних ресурса, је дефинисана важећим *Законом о рударству и геолошким истраживањима*¹⁹ (Službeni glasnik RS, 101/2015).

Ова документација подразумева следеће документе (члан 84. *Законом о рударству и геолошким истраживањима*):

1. Претходна студија оправданости;
2. Студија изводљивости експлоатације лежишта минералне сировине;
3. Дугорочни програм експлоатације;
4. Рударски пројекти;
5. Годишњи оперативни план.

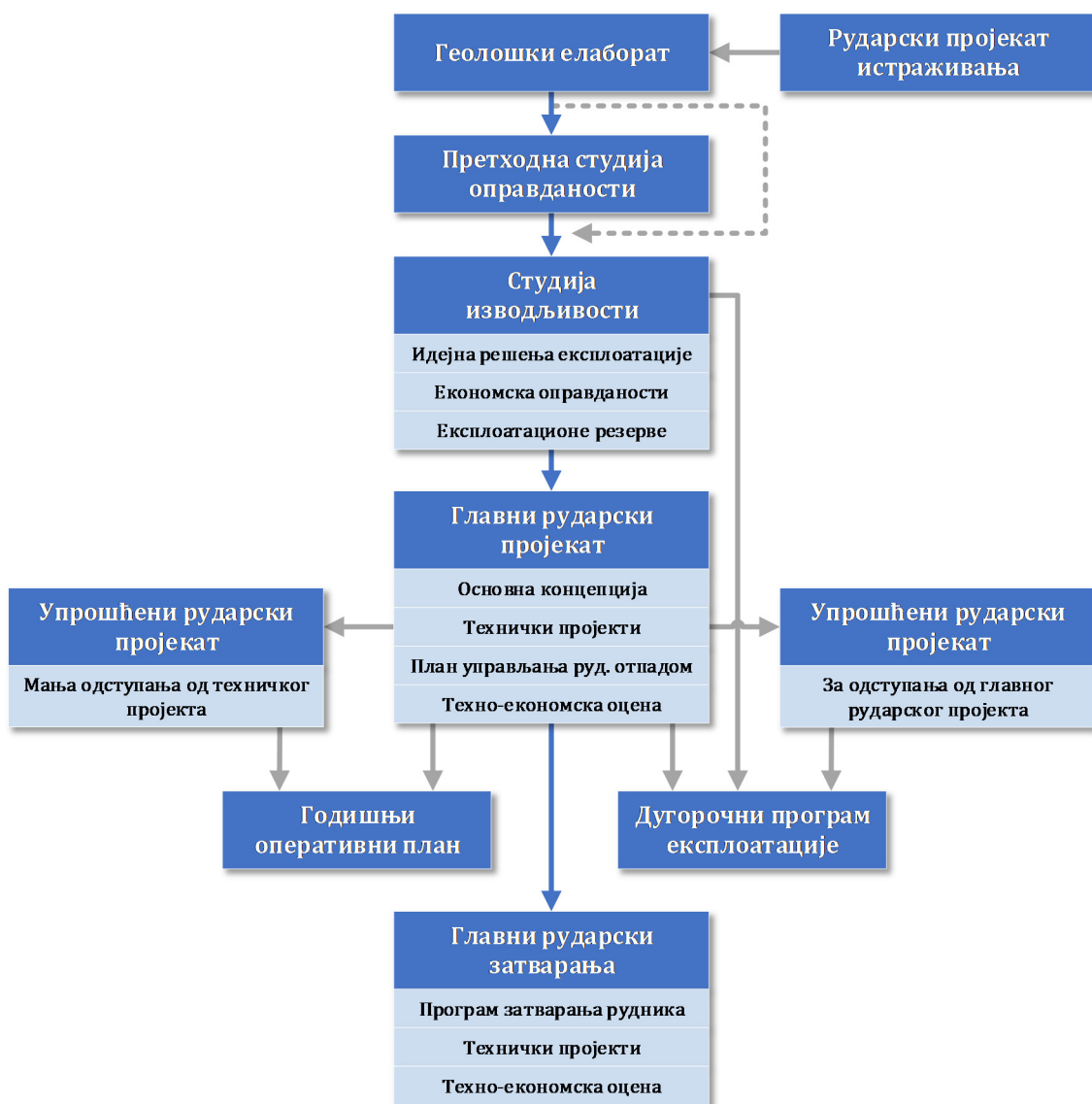
Након техно-економске евалуације рударског пројекта кроз студију изводљивости, следи израда рударских извођачких пројеката, који у зависности од тренутног статуса рудника, могу бити:

- Главни рударски пројекат;
- Допунски рударски пројекат;
- Технички рударски пројекат;
- Технички рударски пројекат за експлоатацију минералних ресурса за добијање природних грађевинских материјала;
- Рударски пројекат на истраживању чврстих минералних сировина;
- Упрошћени рударски пројекат.

Структура, условљеност израде и везе између рударских докумената је приказана на слици 4.3.

За потребе ове дисертације у наставку се даје кратак опис намене ове документације и референцирање на важеће правилнике који детаљно дефинишу садржину ових докумената.

¹⁹ http://romecka.rgf.rs/files/original/Zakoni/Zakon_o_rudarstvu_i_geoloskim_istrazivanjima/Zakon_o_rudarstvu_i_geoloskim_istrazivanjima.1.pdf



Слика 4.3. Структура, редослед израде и везе међу рударским пројектима

Претходна студија оправданости је документ који се израђује у току детаљних геолошких истраживања у циљу техно-економске евалуације оправданости експлоатације лежишта и наставка финансирања даљих активности на реализацији рударског пројекта. Тачност процене техничко-технолошких решења, инвестиција и трошкова је на нивоу од $\pm 35\%$.

Овом студијом се индицирани и измерени ресурси преводе у вероватне и доказане резерве минералних сировина на основу варијантних решења експлоатације и прераде минералних сировина, екологије, тржишта и

економске оцене и доноси одлука о оправданости улагања у додатне истражне радове и израду студије изводљивости (Bullock, 2017).

Студија изводљивости експлоатације лежишта минералне сировине – представља свеобухватну техничку и економску анализу методе развоја рударског пројекта експлоатације лежишта минералних сировина у којој су дате детаљне процене о реалним претпоставкама о: геологији лежишта, експлоатацији, припреми, металуршкој преради, економији, маркетингу, законским условима, екологији, социјалним условима и одлукама владе заједно са детаљном економском анализом, које су неопходне да би се показало да је процес експлоатације у тренутку израде студије технички изводљив и економски оправдан (Bullock, 2017).

Резултати студије изводљивости служе као основа за финалну одлуку инвеститора или финансијске институције да се настави са финансирањем даљег развоја рударског пројекта.

Овај документ се израђује на довољно поузданим геолошким, рударским и финансијско-економским параметрима тако да коначна решења и закључци су са нивоом тачности у опсегу од $\pm 15\%$ (Bullock, 2011).

Овом студијом се измерени ресурси преводе у доказане резерве минералних сировина, које у датом тренутку представљају експлоатабилне резерве руде и где постоји тржишна оправданост за њихову експлоатацију и прераду.

Детаљан садржај израде овог документа је прописан *Правилником о садржини студије изводљивости експлоатације лежишта минералних сировина*²⁰ (Službeni glasnik RS, 108/2006).

У складу са *Законом о процени утицаја на животну средину*²¹ (Službeni glasnik RS, 135/2004, 36/2009), студију изводљивости експлоатације лежишта минералне сировине прати спровођење поступка процене утицаја експлоатације лежишта минералне сировине на животну средину. Поступак процене утицаја

²⁰ http://romeka.rgf.rs/files/original/Pravilnici/Pravilnik_o_sadrz_stud_izvodljiv_eksploat_lez_min_sir/Pravilnik_o_sadrz_stud_izvodljiv_eksploat_lez_min_sir.2.pdf

²¹ http://romeka.rgf.rs/files/original/Zakoni/1482/Zakon_o_procena_uticaja_na_ziv_sred.1.pdf

на животну средину је превентивна мера заштите животне средине заснована на изради документа Студије о процени утицаја на животну средину, којом се анализира: квалитет чинилаца животне средине, осетљивост и међусобни утицаји чинилаца животне средине на одређеном простору, а на основу оцене утврђују и предлажу мере којима се штетни утицаји могу смањити и отклонити. Предмет поступка процене утицаја, сагласно закону, поред пројеката који се планирају и изводе су и промене технологије, реконструкције, проширење капацитета, престанак рада и уклањање пројеката који могу имати значајан утицај на животну средину. У том случају се Студија о процени утицаја на животну средину ради и за друге категорије рударских пројеката како је то приказано на слици 4.1.

Дугорочни програм експлоатације – је документ који се израђује за експлоатацију минералних резерви које су од посебног значаја за Владу или је држава већински власник. Детаљан садржај израде овог документа је прописан *Правилником о садржини дугорочног програма експлоатације лежишта минералних сировина и годишњих планова извођења рударских радова*²² (Službeni glasnik RS, 27/97a).

Годишњи оперативни план – је документ који се израђује за потребе планирања извођења рударских радова на годишњем нивоу, као и годишњи извештај о пословању за претходну календарску годину.

Експлоатација рудних резерви се врши према рударским пројектима, који се израђују за дати рудник зависно од тренутног статуса рударског пројекта. Детаљан садржај према коме се израђују рударски пројекти је прописан *Правилником о садржини рударских пројеката*²³ (Službeni glasnik RS, 27/97b).

У томе смислу рударски радови на терену се могу изводити према следећим категоријама рударских пројеката (Službeni glasnik RS, 101/2015):

²² http://romeka.rgf.rs/files/original/Pravilnici/Pravilnik_o_sadrzini_dugoroc_programa_eksplot_vez_min_sir/Pravilnik_o_sadrz_dugoroc_programa_eksplot_vez_min_sir.1.pdf

²³ http://romeka.rgf.rs/files/original/Pravilnici/Pravilnik_o_sadrz_rud_proj/Pravilnik_o_sadrz_rud_proj.1.pdf

Главни рударски пројекат – је извођачки пројекат и израђује се у складу са техничким решењима датим у студији изводљивости. Овим пројектом се детаљно дефинишу техничко-технолошка решења експлоатације и прераде минералне сировине до нивоа финалног производа. Пројекат садржи основну концепцију експлоатације, техничке пројекте за поједине технолошке процесе на основу којих се изводе рударски радови, граде рударски објекти, рудничка инфраструктура, техничко-технолошке целине и техно-економску оцену пројекта.

Допунски рударски пројекат – овај пројекат се израђује само у случајевима значајних техничко-технолошких одступања од решења датим у главном рударском пројекту. Овај пројекат садржи све елементе и целине и израђује се према истом правилнику као и Главни рударски пројекат.

Технички рударски пројекти - израђују се, у складу са идејним решењима датим у основној концепцији главног или допунског рударског пројекта. Они детаљно разрађују сваку производну технолошку операцију на руднику, и раде се сагласно истом правилнику.

Рударски пројекат на истраживању чврстих минералних сировина - представља пројекат на основу којег се изводе рударски радови у оквиру одобреног пројекта геолошких истраживања на истражном простору или експлоатационом пољу, изван простора обухваћеног главним или допунским рударским пројектом. Циљ ових рударских радова је:

- узимање узорка за лабораторијска и технолошка испитивања на лицу места или у индустријским условима,
- прикупљање геомеханичких параметара и других података о геолошкој грађи и лежишним условима,
- утврђивање просторних параметара минералних ресурса и
- остала истраживања за потребе градње.

Рударски пројекат на истраживању минералних сировина је извођачки пројекат који нарочито садржи: основни концепт са динамиком извођења рударских радова, техничка решења по којима ће се изводити рударски

радови, предмер и предрачун радова, мере безбедности и здравља на раду, план управљања рударским отпадом, као и мере за санацију и рекултивацију простора обухваћеног радовима у случају обуставе истраживања. У случају извођења истражних подземних рударских радова на основу овог пројекта потребно је предвидети мере за одржавање подземних просторија, објеката и инсталација по завршетку истражних радова

Рударски пројекат на извођењу рударских радова ван експлоатационог поља - је извођачки пројекат који се ради самостално или у склопу пројекта за потребе изградње рударских или грађевинских објеката и који нарочито садржи: урбанистичко-техничке услове, основни концепт са динамиком извођења рударских радова, техничка решења по којима ће се изводити рударски радови, предмер и предрачун радова, мере безбедности и заштите на раду, као и мере заштите животне средине.

Упрошћени рударски пројекат израђује се за:

1. сва мања одступања од усвојених техничких решења обрађених у техничком рударском пројекту, с тиме да извођење рударских радова према упрошћеном рударском пројекту може трајати до једне године;
2. израду појединачних истражних бушотина за нафту, гас, подземне воде, геотермалних ресурса, радове у њима, као и за надземна постројења и уређаје за експлоатацију, припрему и транспорт нафте, гаса и подземне воде до сабирне станице; технолошка опробавања рударског отпада.

4.2. ИНВЕСТИЦИОНО-ТЕХНИЧКА ДОКУМЕНТАЦИЈА ПРЕМА СВЕТСКИМ СТАНДАРДИМА У РУДАРСТВУ

Инвеститори у рударству се суочавају са изазовом одлучивања у који пројекат компанија треба да инвестира, с обзиром на вишеструку расположивост међународних пројеката, због различите природе стандарда реализације и услова инвестирања широм света. Са садашњим растом у рударској индустрији, компаније обично припремају извештаје за инвеститоре, на основу берзанских извештаја према одговарајућим стандардима. Постоји много стандарда за процену ресурса/резерви, односно евалуацију рударских пројекта широм света (Stevens, 2010).

Већина водећих земаља које имају јаку рударску индустрију присутне су на већини светских берзи у оквиру којих обављају свој бизнис под окриљем интернационалне организације CRIRSCO (*Committee for Mineral Reserves International Reporting Standards*) (CRIRSCO, 2018). Ова организација је формирана 1994. године под покровитељством Већа рударских и металуршких института (CMMI - *Council of Mining and Metallurgical Institutes*) и представља групу представника организација које су одговорне за израду кодекса и смерница за извештавање о ресурсима у (Abbott Jr, 2017):

- JORC (*Joint Ore Reserves Committee*) (JORC, 2018), Аустралазија²⁴,
- CBRR (*Comissão Brasileira de Recursos e Reservas*) (CBRR, 2018), Бразил,
- CIM (*Canadian Institute of Mining, Metallurgy and Petroleum*) (CIM, 2018), Канада,
- *Comisión calificadora de competencias en recursos y reservas mineras* (Comisionminera, 2018), Чиле,
- PERC (*Pan-European Reserves & Resources Reporting Committee*) (PERC, 2018), Европа,
- MPIGM (*Mongolian Professional Institute of Geosciences and Mining*) (Mining), Монголија,
- NAEN (*Национална асоцијација по експертизи недр*) (НАЭН, 2018), Русија,

²⁴ Аустралазија је подручје које обухвата Аустралију, Нови Зеланд, Нову Гвинеју и околна острва од којих се већина налази у источном делу Индонезије

- SAMREC (*South African Code For The Reporting Of Exploration Results, Mineral Resources And Mineral Reserves*) (SAMREC, 2018), Јужна Африка и
- SME (*Society for Mining, Metallurgy, and Exploration*) (SME, 2018), САД.

Комбинована вредност рударских компанија које се котирају на берзама ових земаља чини више од 80% капитала наведеног у рударској индустрији.

Преглед националних стандарда извештавања који су у складу са принципима CRIRSCO је приказан у табели 4.1, а на слици 4.4 су приказане земље чланице CRIRSCO које користе јединствене критеријуме и принципе у процени ресурса и резерви, као и изради студија оправданости.

Табела 4.1. Стандарди за процену ресурса и резерви

Стандард	Извор
CRIRSCO	http://www.crirSCO.com/news_items/5_standard_definitions.pdf
JORC Code, Аустралазија	http://www.jorc.org/docs/jorc_code2012.pdf
CBRR, Бразил	http://www.crirSCO.com/docs/CBRR_Documents_Guide.pdf
CIM GUIDELINES (NI43-101), Канада	http://www.crirSCO.com/docs/cim_definition_standards_20142.pdf
PERC Reporting Code, Европа	http://www.crirSCO.com/perc_reporting_standard_2013_rev2.pdf
KAZRC, Казахстан	http://www.crirSCO.com/docs/Kazakhstan_Code.pdf
MPIGM, Монголија	http://www.crirSCO.com/docs/The_MRC_Code.pdf
NAEN Code, Русија	http://www.crirSCO.com/news_items/naen_code.pdf
SAMREC Code, Јужна Африка	http://www.crirSCO.com/docs/SAMREC_2016.pdf
SME GUIDE, САД	http://www.crirSCO.com/docs/2014_sme_guide_%20june_10_2014_appendix_a_update_march_2016.pdf
UNFC, Уједињене нације	http://www.unecsc.org/fileadmin/DAM/ie/se/pdfs/UNFC/UNFC.pdf
CERTIFICATION Code, Чиле	http://www.crirSCO.com/docs/2015_Codigo_Ingles.pdf

Сличност различитих националних стандарда и смерница за извештавање омогућило је CRIRSCO да развије Међународни шаблон за извештавање о ресурсима. Овај стандард се користи као "основни оквир и смернице" за сваку земљу која жели прилагодити свој властити стандард за извештавање под окриљем CRIRSCO, након што укључи одредбе специфичне за дату земљу, као што су правне и инвестиционе регулативе.



Слика 4.4. Земље чланице CRIRSCO асоцијације²⁵

Такође, поред стандарда за процену ресурса постоје и међународни стандарди за евалуацију рударских пројеката, како би се на јединствен начин и према истој или сличној методологији извршила оцена техничких и економских параметара пројекта. У табели 4.2 су приказани стандарди за евалуацију рударских пројеката.

²⁵ <http://www.criresco.com/welcome.asp>

Табела 4.2. Стандарди за евалуацију рударских пројеката

Стандарди за израду студије оправданости	Извор
<p>CIMVAL Standards and guidelines Canadian Institute of Mining, Metallurgy and Petroleum on Valuation of Mineral Properties National Instrument 43-101, Standards of Disclosure for Mineral Projects www.cim.org</p>	
<p>VALMIN Code and Guidelines Australasian Institute of Mining and Metallurgy (AusIMM) http://www.valmin.org/docs/VALMIN_Code_2015_final.pdf</p>	
<p>SAMVAL Code South African Mineral Asset Valuation Committee https://www.samcode.co.za/</p>	
<p>SME Valuation Standards Committee of the Society for Mining, Metallurgy, And Exploration, Inc. http://www.smenet.org/publications-resources/resources/sme-valuation-standards-tutorials</p>	
<p>IASC - International Accounting Standards Committee IVSC Code - International Valuation Standards Committee www.ivsc.org IMVC - International Mineral Valuation Committee International Mineral Property Valuation Standards Template http://web.cim.org/standards/documents/Block724_Doc187.pdf</p>	

4.3. ПРАТЕЋА РУДАРСКА ДОКУМЕНТАЦИЈА

Поред докумената наведених у претходним поглављима за рад и праћење успешности рада неког рудника потребно је прибавити, располагати и пратити и више десетина додатних докумената. Ти документи се пишу у облику прегледних и детаљних извештаја, елабората о обављеним радовима и уоченом (забележеном) реалном стању, решења о прихватању или одбијању неке иницијативе, жалби и приговора на неко решење, захтева за набављање или организовање неке активности, услова за остваривање неког права, одобрења неких радова, дозвола за рад по неким условима итд.

Суштински се сва пратећа документација може сврстати у следеће групе:

- **Административна документација**, која обухвата: услове, решења, дозволе, захтеве, жалбе, предлоге, препоруке, технички и еколошки аудит, односно обухвата сву документацију чије поседовање и усаглашавање је неопходно вршити са надлежним органима власти.
- **Додатни истражни радови**, који обухватају: елаборате, цртеже, табеле и прилоге везане за додатно и континуирано истраживање руде, услове прераде, начин транспорта, припреме, депоновања итд.
- **Периодични извештаји о раду рудника и појединих делова**, који обухватају: извештаје надлежног особља о раду, стању опреме, постигнутим ефектима и капацитетима, поремећајима, билансе концентрације, билансе потрошње појединих норматива итд.
- **Извештаји о техничком одржавању**, који обухватају: систематизацију рокова и програма планираног одржавања, радове изведене као непланирано одржавање, препоруке за унапређење процеса одржавања, препоруке за набавку резервних делова итд.
- **Извештаји из области безбедности и здравља на раду**, који обухватају: акт о процени ризика, елаборат о испитивању услова радне околине, елаборат о периодичним прегледима и испитивању средстава за рад и опреме личне заштите на раду итд.

- **Мониторинг** обухвата извештаје о стању појединих елемената животне средине на које утицај има или може имати рудник, елаборате о постигнутим резултатима и трендове загађивања, извештаје припремљене за надлежне државне органе, планове за адаптацију и прилагођавање програма мониторинга, планове за усаглашавање вредности појединих индикатора итд.).

Сва напред побројана пратећа документација се допуњује из дана у дан, врло често показује само затечено стање, употребљивост јој је ограничена на унапред дефинисано време, али представља неопходни део документације која се прати, упоређује са законским нормама и условима из дозволе за рад, повремено јавно показује да би се нешто демантовало или потврдило. Уз наведено, ова документација служи за вођење и планирање посла око расподеле радног времена и времена одржавања, а такође представља и основу за планирање трошкова производње итд.

По правилу сва поменута документација се доставља у папирном и електронском облику. Вештим управљањем документацијом у електронском облику убрзава се процес, смањује број радника задужен за чување, разврставање и налажење те документације уз повећање ефикасности њеног коришћења.

5. МОДЕЛ УПРАВЉАЊА РУДАРСКОМ ПРОЈЕКТНОМ ДОКУМЕНТАЦИЈОМ

Када је у питању документација на српском језику, посебна карактеристика рударске документације, осим мултимедијалног садржаја, је чињеница да се текстуална документација чува на два равноправно коришћена писма: ћирилици и латиници. Ово је нешто што стандардни рачунарски системи за руковање текстуалним документима у основи не подржавају. Комплексна граматика српског језика и различити облици речи представљају други изазов са којим се корисници срећу сваки пут када се претражују документи на српском. Други словенски језици деле сличне проблеме.

Многи од ових проблема се могу решити помоћу (природно) језичких технологија (енг. *Human Language Technology*, HLT). Циљ језичких технологија је да савлада све облике писаног и говорног језика користећи информационе технологије (енг. *information technology*, IT), помажући људима да сарађују, послују, размењују знање и учествују у политичким и друштвеним дебатама, без обзира на језичке баријере или њихове IT вештине. Језичке технологије обухватају неколико основних апликација које омогућавају обраду језика у оквиру ширих програмских система (Vitas *et al.*, 2012). У склопу Групе за језичке технологије на Математичком факултету Универзитета у Београду развијен је низ језичких алата алата и ресурса за српски језик, међу којима су свеобухватни електронски морфолошки речници и корпуси који покривају српску општу лексику, док је развој ресурса за одређене домене стални задатак.

С обзиром на наведене проблеме и могућности које нуде језичке технологије, ово истраживање је имало за циљ развој система који би омогућио ефикасно управљање рударском пројектном документацијом уз помоћ проналажења информација (IR, енг. *information retrieval*) и екстракције информација (енг. *information extraction*, IE). Проналажења информација подразумева проналажење и повлачење текстова који одговарају информационој потреби корисника исказаној кроз захтев за информацијом (кориснички упит) из скупа датих текстова (Manning *et al.*, 2008), док екстракција информација подразумева

анализирање информација садржаних у тексту, њихово одабирање, обележавање и организовање у структурираним скуповима података, као што су онтологије и базе података, за даљу обраду (Jurafsky & Martin, 2014). Ове две области, проналажење информација и екстракција информација, иако различите, често користе исте ресурсе, као и алате.

Систем управљања рударском документацијом, који је описан у овој докторској дисертацији, креиран је тако да буде независан од писма (ћирилица, латиница) и да подржава претраживање кључних речи датих у упиту у свим граматичким формама. Поред тога, систем омогућава проширење упита на синонине, као и претраживање помоћу лексичких маски. Систем омогућава систематизацију, структурирање и складиштење законских прописа заједно са пројектном документацијом како би се обезбедило унакрсно упоређивање и провера усклађености сегмента пројектне документације са важећим прописима. Овако конципиран, систем је значајан за рударску индустрију, јер омогућава успешно управљање пројектном документацијом, које није само брзо, флексибилно и једноставно за коришћење и примену, већ покрива и цели животни циклус рударских активности, почевши од фазе анализе, кроз истраживање, процену добијених резултата, пројектовање, производњу, и кроз фазу затварања рудника. Тиме се постиже неопходан ниво сложености који омогућава позиционирање ризика и контролу пројектног циклуса и координацију рада на сложеним пројектима са више уговорних страна и неколико заинтересованих страна.

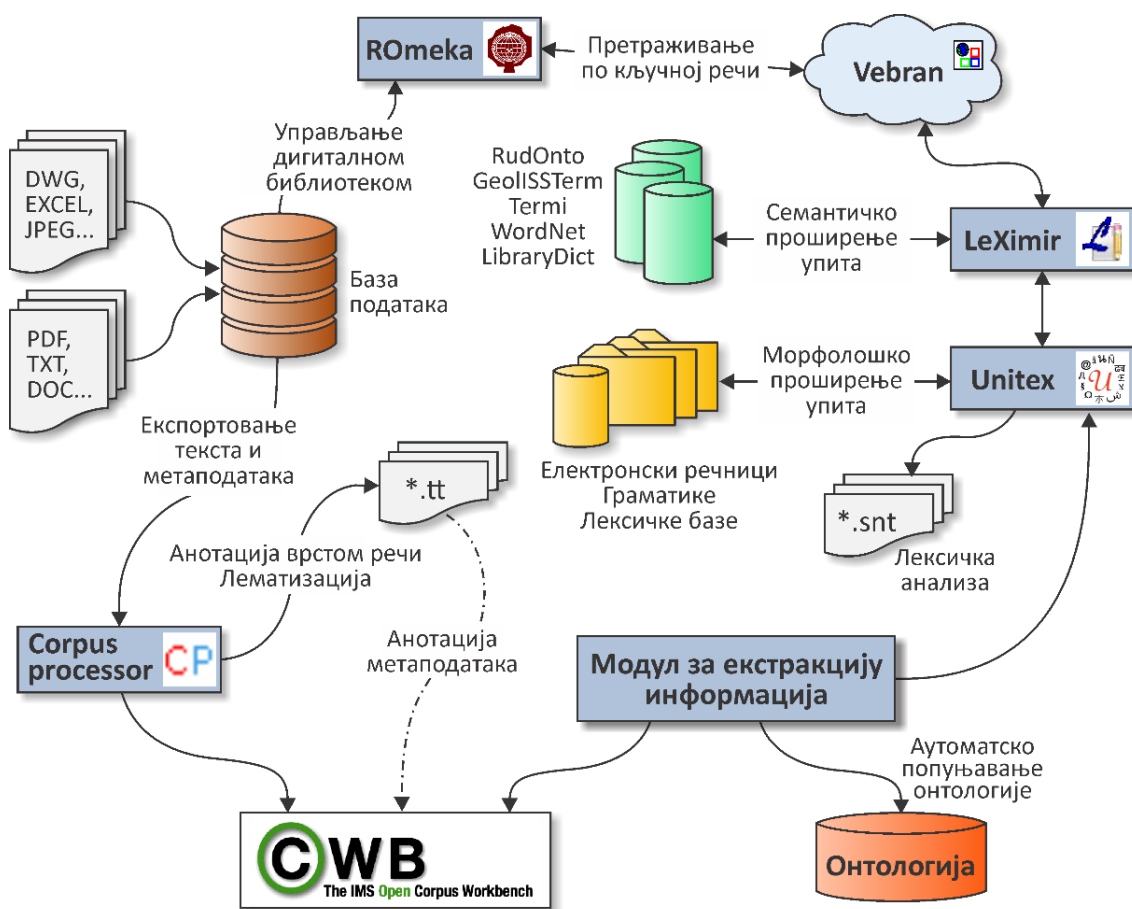
Да би се обезбедило проналажење информација и екстракција информација, развијена су два главна извора који садрже релевантну документацију: дигитална библиотека и пратећи корпус сакупљених и обрађених текстова из ове библиотеке. Поред тога, коришћени су бројни језички ресурси (Krstev *et al.*, 2008b), алати и технике (Krstev, 2008), интегрисани у систем који је шематски приказан на слици 5.1. Систем се састоји од различитих типова компоненти: развојне компоненте за припрему, побољшање и допуњавање ресурса (на пример Leximir, Corpus processor), као и компоненте које се користе за екстракцију и проналажење информација (на пример IMS Open

Corpus Workbench, Unitex, ROmeка@RGF). Интеракција и размена података између компоненти система биће детаљније описани у наредним одељцима.

У даљем тексту биће описани основни елементи система за управљање документима у електронском облику помоћу језичких ресурса и технологија, илустрован примером управљања рударском пројектном документацијом, и то:

- дигитална библиотека, која је централни репозиторијум за складиштење рударске пројектне документације,
- рударски корпус,
- лексички и терминолошки ресурси,
- онтологија.

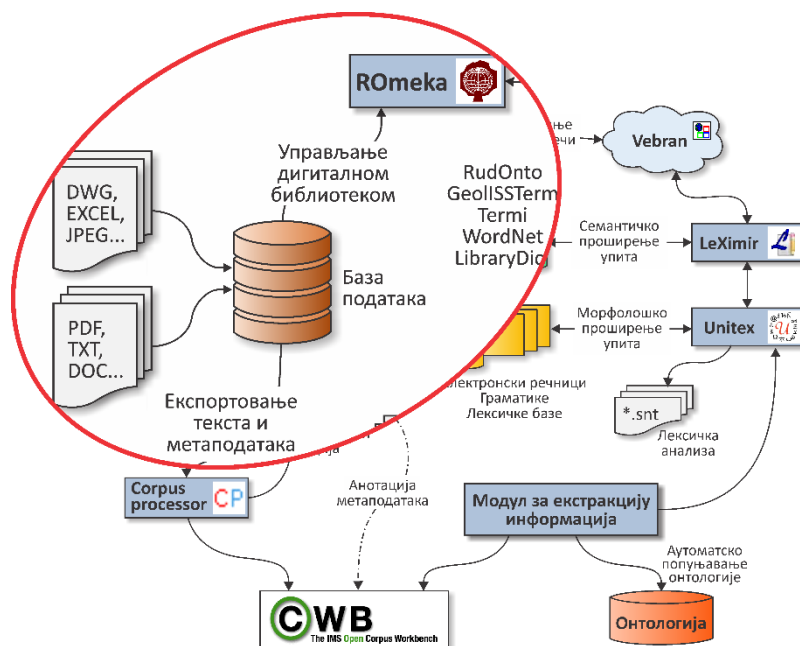
Посебна пажња посвећена је примерима (сценаријима) коришћења система, где су представљени различити приступи за претраживање пројектне документације.



Слика 5.1. Система управљања документацијом заснован на језичким технологијама

5.1. RОmeка@RGF – РУДАРСКА ДИГИТАЛНА БИБЛИОТЕКА

Централни репозиторијум за складиштење рударске пројектне документације је дигитална библиотека RОmeка@RGF²⁶. Њена позиција у систему за управљање рударском пројектном документацијом приказан је на слици 5.2.



Слика 5.2. Централни репозиторијум за складиштење рударске пројектне документације

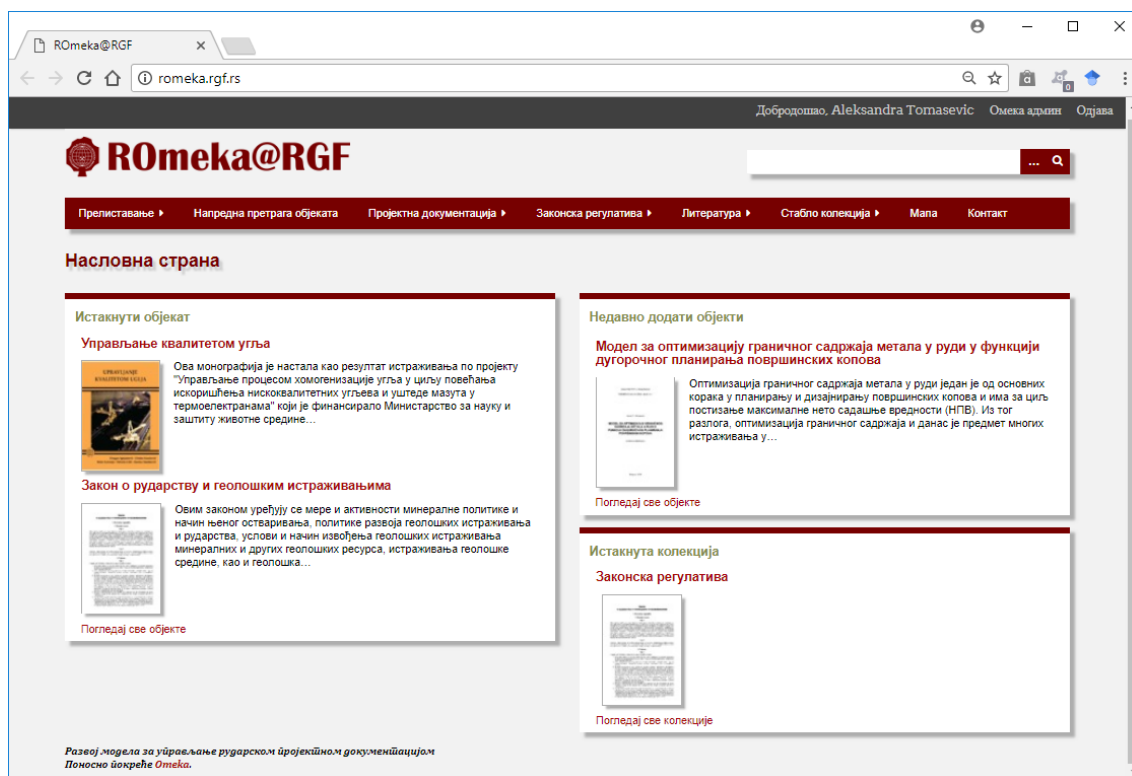
Користећи веб платформу Омека, описану у поглављу 3.3.7, развијена је дигитална библиотека RОmeка@RGF, намењена за складиштење, приказивање и управљање пројектном документацијом из области рударства, као и блиско повезаних домена: безбедности и заштите на раду, заштите животне средине и процене ризика. Изглед почетног панела приказан је на слици 5.3.

Објекти су сврстани у 4 главне, надређене колекције и 15 подколекција. Хијерархијска структура колекција приказана је у табели 5.1.

Сви објекти детаљно су описани свим типовима метаподатака, чији је детаљан опис дат у поглављу 3.3.4, изузев метаподатака о праву коришћења чија се

²⁶ Назив RОmeка@RGF је настао комбиновањем скраћеног назива Рударског одсека Рударско-геолошког факултета из Београда и веб платформе Омека. <http://romeка.rgf.rs>

употреба планира у наредном периоду. Коришћен је скуп елемената метаподатака Даблинског језгра, с обзиром на чињеницу да је ова шема не само широко прихваћена, већ и једноставна, функционална и стандардизована (ISO 15836-1:2017).



Слика 5.3. Почетни панел дигиталне библиотеке ROMeka@RGF

Видљивост на вебу омогућена је за већину дигиталних објеката, али ограничења су уведена за све објекте из колекција *Пројектна документација* и *Стандарди*, и неколико објеката из колекције *Литература*. Разлог за ово ограничење везано је за питање права публикавања, било због поверљивости података било због ауторских права.

Дигитална библиотека садржи 209 објеката односно докумената, а селекција се базирала на доступним дигиталним ресурсима, при чему за потребе ове дигиталне библиотеке није рађено додатно скенирање докумената расположивих искључиво у папирном облику.

Дигитална библиотека садржи целокупну актуелну законску регулативу Републике Србије из домена рударства и безбедности и заштите на раду, као

и литературу која би пројектантима и инжењерима омогућила проналажење потребних информација.

Документација из области рударства, коју чини 172 документа, послужила је као подлога за истраживачки рад везан за екстракцију рударске терминологије, обележавање текста, екстракцију информација и др.

Окосницу дигиталне библиотеке чине пројекти и студије који се, почев од 2009. године, раде на Катедри за пројектовање и планирање површинских копова Рударско-геолошког факултета из Београда, а финансирани су од стране ЈП Електропривреда Србије из Београда. Као најзначајније треба напоменути 51 документ (пројектни задаци, студије и пројекти) из колекције *Студије и пројекти*, који чине "Главни рударски пројекат површинског копа "Дрмно" за капацитет од 9×10^6 тона угља годишње".

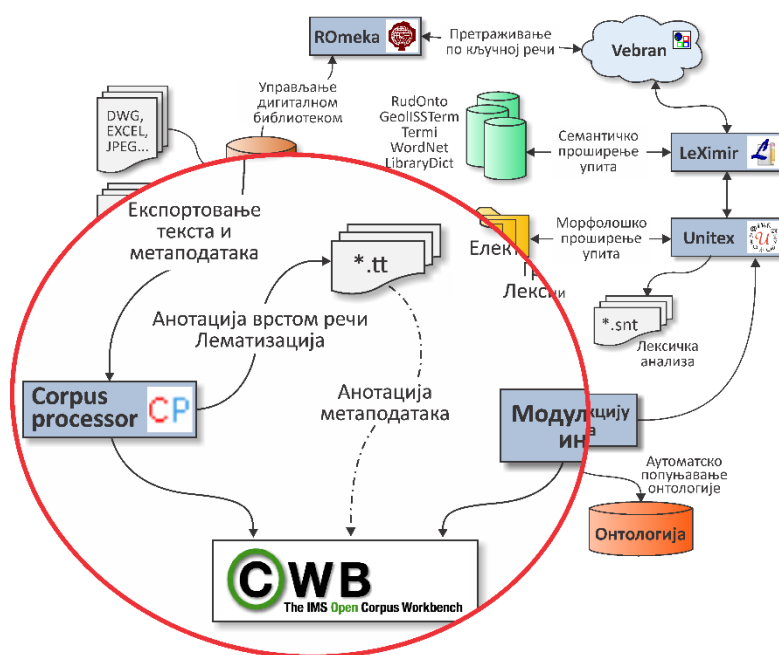
Табела 5.1. Структура колекција у дигиталној библиотеци

Колекција	Подколекције
Законска регулатива	Закони
	Правилници
	Уредбе
	Стратегије и смернице
Пројектна документација	Студије
	Пројекти
Стандарди	Међународни стандарди
	Домаћи стандарди
Литература	Монографије
	Уџбеници
	Докторске дисертације
	Радови
	Предавања

Нарочита пажња посвећена је безбедности и заштити докумената у дигиталној библиотеци, посебно са аспекта поштовања ауторских права и поверљивости података. Ово је у директној вези са улогом која се додељује корисницима. Већина докумената је јавно доступна, али је приступ пројектној документацији и неким публикацијама ограничен само за регистроване кориснике, како је већ поменуто, због поверљивости података и права публиковања. Регистрованим корисницима могу се доделити следеће улоге: *super users* (имају потпун приступ дигиталној библиотеци уз могућност измена њеног изгледа и функционалности, али и потпуни приступ објектима), *administrator users* (имају потпун приступ свим објектима дигиталне библиотеке), *contributor users* (могу креирати нове објекте и вршити измене само на њима) и *researcher users* (имају потпуни увид у све објекте и колекција дигиталне библиотеке али немају могућност измена садржаја) (Омека, 2017). Нерегистровани корисници могу прегледати само јавно доступне документе.

5.2. КОРПУС ТЕКСТОВА ИЗ ДОМЕНА РУДАРСТВА

Под корпусом се подразумева колекција текстова у електронској форми која је репрезентативна по: тематици, величини, функционалном стилу и временском периоду у коме су текстови настали. Репрезентативност одређеног корпуса зависи од његове намене (Ђорђевић, 2017). Позиција корпуса у систему за управљање рударском пројектном документацијом приказан је на слици 5.4.



Слика 5.4. Рударски корпус

5.2.1. Припрема корпуса

За потребе овог истраживања, формиран је корпус рударских текстова сачињен од докумената из домена рударства и заштите животне и радне средине. С обзиром да је овај корпус намењен специфичном доменском истраживању, он се сврстава у специјализоване корпуре, за разлику од општих корпуса који се формирају да би представљали основу за произвољно будуће, обично лингвистичко, истраживање (Utvić, 2014).

Сви прикупљени, систематизовани и метаподацима опремљени документи, ради даље обраде, морали су бити адекватно припремљени и преведени у облик који је погодан за претраге и анализе коришћењем одговарајућих софтверских алата.

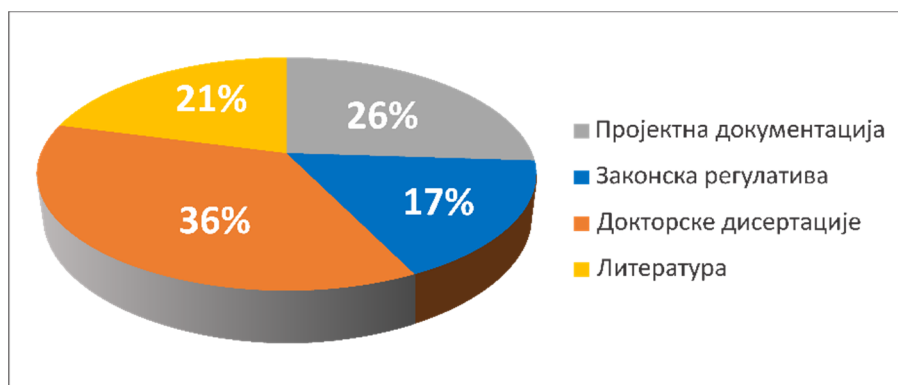
Стога је урађено оптичко препознавање текста и његово пречишћавање. Уклоњени су делови на страном језику, слике, референце и линкови. Формирана је једна текстуелна датотека величине 39 МВ, са око 6200 страна "чистог" текста формата А4. Текст је обрађен електронским речницима и локалним граматикама²⁷, при чему је поступак обраде текста и допуне речника итеративан процес. Након прве обраде текста добијено је: 150.365 реченица, 2.719.086 (100.414 различитих) речи. Око 1900 речи (не укључујући вишечлане термине), специфичних за области рударства, безбедности и заштите на раду и процене ризика, су обрађени и укључени у систем морфолошких речника (Krstev *et al.*, 2015), чиме је омогућена даља екстракција вишечланих термина према методологији описаној у раду (Stanković *et al.*, 2012).

Преглед структуре корпуса рударских текстова у односу на тип документа према броју докумената, броју реченица, броју речи, броју различитих речи, броју канонских облика речи (лема), величини текста у МВ и лексичкој разноврсности дат је у табели 5.2. Структура корпуса према броју речи за одређени тип текста приказана је на слици 5.5.

Табела 5.2. Резултати обраде текстова за рударски корпус

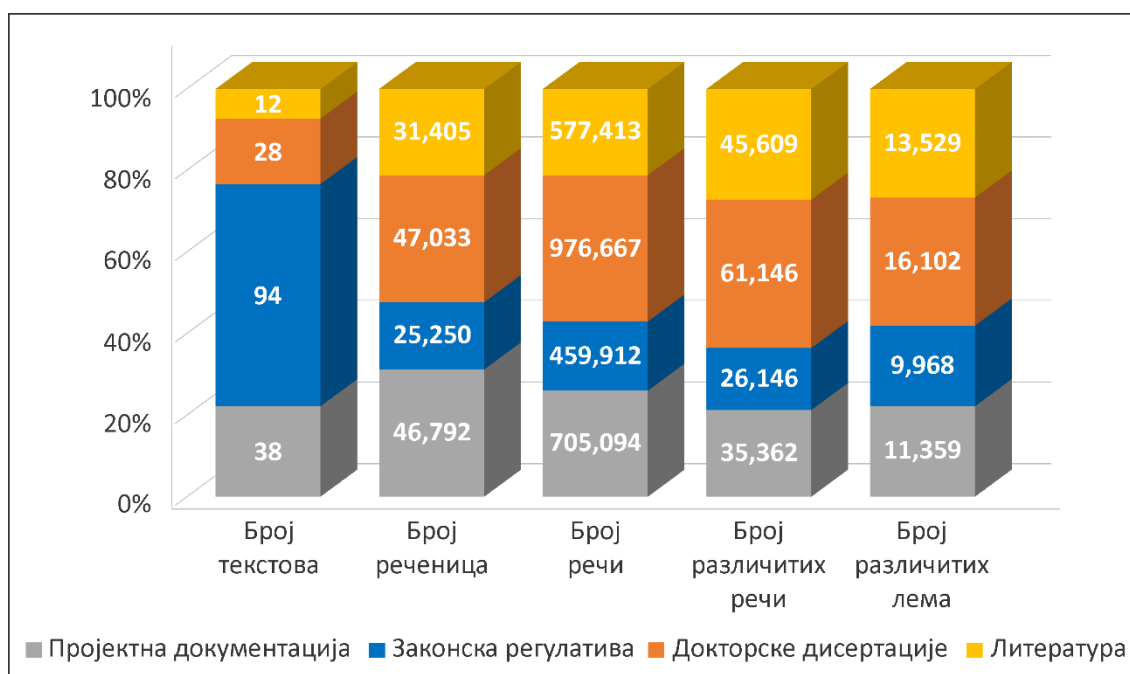
Врста текста	Број текстова	Број реченица	Број речи	Број различитих речи	Број различитих лема	Величина текста (МВ)	Индекс LR (%)
Пројектна документација	38	46.792	705.094	35.362	11.359	10,1	1,61
Законска регулатива	94	25.250	459.912	26.146	9.968	6,6	2,17
Докторске дисертације	28	47.033	976.667	61.146	16.102	14,2	1,65
Литература	12	31.405	577.413	45.609	13.529	8,2	2,34
Корпус	172	150.365	2.719.086	100.414	22.875	39,1	

²⁷ Локална граматика (енг. *local grammar*) је коначан аутомат који описује исправне (енг. *well-formed*) низове у тексту и означава их.



Слика 5.5. Расподела корпусних речи према типу текста

Лексичка разноврсност у најширем смислу дефинише се као опсег различитих речи коришћених у тексту (McCarthy & Jarvis, 2010) и може се изразити на неколико начина. У табели 5.2 дат индекс лексичке разноврсности (LR) рачунат као проценат броја различитих лема у односу на број речи. Уочава се да је индекс највећи код литературе, потом следи законска регулатива, дисертације и коначно пројектна документација, што је и очекиван резултат. Лексичка разноврсност текстова за рударски корпус је приказана на слици 5.6.



Слика 5.6. Лексичка разноврсност текстова за рударски корпус

Релативно мала лексичка разноврсност код расположиве пројектне документације је и била мотив да се корпус прошири и другим документима, како би се развила платформа и за будуће пројекте, за све области рударства.

5.2.2. Анотација и управљање корпусом

У развоју рударског корпуса, коришћени су два различита система за креирање верзија корпуса за два различита сценарија коришћења. Први је креиран помоћу CWB (IMS Open Corpus Workbench), колекције алата отвореног кода за креирање и управљање корпусом (Evert & Hardie, 2011) и додатка CQPweb, веб апликације развијене специјално за CWB процесор упита - CQP, као и за онлајн управљање корпусом (Hardie, 2012). У десктоп окружењу, Unitex (Paumier, 2016) је коришћен за креирање другог корпуса од истих текстова за решавање различитих задатака: екстракција информација, екстракција термина, претраживање и развој локалних граматика.

CWB је погодан за кодирање, индексирање, компресију и декодирање великих текстуалних корпуса (до 2 милијарде речи), са вишеслојном анотацијом на нивоу речи. CQP је моћан и ефикасан систем за генерисање конкорданци²⁸ који може да обрађује шаблоне (упите, односно обрасце) (енг. *query patterns*) специфициране како на нивоу карактера (наводећи облик појединачне речи или анотацију), тако и на нивоу токена²⁹ (одређујући синтаксичке везе између токена). Кроз CQP веб корисници могу навести шаблоне упита и добити статистичке информације о корпусу.

CWB-верзија корпуса је анотирана. Свакој речи у корпусу су додељиване речи у следећем редоследу:

- анотација врстом речи (именица, глагол, придев, итд.) - PoS tagging (Utvić, 2011);
- лема (номинатив једине за именицу, инфинитив за глагол, итд.) – придружена поступком који се зове лематизација.

²⁸ Конкорданца је формат у ком се најчешће презентују аутоматски пронађени примери употребе дате речи у контексту: (леви контекст) → (кључна реч) → (десни контекст)

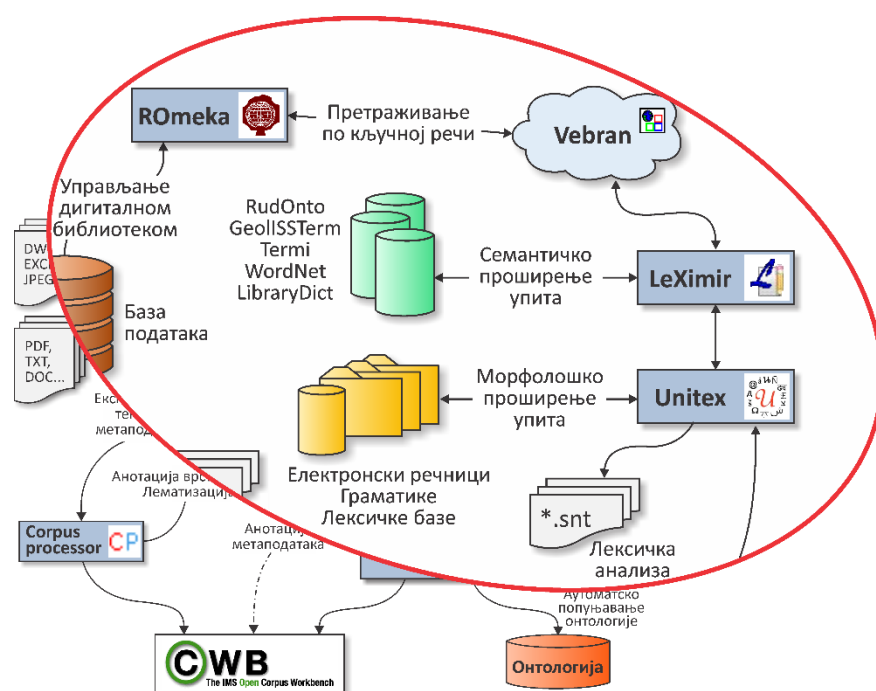
²⁹ Језичка јединица, најчешће реч

У зависности од величине корпуса, анотација може бити ручна (за мале корпусе) и аутоматска (за велике корпусе). Уобичајено је да се комбинују оба типа анотација, тј. прво се ручно аотира скуп за обучавање, а онда се тренира тагер³⁰ (енг. *tagger*), на пример PoS (енг. *Part of Speech*), за обележавање врста речи и након обучавања се користи за аутоматско аотирање остатка корпуса. Ако је потребно, грешке аотације могу се ручно исправити, након аутоматског аотирања, што се најчешће врши селективно. У нашем истраживању примењена је аутоматска анотација програмом TreeTagger (Schmid, 1999, Schmid, 2013) и српском параметарском датотеком. Српска параметарска датотека TreeTagger-а је генерисана коришћењем морфолошких електронских речника за српски језик и обучавањем заснованим на ручно аотираном корпусу који садржи милион речи (Utvić, 2011). С обзиром да је скуп за обуку ручно аотиран само на нивоу врсте речи и лема, они представљају једину морфолошку информацију којом би корпус могао бити означен. Прецизније, сваки токен корпуса индексиран у CWB-у има три атрибута: *реч* (токен вредност, флективни облик речи), *PoS* (врсту речи) и *лему*. Тренутно атрибут *PoS*, односно скуп етикета корпуса, има 16 могућих вредности (10 ознака за врсту речи на српском, као и скраћенице, римске бројеве, интерпункције, префиксе, суфиксе итд.), на пример: N (*Noun*, именица), PRO (*Pronoun*, заменица), A (*Adjective*, придев), V (*Verb*, глагол) итд. За управљање Unitech корпусом узета је у обзир додатна анотација, а то је вредност флективних категорија (граматички род, број, падеж, глаголски облик итд.) и семантички маркери који омогућавају анализу текста за све морфолошке, доменске и семантичке вредности дате у е-речницима. Међутим, треба напоменути да у овом процесу сам корпус није интерно означен. Уместо тога, Unitech креира речник корпуса на основу текста корпуса, са аотацијама из система е-речника за српски језик, а затим користи пресек речника корпуса и текста корпуса за претрагу и екстракцију конкорданци из корпуса користећи коначне трансдукторе.

³⁰ Тагери су програми који врше обележавање информација у тексту. Оне могу бити: опште библиографске информације (аутор, функционални стил, процесу дигитализације и др.), информације о логичкој структури текста (почетку и крају поглавља, параграфа и реченица), језичке информације (морфосинтаксичке категорије речи у тексту, њихова функција или значење) (Ђорђевић, 2017).

5.3. ЛЕКСИЧКИ И ТЕРМИНОЛОШКИ РЕСУРСИ

Посебна пажња код осмишљавања система за управљање рударском пројектном документацијом, посвећена је допуни постојећих лексичких ресурса терминима из области рударства. Позиција лексичких и термиолошких ресурса у систему за управљање рударском пројектном документацијом приказана је на слици 5.7.



Слика 5.7. Лексички и термиолошки ресурси

Термиолошки ресурси, који су имплементирани у приказани систем, су: *RudOnto*, *GeolISSTerm*, *Termi*, *Serbian WordNet* и *LibraryDict*. *RudOnto*, као термиолошки ресурс, садржи терминологију, на српском и енглеском језику, из домена заштите на раду и рударства, али је отворен за проширење терминима и дефиницијама термина из других сродних области (Kolonja, 2016). *GeolISSTerm* је организован као таксономија са: дефиницијама за сваки општеприхваћени геолошки термин, синонимима и библиографским референцама извора из којег су преузети, као и еквивалентним терминима на енглеском језику (Stanković *et al.*, 2011b). *Termi*³¹ је термиолошки ресурс који подржава развој термиолошких речника у разним научним областима (математика, рачунарске науке, рударство,

³¹ <http://termi.rgf.bg.ac.rs/>

библиотекарска наука, рачунарска лингвистика...). Подржава обраду и приказ термина на српском и енглеском језику, а у плану је развој приказа на другим језицима. Као лексички ресурс, *Serbian WordNet* се примењује у истраживањима вишечланих лексичких јединица, класификацији, претраживању вишејезичних дигиталних база података, препознавању реторичких фигура и др. (Mladenović, 2016). *LibraryDict*³² обухвата терминологију из теорије и праксе библиотекарства и информационих наука и широког спектра блиских или сродних области, на српском, енглеском и немачком језику. Језици у овом речнику имају једнак статус. Online верзија овог речника садржи: око 40.000 термина (око 14.000 на српском језику, око 12.400 термина на енглеском језику и око 14.000 на немачком језику); 900 дефиниција и објашњења термина; 2.300 акронима домаћих и међународних ентитета; 190 линкова релевантних веб страна.

Систем српских морфолошких речника обухвата општу лексику, властита имена и одређене доменски специфичне термине са флективним облицима генерисаним од ~140.000 једночланих речи и ~18.000 вишечланих лексичких јединица. Термина који се директно односе на домен рударства јако је мало и углавном су општег типа. Овај систем обухвата једночлане речи (енг. *simple words*), вишечлане лексичке јединице (енг. *multi-word units*, MWUs) и коначне трансдукторе (енг. *finite-state transducers*, FSTs), који се користе за екстракцију информација и анотацију и обележавање текста у оквиру различитих апликација природног језика (Krstev *et al.*, 2008a).

Сваки морфолошки речник се састоји од листе лексичких записа, који могу садржати: облик речи, лему, морфосинтаксичке, семантичке, доменске и друге информације. Речник канонских облика речи, односно лема је следећег облика:

`lemma, POS#fst [+Marker]*`

где lemma означава канонски облик речи (једночлане или вишечлане лексичке јединице), POS#fst је кôд флективне класе који се састоји од врсте речи и редног броја флективног трансдуктора³³ (генеративне граматике)

³² Dictionary of Library and Information Sciences Serbian-English-German <http://rbi.nb.rs/en/dict.html>

³³ Трансдуктори дефинишу релације између два скупа ниски карактера, односно трансформишу задату ниску у другу ниску над истом или неком другом азбуком; препознају једну ниску карактера а генеришу другу. <http://jerteh.rs/wp-content/uploads/2015/12/Jacimovic.pdf>

којим се производе сви граматички исправни облици леме. У наставку се може наћи више маркера. који ће бити детаљније описани у поглављу 5.5.2, а у изразу су означени са: [+Marker]*.

Већина флективних облика речи у српским морфолошким е-речницима су снабдевени не само вредностима граматичких категорија, већ и додатним маркерима који су наслеђени из записа лема из којих се генеришу. Ови маркери могу бити граматички, деривациони, доменски и семантички. На пример: лема глагола "rudariti" (у руднику, ископати) из речника канонских форми има следећи облик:

`rudariti,V551+Imperf+It+Iref+DOM=Mining`

где маркери имају следећа значења: +Imperf = несвршени глагол, +It = прелазни глагол, +Iref = неповратни глагол и +DOM = Mining значи да је глагол из области рударства. Од леме трансдуктор производи на излазу све флективне облике у следећем формату:

`form,lemma[:categories]*`

Дакле, иза облика се раздвојена запетом налази лема и опционо, иза две тачке листа граматичких категорија које одговарају наведеном облику. У претходном примеру, за лему "rudariti" која припада флективној класи V551 (Krstev, 2008), производи се речних флективних облика у следећем формату:

```
rudarim,rudariti.V:Pxsat
rudarimo,rudariti.V:Pxp:Yxp
rudarila,rudariti.V:Gpn:Gsf:Gwm:Gwn
rudarile,rudariti.V:Gpf:Gwf
rudarismo,rudariti.V:Axp
[...]
```

Грамматичке категорије повезане са формама су: P - садашње време, G - перфекат, Y - императив, A - аорист, s - једнина, p - множина, x - прво лице једине, а – позитив (основни облик придева), n - средњи род, f - женски род, m - мушки род, w - паукал³⁴.

³⁴ Паукал је облик који иде уз бројеве два, три и четири, остатак је старе словенске двојине.

5.3.1. Обележавање речничких одредница

Маркери су носиоци информација од изузетног значаја за проналажење и екстракцију информација из текстуалних ресурса. Маркери наведени у речницима означавају домен употребе дате речи или их ближе одређују по свом значењу, тако да могу бити доменски, семантички или синтактички. Тако, на пример, маркер +Nun означава нешто живо, маркер +Org означава врсту организације итд. Значај маркера се огледа у томе што они у комбинацији са граматичким ознакама ближе дефинишу речи. Њима се прави финија дистинкцију међу речима неопходна за прецизнију екстракцију информација.

Маркере је могуће користити кроз регуларне изразе, за неке једноставније упите над корпусом или кроз постављање упита конструкцијом аутомата са сложенијим захтевима. Пример истраживања при коме су коришћени маркери јесте систем NERosetta за екстракцију именованих ентитета NERosetta (назива установа, личних имена, улица...) (Krstev *et al.*, 2013). Употреба маркера могућа је и за потребе обраде текста ради екстракције доменске терминологије, мерних јединица, успостављање релација међу речима... (Obradović *et al.*, 2017)

Рударство је веома сложена индустријска грана, са активностима које се међусобно прожимају и надопуњују, али се истовремено и веома разликују. У (Krstev, 2008) дефинисан је општи домен рударства $DOM=+Mining$, а у овом раду се уводи детаљније обележавање рударских термина

5.3.2. Рударски доменски маркери

Пракса је показала да је потребно увести финију поделу домена на гране рударства, као и специфичне семантичке маркере и тиме омогућити прецизнију екстракцију информација из текстова рударског домена. Како би свака од рударских активности била ближе описана и како би се омогућила прецизнија екстракција информација из текстова рударског домена, дефинисани су нови доменски и поддоменски маркери.

У табели 5.3 приказани су постојећи и новопредложени доменски маркери значајни за област рударства.

Табела 5.3. Доменски маркери из области рударства

Маркер	Опис	Статус	Примери
+Mining	Рударство	Постојећи маркер	рудник, рудар, руда, експлоатација, угаљ, минерална сировина, руда бакра, етажа, јаловина, окно, бушење, цевовод, рударска мерења, вентилација, нафта, бушотина, одводњавање, млевење,...
+Mach	Машинство	Постојећи маркер	роторни багер, багер ведричар, камион, дозер, рипер, транспортна трака, цевополагач, железница, одлагач, транспортни мост, грејдер, комбајн, хидромонитор, депонијска машина,...
+Safety	Заштита на раду	Постојећи маркер са предлогом промене назива +Safe	вентилација, заштитна опрема, бука, вибрације, минерална прашина, штетни гасови, повреда на раду, професионално обољење, пожар, заштита од пожара, експлозија, служба спасавања,...
+Transport	Транспорт	Предложен нови маркер	транспортна трака, камион, дампер, хидраулички транспорт, извоз, извозно постројење, самоходно транспортно средство,...
+RockMech	Механика стена	Предложен нови маркер	притисак, смицање, напонско стање, носивост стене, физичко-механичке особине, деформабилност стена, деформабилност тла, стабилност косина, носивост тла,...
+Surveying	Геодезија	Предложен нови маркер	рудничка мрежа, јамски полигон, нивелмански влак, рударска viseћа бусола, теодолит, жиротеодолит,...
+EnvProt	Заштита животне средине	Предложен нови маркер	аерозагађење, бука, загађење, мониторинг воде мониторинг ваздуха мониторинг земљишта, ремедијација, рекултивација, еколошки аудит, одлагање отпада, ...

Ради прецизнијег описивања рударских термина предложена је и листа поддоменских маркера. Њима би се ближе дефинисале четири велике области унутар рударског домена (површинска експлоатација, подземна експлоатација, припрема минералних сировина и експлоатација нафте и гаса), које се разликују у толикој мери да оправдавају увођење додатних маркера. Поддоменски маркери би се придруживали доменском маркеру, па би тако, на пример, маркер +DOM=Mining+DOM=Surface означавао појмове који припадају домену рударства и поддомену површинска експлоатација. У табели 5.4 приказани су предложени поддоменски маркери значајни за област рударства (у речницима уместо +Маркер се бележи +DOM=Маркер).

Табела 5.4. Поддоменски маркери из области рударства

Маркер	Опис	Примери
+Surface	Површинска експлоатација лежишта минералних сировина	површинска експлоатација, површински истражни радови, етажа, косина, завршна косина, стабилност косина, откривка, коефицијент откривке, БТО систем, депонија, минирање, бушење, граница копа, јаловина, рекултивација, нагиб косине, багер,...
+Underground	Подземна експлоатација лежишта минералних сировина	подземна експлоатација, подземни истражни радови, окно, ходник, ускоп, нископ, сипка, бункер, заштитни стуб, широко чело, вентилација,...
+MinProcess	Припрема минералних сировина	кретање масе, узорак, уситњавање, дробљење, сито, млевање, млин, класирање, концентрација, флотација, флотацијска пулпа, згушњавања, лужење, центрифугирање, депоновање,...
+Petroleum	Експлоатација нафте и гаса	нафта, гас, угљоводонични флуид, бушотина, цевовод, динамика протока, гас лифт, гасоводни систем,...

5.3.3. Рударски семантички маркери

Како би се омогућила екстракција специфичних концепата и релација међу концептима креирањем лексичких маски, предложени су и нови семантички маркери значајни за област рударства. Листа ових маркера, са описом и примерима приказана је у табели 5.5. Тако на пример сложени термини: роторни багер, вентилација рудника, јамска просторија, означили би се маркерима на следећи начин:

роторни багер	+Mining+Surface+Instrum
вентилација рудника	+Mining+Underground+Activity
јамска просторија	+Mining+Underground +Object

Инспирација за семантичке маркере је био *EarthResourceML* (EarthResourceML), стандард за размену дигиталних информација заснован на XML-у за минералне појаве, ресурсе и резерве, потом руднике и рударске активности, као и производњу концентрата, излазних производа, рударског отпада. Речници који су послужили за креирање семантичких маркера су: *Mine Status*, *Mining Activity*, *Processing Activity*, *Exploration Activity Type*, *Raw Material Role*, *Reserve Category*, *Resource Category*, *Exploration Result*, *Mineral Occurrence Type* и *Environmental Impact*. Ови речници су јавно доступни као RDF и XML датотеке, а постоји и SPARQL приступна тачка за RESTful APIs. У оквиру лексичке базе је предвиђено мапирање маркера из домена рударства са поменутиим стандардним речницима (Stanković *et al.*, 2018).

Табела 5.5. Семантички маркери за област рударства

Маркер	Опис	Примери
+MinStatus	Статус рудник	активан рудник, неактиван рудник, затворен рудник, напуштен рудник, конзервиран рудник, рудник у развоју,...
+Ore	Минералне сировине	угаљ, лигнит, мрки угаљ, камени угаљ, руда гвожђа, хематит, магнетит, руда бакра, халкопирит, ковелин, руда олова, галенит, нафта, гас, геотермална вода, шљунак, песак, камен,...
+Activity	Рударска активност	пројектовање рудника, минирање, бушење, експлоатација минералне сировина, вентилација рудника, одлагање откопаног материјала, дробљење руде, млевење руде, уситњавање руде,...
+Object	Рударски објекти	окно, ходник, ускоп, нископ, бункер, усек, поткоп, засек, јама, дробилана, бушотина, хоризонт, сипка, јамска просторија,...
+Prof+Hum	Професије	рударски инжењер, рудар, геолог, мерач, рударски техничар, геолошки техничар, багериста, копач, минер, површинац, подземљаш, припремаш, помоћни радник,...
+Org	Организације	Колубара, Костолац, Рудник, РТБ Бор, Дрмно, ЕПС,...
+Instrum	Рударска опрема	багер, камион, транспортна трака, дозер, утоварач, скрепер, концентратор, теодолит, бушаћи чекић, дробилица, одлагач, дреглајн, додавач,...
+Exploration	Истраживање минералних сировина	експлоатационо истраживање, површински истражни радови, подземни истражни радови, истражно бушење,...
+Conc+Fashion	Средства личне заштите	лампа, самоспасилац, маска, шлем, чизме, заштитна опрема, заштитно одело,...
+Text	Типови рударских докумената	рударска пројектна документација, геолошки елаборат, претходна студија оправданости, студија оправданости, рударски пројекат, технички пројекат,...

5.4. ПРЕТРАЖИВАЊЕ, ПРЕЛИСТАВАЊЕ, ПРОНАЛАЖЕЊЕ И ЕКСТРАКЦИЈА ИНФОРМАЦИЈА

Системи за проналажење информација базирани су на два основна концепта:

- објект - ентитет који садржи тражене информације, на пример документи и веб стране,
- упит - формални захтев за потребним информацијама које корисник уноси у систем за претраживање информација, које су најчешће ускладиштене у базама података.

У зависности од ресурса који се претражује, одговори могу бити, на пример, листа докумената, листа веб страница, конкорданце (алфабетски уређена листа облика речи екстрахованих из неког текста која уз сваки елемент укључује и упућује на његову позицију у изворном тексту, као и на изабрани део текста испред и иза датог облика речи) или пак издвојене реченице у паралелним вишејезичним текстовима.

Једноставно проналажење информација у текстуалним ресурсима се имплементира као *string matching*, не узимајући у обзир синтаксичка или семантичка својства тражене речи. Код претраживања дигиталне библиотеке резултат овакве претраге би био листа докумената.

Када је у питању претраживање корпуса, формулисање сложенијих упита је омогућено преко проширених регуларних израза, а резултат претраге је листа конкорданци.

Након имплементације система за проналажење информација, било ког типа, и генерално, код развоја нових функција претраживања, веома је важно проценити не само колико је добро развијено решење и колико је скуп враћених резултата свеобухватан, већ и колико је у њему одговора који одговарају полазној информационој потреби. Нарочиту пажњу треба обратити на чињеницу да се може догодити да модел превише прецизан, да су сви резултати тачни, али да изостави неке резултате који могу бити релевантни.

Процес евалуације је поређење унапред познатих одговора (резултата) са онима које је вратио систем. Тиме се добијају исправно и неисправно понуђени одговори, односно враћени документи или одзиви на упит. Могући исходи су (слика 5.8):

- издвојени релевантни документи (Ir),
- издвојени нерелевантни документи (In),
- неиздвојени релевантни документи (Nr) и
- неиздвојени нерелевантни документи (Nn).

При евалуацији претраживања се морају рачунати основне мере ефикасности:

- одзив (енг. *recall*),
- прецизност (енг. *precision*),
- Ф-мера (енг. *F-measure*) и
- рангирање (енг. *ranking*).

Прецизност је проценат примера који су исправно класификовани као релевантни за постављени упит међу свим примерима који су понуђени. Прецизност, дакле, изражава меру тачности одговора који су добијени на постављени упит и представља количник укупног броја релевантних пронађених докумената и укупног броја пронађених (и кориснику понуђених) одговора.

Израчунава се по формули (Baeza-Yates & Ribeiro-Neto, 1999):

$$P = \frac{Ir}{Ir + In} \quad (5.1)$$

где је: P – прецизност,

Ir – издвојени релевантни документи,

In – издвојени нерелевантни документи,

Ir+In – укупан број одговора на упит.

Одзив изражава меру потпуности одговора који су добијени на постављени упит и изражава се као количник укупног броја релевантних пронађених докумената и укупног броја релевантних докумената.

Израчунава се по формули (Baeza-Yates & Ribeiro-Neto, 1999):

$$O = \frac{I_r}{I_r + N_r} \quad (5.2)$$

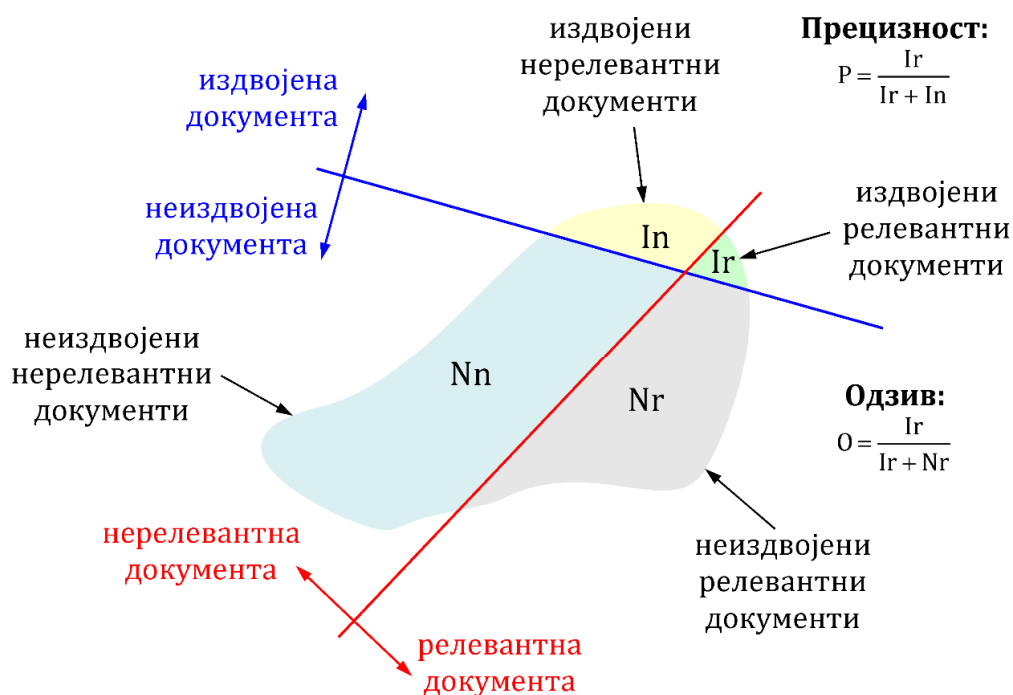
где је: O – одзив,

I_r – екстраховани релевантни документи,

N_r – неиздвојени релевантни документи,

$I_r + N_r$ – укупан број одговора релевантних за постављени упит.

Одзив указује на то колико је систем свеобухватан током претраживања релевантних информација, док прецизност указује на тачност с којом систем ради. Илустрација издвајања релевантних и нерелевантних докумената је илустрована сликом 5.8, где је приказана подела скупа докумената на четири дела. Скупови I_r и N_n би требало да буду што већи, а скупови I_n и N_r треба да буду што мањи код ефикасне претраге.



Слика 5.8. Подела скупа докумената упитом³⁵

Однос између одзива и прецизности показује њихову тенденцију да буду обрнуто пропорционални (ако је одзив система већи, то ће утицати на

³⁵ http://poincare.matf.bg.ac.rs/~cvetana/Nastava/Materijal/IR_05.ppt

смањење његове прецизности; и обрнуто, ако је одговор система прецизнији, одзив ће се смањити). У пракси се ове две мере комбинују кроз тзв. Ф-меру на следећи начин (Jurafsky & Martin, 2014, Stanković, 2009):

$$F_{\beta} = (1 + \beta^2) \cdot \frac{P \cdot O}{\beta^2 \cdot P + O} \quad (5.3)$$

где је: P – прецизност

O – одзив

β - параметар који регулише однос између прецизности и одзива:

ако је $\beta = 1$, прецизност и одзив имају исту тежину,

ако је $\beta > 1$ већу тежину има прецизност,

ако је $\beta < 1$ већу тежину има одзив.

У пракси се често користи опција прецизност и одзив имају исту тежину ($\beta = 1$), који онда даје хармонијску средину прецизности и одзива, па формула (5.3) добија облик:

$$F_{\beta} = \frac{2 \cdot P \cdot O}{P + O} \quad (5.4)$$

Хармонијска средина нагиње мањој вредности и она тежи да неутралише утицај већег, а потенцира значај мањег.

Тачност (Т) је мера успешности класификације докумената према релевантности. Овом мером се утврђује колики део докумената је исправно класификован као релевантан односно нерелевантан.

$$T = \frac{I_r + N_n}{I_r + I_n + N_r + N_n} \quad (5.5)$$

где је: A – издвојени релевантни документи,

B – издвојени нерелевантни документи,

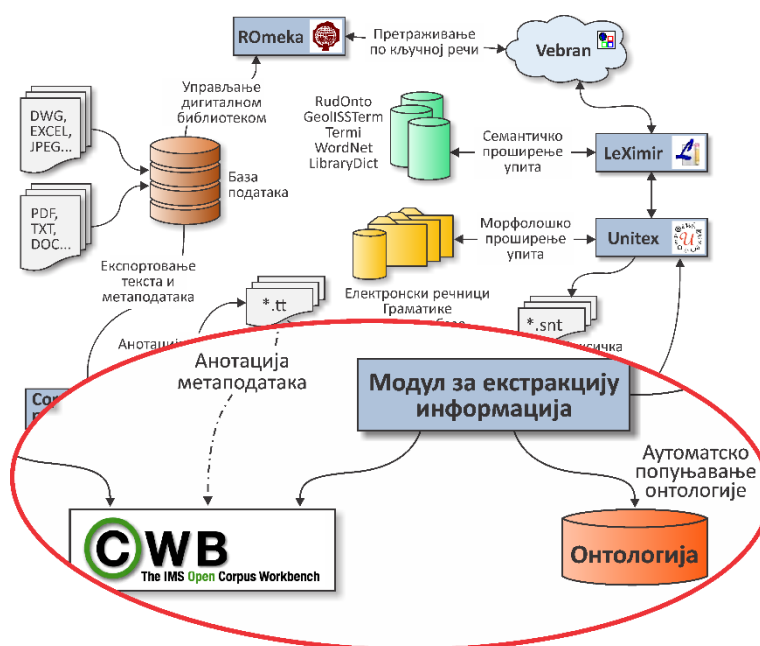
C – неиздвојени релевантни документи,

D – неиздвојени нерелевантни документи,

S – сви документи посматраног скупа.

5.5. ИЗГРАДЊА ОНТОЛОГИЈЕ

Рударска доменска онтологија *RuDokOnto* изграђена је са циљем да прикупља, описује и систематизује рударску документацију по фазама пројекта, као и да се повезује са другим сродним онтологијама, као што су на пример *EarthResource*, *MinExOnt* и др. Њена позиција у систему за управљање рударском пројектном документацијом приказана је на слици 5.9.



Слика 5.9. Позиција онтологије у систему

5.5.1. Protégé – алат за изградњу онтологија

Онтолошки едитори представљају алате који се користе за креирање и одржавање како нових, тако и постојећих онтологија. За потребе овог рада одабран је едитор онтологија Protégé 5.2, који се развија на Stanford University (USA) (Horridge *et al.*, 2004). Protégé је интегрисани софтверски алат отвореног кода, за развој система заснованих на знању, који омогућује паралелан рад са класама и инстанцама. Изградња онтологија у Protégé едитору састоји се из следећих корака (Musen, 2015):

- изградња иницијалне, мале онтологије која ће даље бити надограђивана,
- креирање алата за сакупљање знања (енг. *knowledge-acquisition*),

- унос специфичних инстанци података и креирање базе знања,
- извршавање апликације и њено тестирање.

Protégé омогућује додавање постојеће онтологије другој која се тек развија. На тај начин се од више мањих онтологија може саставити једна велика онтологија. Поступак креирања онтологија биће објашњен у наредном поглављу, у којем се описује развој рударске онтологије.

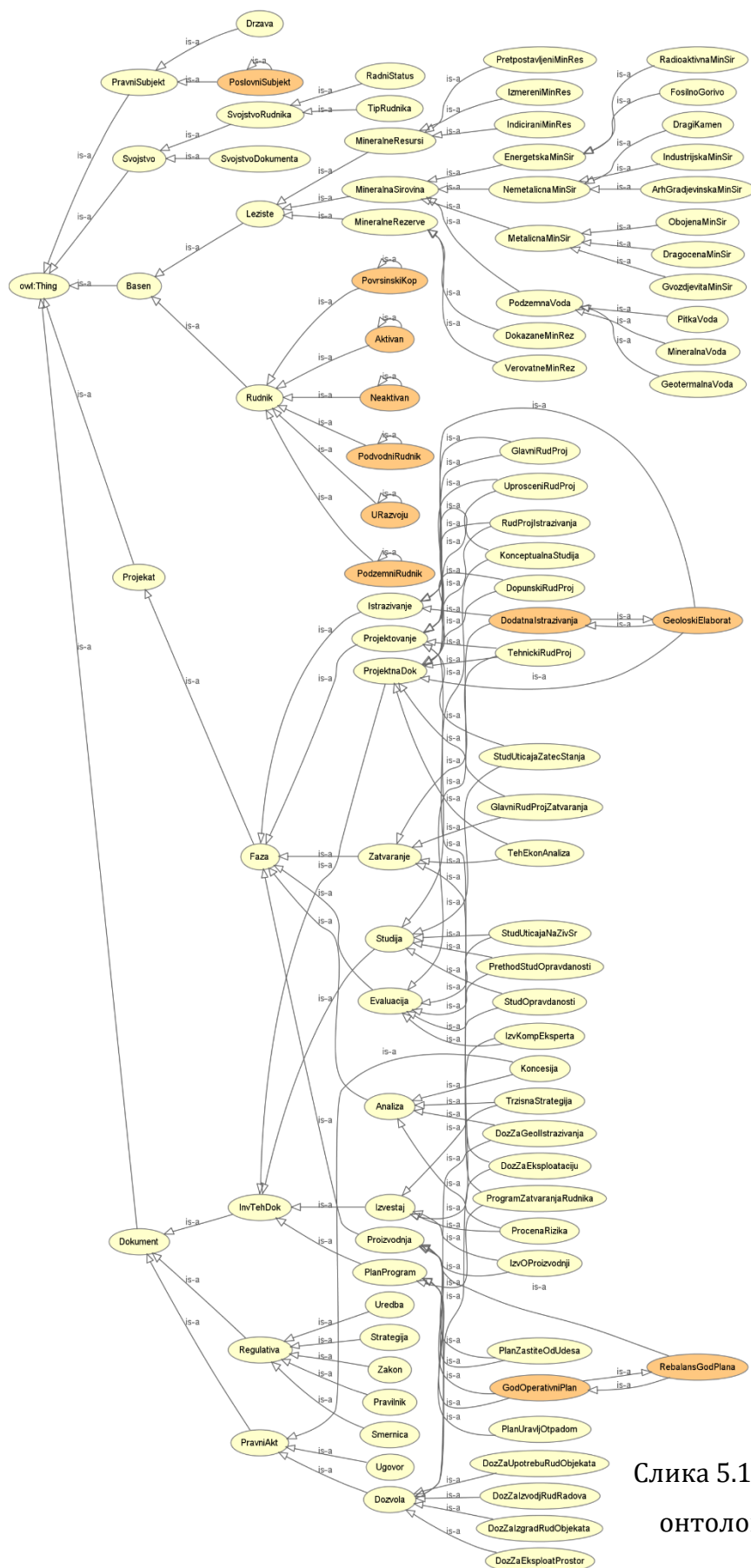
5.5.2. Доменска онтологија *RuDokOnto*

Техника моделовања која је примењена у овом раду је техника "одозго-наниже" (енг. *top-down*). Онтологија *RuDokOnto* креирана је помоћу алата Protégé 5.2. Визуелизација онтологије рађена у програмском додатку OWLViz и приказана је на слици 5.10.

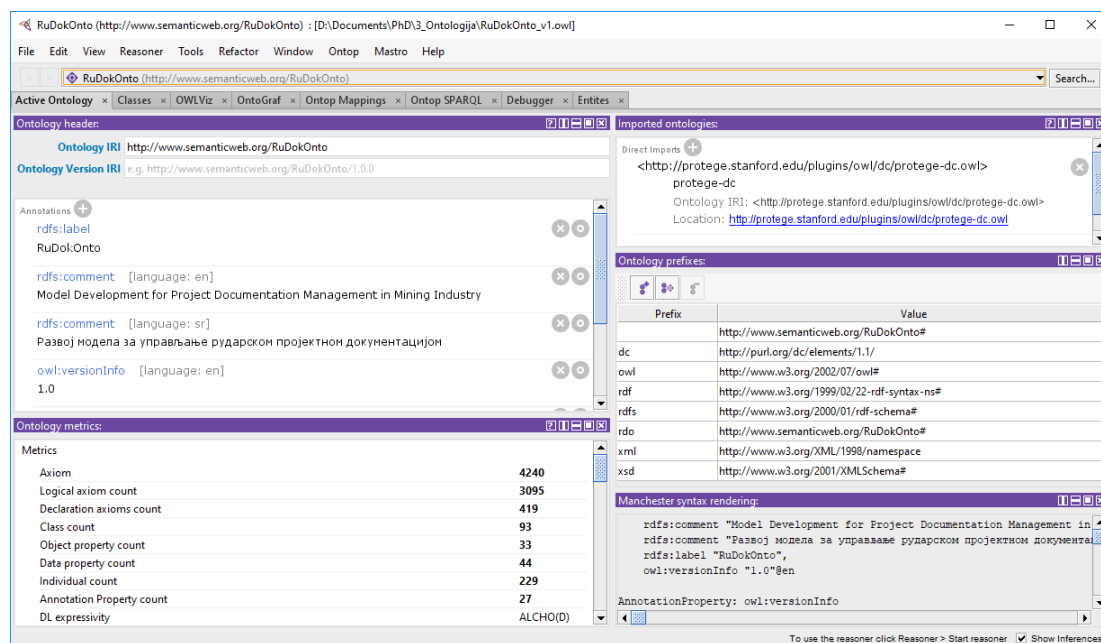
Дефинисање и креирање онтологије одвијало се по следећим корацима (Petrušić, 2016):

1. дефинисање класа,
2. одређивање хијерархије класа,
3. дефинисање особина класа,
4. успостављање релација,
5. креирање примерака класа.

На самом почетку креирања онтологије, у делу *Ontology header*, дефинисани су општи подаци о онтологији: IRI, назив онтологије, коментари који је ближе описују, верзија, аутори итд. Прозор *Imported ontologies* приказује увезене онтологије. У овом случају, то је онтологија *protege-dc.owl*, која омогућава повезивање мета-података Даблинског језгра објеката похрањених у дигиталну библиотеку *RuDokOnto*, са онтологијом. Такође, приказани су и статистички подаци (*Ontology metrics*), као префикси који омогућавају повезивање са стандардним онтологијама, а који се користе за креирање упита. Почетни панел (картица *Active Ontology*), са општим подацима, приказан је на слици 5.11.



Слика 5.10. Визуелизација онтологије RuDokOnto



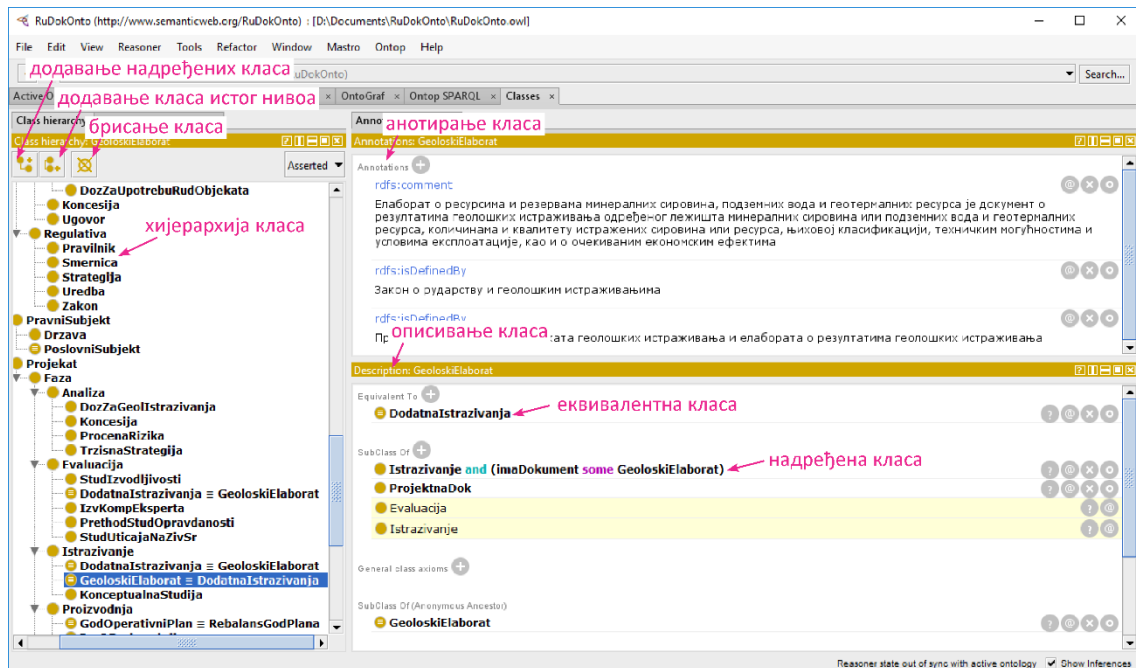
Слика 5.11. Креирање онтологије

5.5.2.1. Класе

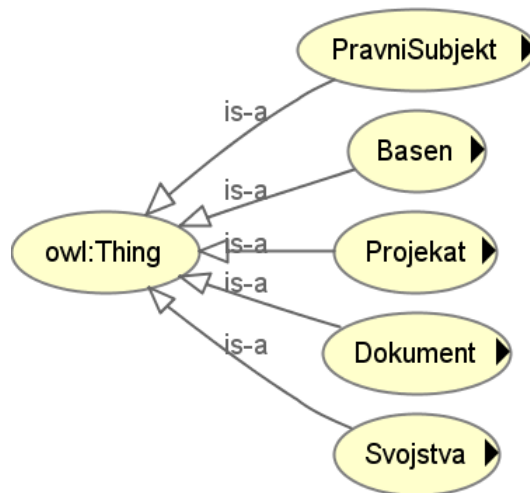
Први корак у креирању онтологије је формирање класа, којима се дефинишу сви концепти у домену. На картици *Class hierarchy* врши се додавање класа и подређених класа, као и њихово брисање. Све класе приказане су хијерархијски, у облику стабла. У делу *Annotations* класе се детаљно могу описати метаподацима, а у делу *Description* дефинишу се еквивалентне класе, надређене класе, примерци односно инстанце класа и дисјунктне класе (енг. *disjoint class*). Панел за креирање класа приказан је на слици 5.12.

Основна класа, надређена свим класама у онтологији је *owl:Thing*. У овом случају њено име је промењено у *RuDokOnto*. Онтологија садржи четири класе, то су: *PravniSubjekt*, *Basen*, *Dokument* и *Projekat* са великим бројем подређених класа. Таксономија онтологије приказана је на слици 5.13.

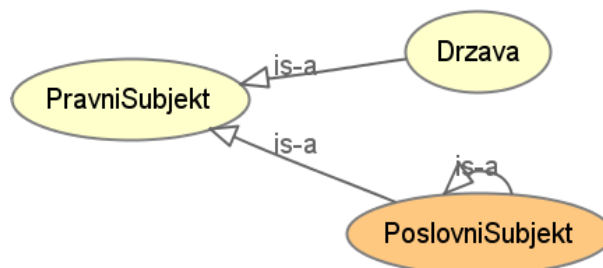
Класа *PravniSubjekt* описује све правне субјекте ангажоване на пројекту. То су Држава, са својим ресорним министарствима, која је власник рудног богатства, и пословни субјекти који су ангажовани на пројекту: компаније које су власници и/или инвеститори неког пројекта, институције овлашћене за геолошка истраживања и/или пројектовање и др. Сходно томе, ова класа садржи две подкласе: *Drzava* и *PoslovniSubjekt* (слика 5.14).



Слика 5.12. Креирање класа

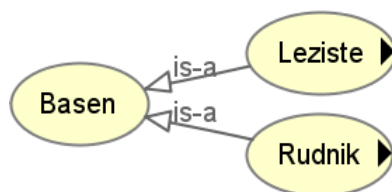


Слика 5.13. Таксономија онтологије RuDokOnto



Слика 5.14. Таксономија класе PravniSubjekt

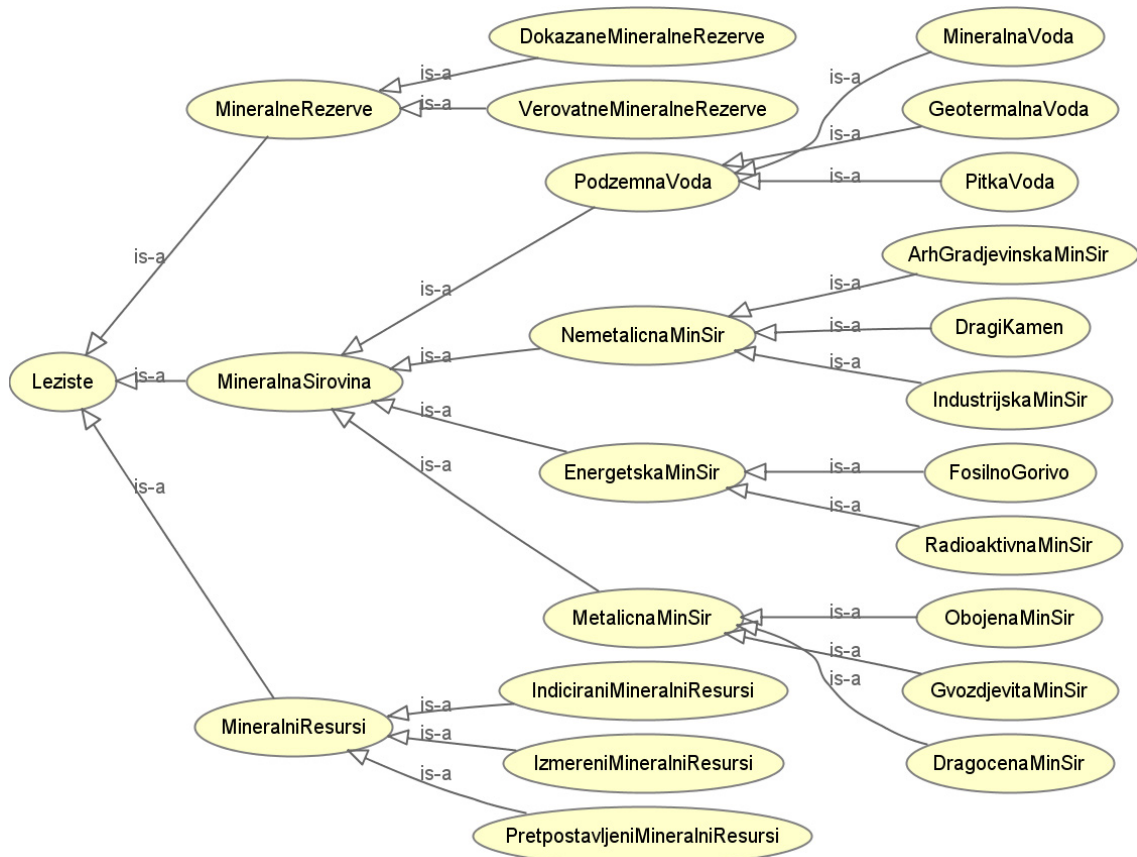
Класа *Basen* описује рударски басен и има две подређене класе: *Leziste* и *Rudnik* (слика 5.15).



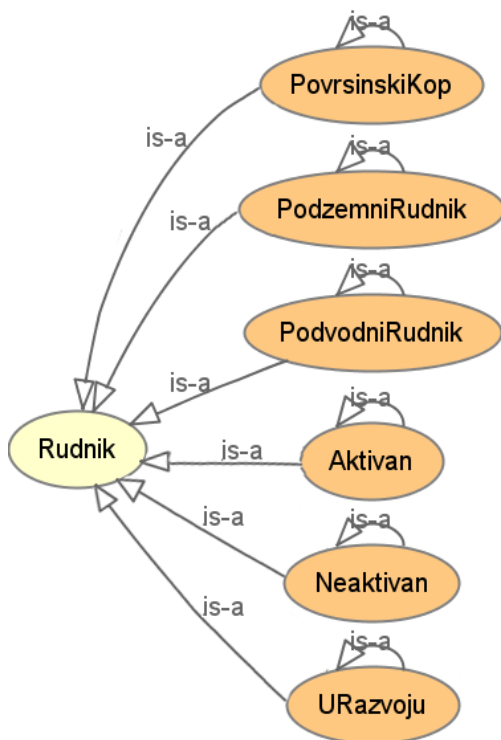
Слика 5.15. Таксономија класе *Basen*

Класа *Leziste* има три подкласе *MineralniResursi*, *MineralneRezerve* и *MineralnaSirovina*. Класа *MineralniResursi* се дели на подкласе: *PretpostavljeniMineralniResursi*, *IndiciraniMineralniResursi* и *IzmereniMineralniResursi*. Класа *MineralneRezerve* се дели на: *VerovatneMineralneRezerve* и *DokazaneMineralneRezerve*. Класа *MineralnaSirovina* се дели на подкласе: *EnergetskaMineralnaSirovina*, *MetalicnaMineralnaSirovina*, *NemetalicnaMineralnaSirovina* и *PodzemneVode*. Свака од ових подкласа даље је подељена на нове подкласе у складу важећим међународним стандардима и домаћим прописима. Таксономија класе *Leziste* приказана је на слици 5.16.

Класа *Rudnik* описује руднике према радном статусу (подкласе *AktivanRudnik*, *NeaktivanRudnik* или *RudnikURazvoju*) и према начину експлоатације (*PovrsinskiRudnik*, *PodzemniRudnik* и *PodvodniRudnik*). Ова класификација у потпуности је усаглашена је са онтологијом *EarthResource*. Таксономија класе *Rudnik* дата је на слици 5.17.

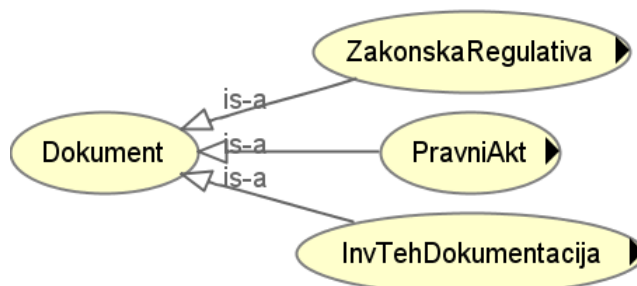


Слика 5.16. Таксономија класе *Leziste*



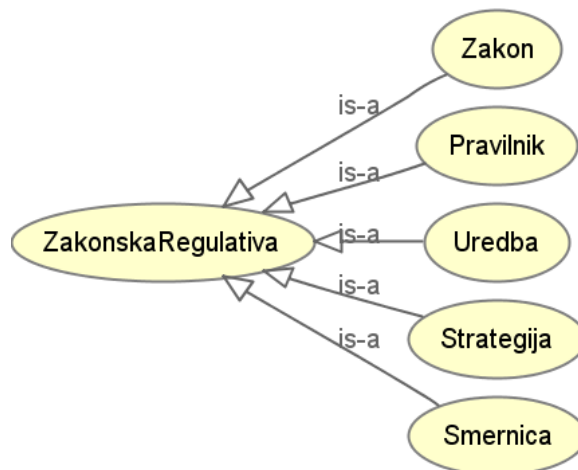
Слика 5.17. Таксономија класе *Rudnik*

Класа *Dokument* описује рударску документацију, законску регулативу и правна акта. Сходно томе, ова класа садржи следеће подкласе: *ZakonskaRegulativa*, *PravniAkt* и *InvTehDok* (слика 5.18).



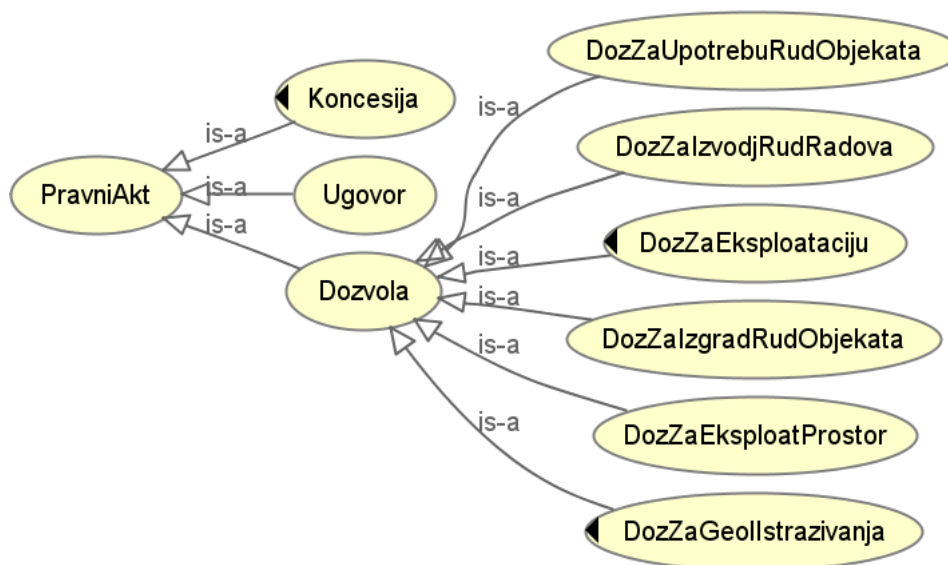
Слика 5.18. Таксономија класе *Dokument*

Класа *ZakonskaRegulativa* описује: законе, правилнике, уредбе, стратегије и смернице из домена рударства и безбедности и заштите на раду. Ова класа је усаглашена са колекцијама дигиталне библиотеке RОmeка@RGF. Таксономија класе *Dokument* дата је на слици 5.19.



Слика 5.19. Таксономија класе *ZakonskaRegulativa*

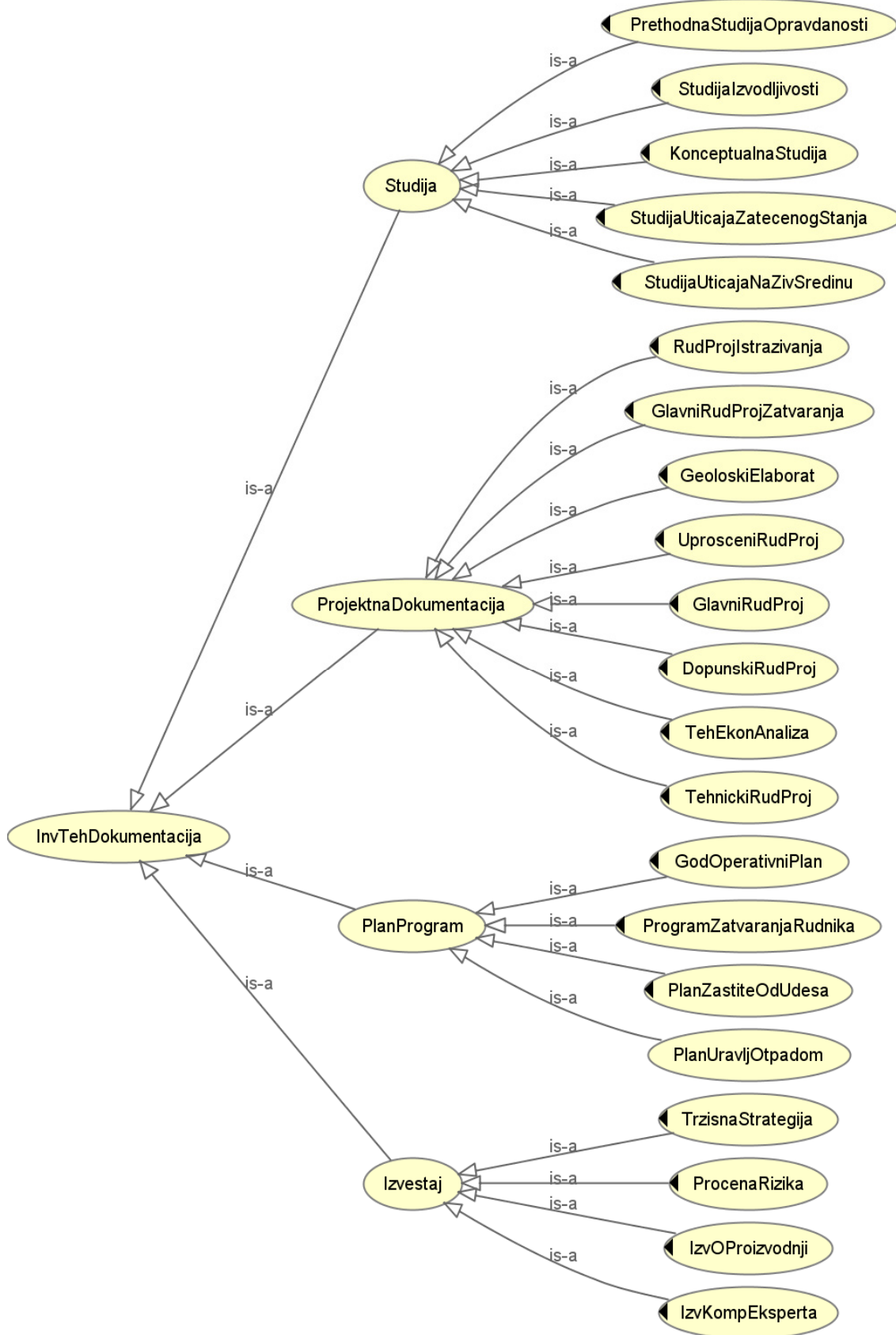
Класа *PravniAkt* описује правна акта: концесије, дозволе и уговоре, којим се регулишу односи између правних субјеката: државе са једне стране и пословних субјеката са друге стране (слика 5.20.).



Слика 5.20. Таксономија класе *PravniiAkt*

Класа *InvTehDokumentacija* описује целокупну рударску пројектну документацију, детаљно описану у поглављу 4, која прати процес развоја једног рударског пројекта. Документација је подељена на: извештаје, студије, планове и програме и пројектну документацију. Таксономија класе *InvTehDokumentacija* дата је на слици 5.21.

Класа *Projekat*, преко подкласе *Faza* гради таксономију фаза животног циклуса рударског пројекта и има следеће подкласе: *Analiza*, *Istrazivanje*, *Evaluacija*, *Projektovanje*, *Proizvodnja* и *Zatvaranje*. Свака од ових подкласа обухвата документа карактеристична за ту фазу развоја пројекта. Ове подкласе су заједничке (дељене) са подкласама класе *InvTehDokumentacija* (слика 5.22).



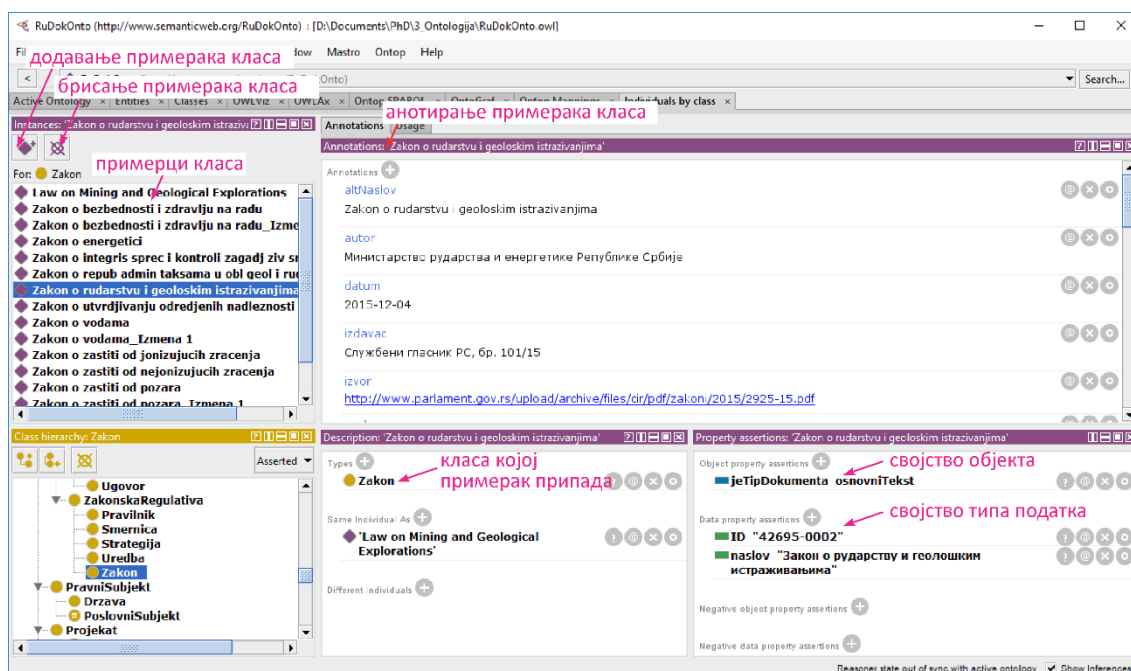
Слика 5.21. Таксономија класе *InvTehDokumentacija*



Слика 5.22. Таксономија класе *Projekat*

5.5.2.2. Примерци класа

Класама се додељују примерци (инстанце), као основне јединице класе. Поступак њиховог креирања, у сличан описаном поступку креирања класа. У делу *Individuals* врши се додавање примерака класа, као и њихово брисање. У делу *Annotations* примерци класа се детаљно могу описати метаподацима. Примера ради, сви примерци класа *ZakonskaRegulativa* и *ProjektnaDok*, с обзиром на то да су директно преузети из дигиталне библиотеке ROneka@RGF, у потпуности су аотирали метаподацима Даблинског језгра. У делу *Description* дефинише се тип примерка класе, односно дефинише се класа којој примерак припада. Такође, дефинишу се и слични примерци, па је тако примерак класе "Закон о рударству и геолошким истраживањима" исти као и примерак "Law on Mining and Geological Explorations". Панел за креирање примерака класа приказан је на слици 5.23.

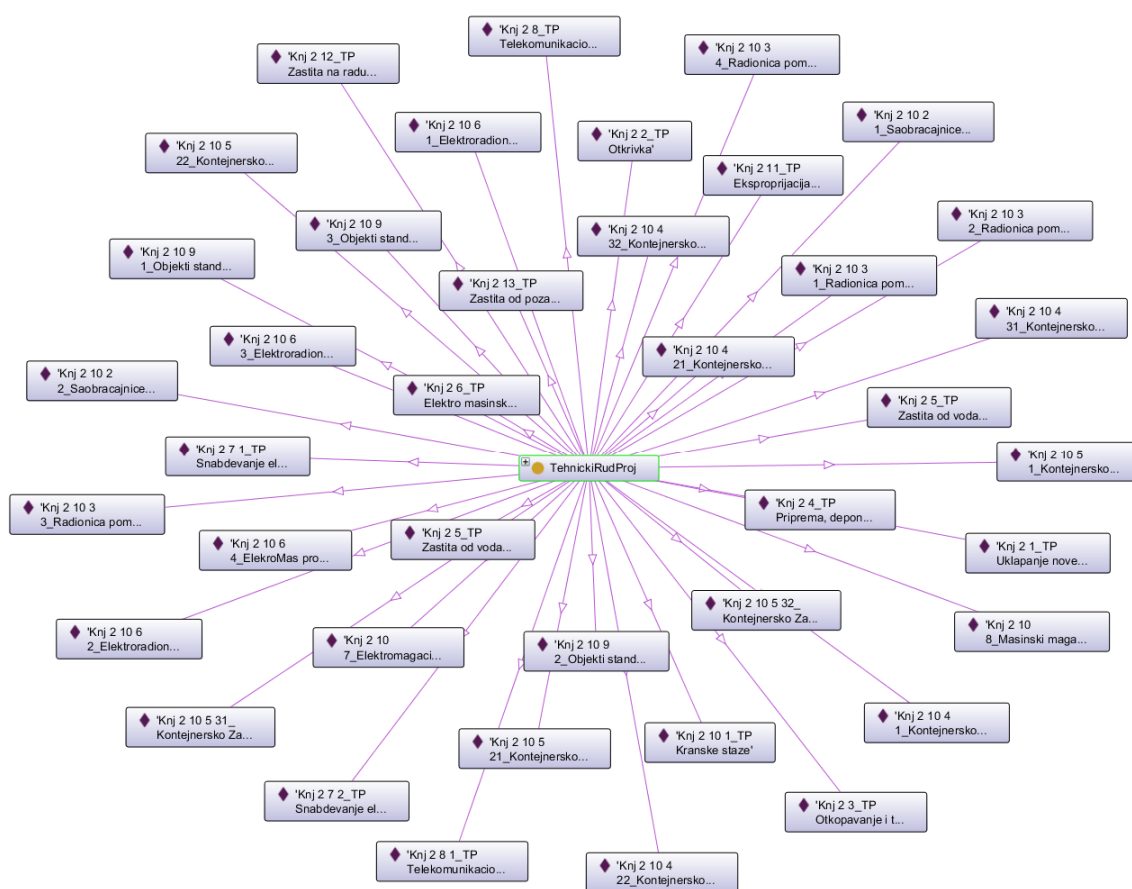


Слика 5.23. Креирање примерака класа

Овде ће бити наведени неки од примера:

- класа *PoslovniSubjekt* има примерке: ЕПС, ЈП ПЕУ Ресавица, РТБ Бор, Геолошки завод Србије, ... ;

- класа *FossilnoGorivo* има примерке: антрацит, камени угаљ, лигнит, мрки угаљ, нафта, природни гас и уљни шкриљац;
- класа *Zakon* има примерке: Закон о рударству и геолошким истраживањима, Закон о енергетици, Закон о заштити животне средине, сет закона који се односе на безбедност и заштиту на раду,...
- класа *TehnickiRudProj* има 40 примерака техничких пројеката који чине "Главни рударски пројекат ПК "Дрмно" за капацитет 9×10^6 тона угља годишње", слика 5.24.

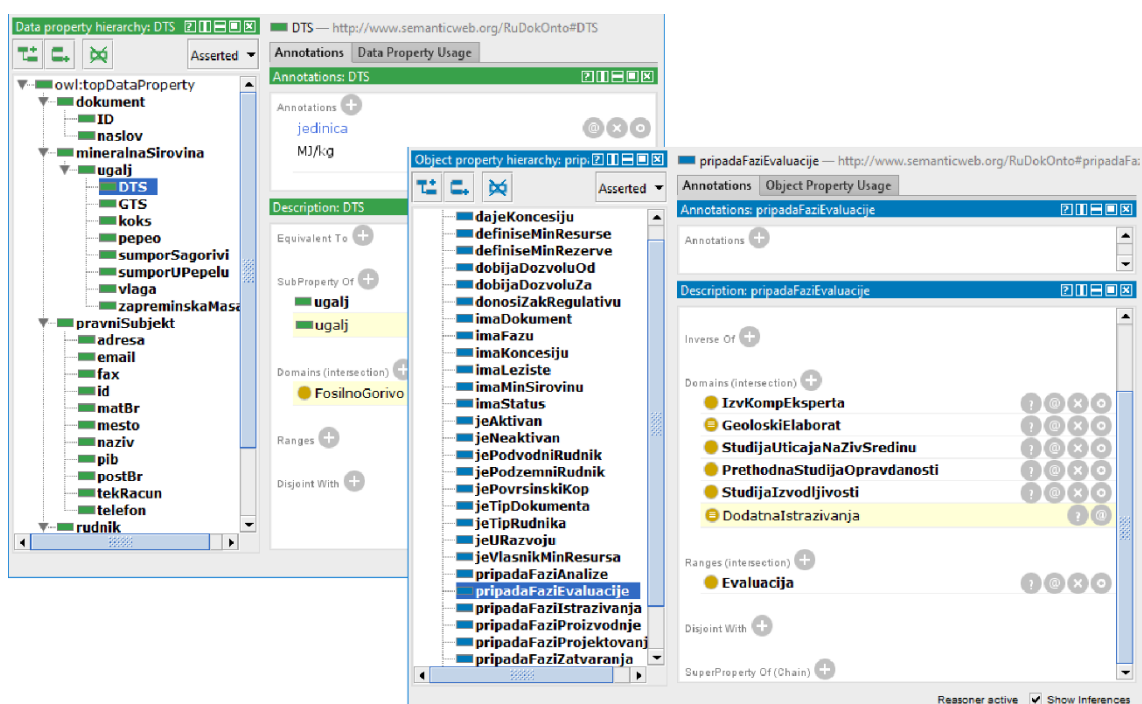


Слика 5.24. Примерци класе *TehnickiRudProj*

5.5.2.3. Релације

Релација између класа међусобно, класа и примерака, као и примерака међусобно, успоставља се преко својства објекта (енг. *Object properties*) и својстава типова података (енг. *Data properties*), као што је поменуто у поглављу 3.5.1.

Креирање и аотирање својстава врши се на исти начин као код класа и примерака класа. Својства се могу хијерархијски структурирати уколико имају неку заједничку особина, на пример заједнички домен или заједнички опсег. Својства типа података, повезују, као што је већ речено, примерке класа са неким податком. Примера ради, класа *ugalj* има следећи тип података: запреминска маса (kg/m³), влага (%), pepeo (%), сумпор у пепелу (%), сумпор сагориви (%), кокс (%), GTS (MJ/kg), DTS (MJ/kg), итд. Панели за креирање релација приказани су на слици 5.25.

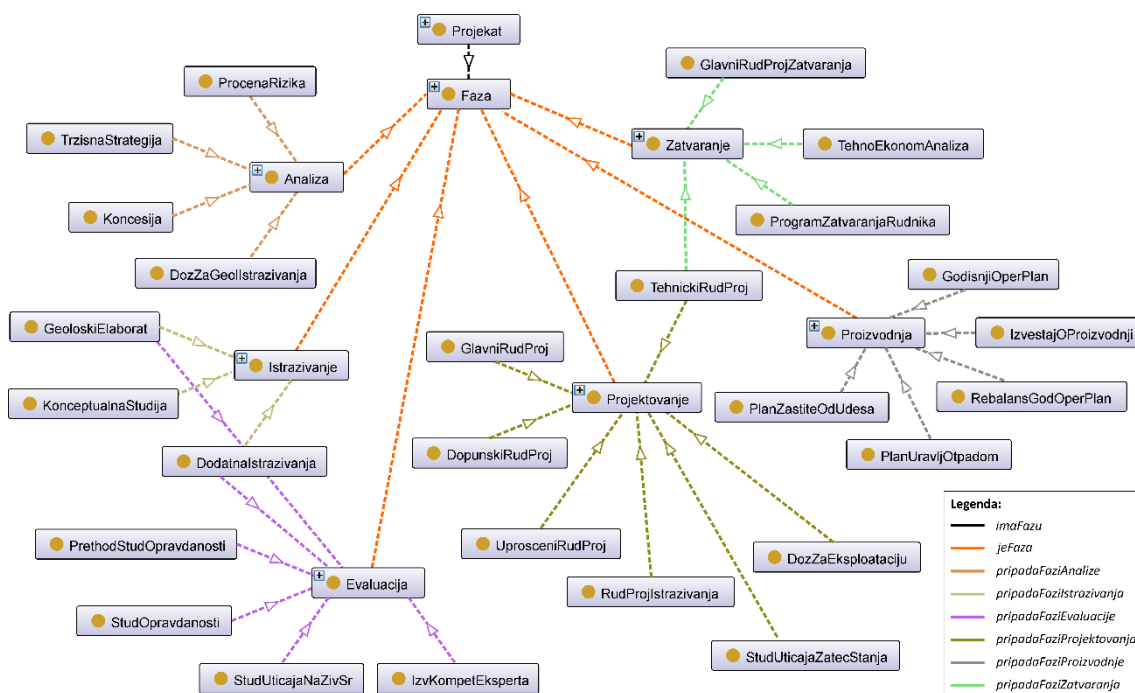


Слика 5.25. Панели за креирање својства објекта и својстава типова података

Све класе и примерци класа повезани су са 62 релације. Неке од њих приказане су на слици 5.25. За систем управљања пројектном документацијом, приказан у овој дисертацији, од највећег значаја су својства објеката који повезују фазе пројекта и документа. Табеларни приказ ових релација приказан је у табели 5.6, а њихова графичка визуализација је приказана на слици 5.26.

Табела 5.6. Повезивање докумената са фазама пројекта

Релација	Домен - документ	Опсег - фаза пројекта
<i>pripadaFaziAnalize</i>	ProcenaRizika	Анализа
	TrzisnaStrategija	
	Koncesija	
	DozZaGeolIstrazivanja	
<i>pripadaFaziIstrazivanja</i>	GeoloskiElaborat	Истраживање
	KonceptualnaStudija	
	DodatnaIstrazivanja	
<i>pripadaFaziEvaluacije</i>	PrethodStudOpravdanosti	Евалуација
	DodatnaIstrazivanja	
	StudOpravdanosti	
	StudUticajaNaZivSr	
	IzvKompetEksperta	
<i>pripadaFaziProjektovanja</i>	GlavniRudProj	Пројектовање
	DopunskiRudProj	
	TehnickiRudProj	
	UprosceniRudProj	
	RudProjIstrazivanja	
	StudUticajaZatecStanja	
	DozZaEksploataciju	
<i>pripadaFaziProizvodnje</i>	GodisnjiOperPlan	Производња
	IzvestajOProizvodnji	
	RebalansGodPlana	
	PlanUpravljOtpadom	
	PlanZastiteOdUdesa	
<i>pripadaFaziZatvaranja</i>	GlavniRudProjZatvaranja	Затварање
	TehnickiRudProj	
	TehnoEkonomAnaliza	
	ProgramZatvaranjaRudnika	



Слика 5.26. Релације између докумената и фаза пројекта

5.5.2.4. Провера онтологије

Завршна фаза развоја онтологије подразумева тестирање хијерархије класа, проверу доследности и логичке конзистентности онтологије, као и креирање упита над њима. У те сврхе коришћени су програмски додаци интегрисани у Protégé (Horridge *et al.*, 2004):

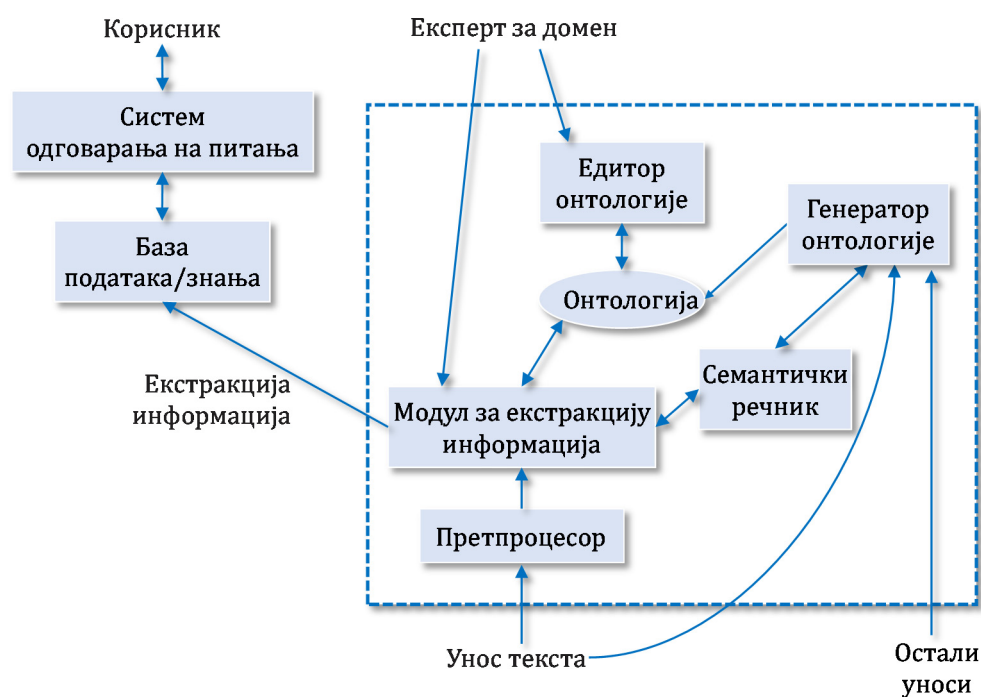
- *Debugger* – додаток за проверу кохерентности и конзистентности онтологије којим се детектују и отклањају логичке грешке;
- *Hermit*³⁶ - додаток за аутоматско закључивање и извођење нових знања у онтологијама за чије креирање је коришћен OWL (Web Ontology Language). Овим алатом се, поред додатне провере конзистентности, проверавају успостављене релације између класа, класа и примерака класа и примерака класа међусобно;
- *Ontop*³⁷ – алат за приступ подацима заснованим на онтологијама (*Ontology Based Data Access* - OBDA) и креирање SPARQL упита над њима.

³⁶ <http://www.hermit-reasoner.com/>

³⁷ <http://ontop.inf.unibz.it/>

5.6. ВЕЗА ОНТОЛОГИЈА-ДОКУМЕНТ

Систем за екстракцију информација заснован на онтологијама, обрађује неструктуриране и полуструктуриране текстове (природног језика) методама и техникама вођеним онтологијама, како би екстраховао одређене типове информација и представио излазни резултат такође уз помоћ онтологија. Слика 5.27 приказује типичну архитектуру као унију различитих компоненти које се јављају у различитим типовима оваквих система (Wimalasuriya & Dou, 2010), при чему у њима не морају да се нађу све наведене компоненте, већ само неке од њих. На пример, системи који користе онтологије које су други креирали, неће имати компоненту за интерно креирање онтологија. Едитовање онтологије и њена лексикализација јесте део система представљеног у овој тези. Опциона компонента је и систем за постављање питања на природном језику, који је обично део комплекснијег система за екстракцију информација. У систему који је развијен у оквиру истраживања ове дисертације, компонента за постављање упита на природном језику није имплементирана, али је њен развој концептуално постављен и развој је у плану.



Слика 5.27. Екстракција информација заснована на онтологијама
(Wimalasuriya & Dou, 2010)

Текст, као улаз у систем за онтолошки засновану екстракцију информација, пролази компоненту претпроцесирања, која конвертује текст у формат којим се може управљати у модулу за екстракцију информација. То значи, ако је на пример улазни текст HTML потребно је уклонити HTML етикете и метаподатке, уколико је PDF потребно је проверити исправност кодирања, интерпретације стубаца и слично. Развијено решење имплементира полуаутоматски режим, где се више различитих врста претпроцесирања могу аутоматски урадити, али финална провера и корекција је увек на крајњем кориснику.

У модулу за екстракцију информација се извршава конкретно препознавање и издвајање сегмената текста од интереса, што се реализује коришћењем система заснованог на правилима (Krstev *et al.*, 2014), али може бити засновано и на статистичким методама. Семантички речници за конкретан језик се користе за подршку оваквим системима, што је у презентованом решењу WordNet (Krstev *et al.*, 2003).

Генерисање и проширење онтологија је полуаутоматско, у смислу да се различитим методама издвајају кандидати инстанци и релација за онтологију, али је обавезно ручна верификација аутоматски генерисаних записа. Ова компонента укључује и сталну допуну онтологије *RuDokOnto*, као и онтологија *RudOnto* (Kolonja, 2016), *Termi*³⁸ и *GeolISSTerm* (Stanković *et al.*, 2011a).

Излаз система за екстракцију информација заснованог на онтологији састоји се од података издвојених из текста. Ови подаци се могу приказати помоћу језика за креирање онтологије, у овом случају то је OWL. Поред тога излаз могу бити и везе са текстуалним документима из којих су подаци издвојени, при чему-излаз може садржати и везе са текстуалним документима из којих су подаци извучени, Ово је нарочито важно јер обезбеђује поузданије одређивање релевантности одговора који је понуђен кориснику на основу екстраховане информације.

Препознавање именованих ентитета (енг. *Named Entity Recognition*, NER) према (Sekine & Ranchhod, 2009) се обично односи на имена особа, локалитета или организација, као и нумеричких израза који укључују датум, време, новац

³⁸ <http://termi.rgf.bg.ac.rs/>

и проценте и представља кључни корак за екстракцију информација. Однедавно се и други типови именованих ентитета укључују, као на пример производи и догађаји, али такође и е-мејл адресе и наслови књига.

Хијерархија именованих ентитета у систему препознавања именованих ентитета (NER) која је коришћена у овом раду се састоји од пет типова највишег нивоа: особе, организације, локације, количине и временски изрази (Krstev *et al.*, 2014). Сваки од њих има један или више нивоа подтипова. Стратегија аотирања дозвољава угњеждавање, што значи да се именовани ентитет могу угњездити у другу именовани ентитет, на пример топоним у називу организације "Институт за грађевинарство Суботица": <org>Институт за грађевинарство <top>Subotica </top></org>. Етикета <top> означава да је у питању топоним, а <org> да је у питању организација.

Према (Stanković *et al.*, 2017) у тексту који се односио на геолошку документацију, око 1,14% текста чине особе (лична имена особа), локације чак 5,05% и организације 0,27%, што укупно чини 6,45%.

Систем за препознавање именованих ентитета за српски језик је заснован на ручно креираним правилима која се ослањају на свеобухватне лексичке ресурсе за српски језик. За препознавање неких типова именованих ентитета, као што су лична имена и локације, у морфолошким електронским речницима су садржане кључне информације, док код рецимо временских израза, локалне граматике у облику коначних трансдуктора су примарне. Трансдукторима се покушава препознати и обележити разноликост синтаксичких облика у којима се именовани ентитет може појавити. Ипак, за све њих је развијен систем локалних граматика који користи шири контекст текста како би решио вишезначност појединих сегмената текста, колико год је могуће (Krstev *et al.*, 2014). Ове локалне граматике се организују као каскаде којима се разрешавају вишезначности (Maurel *et al.*, 2011), тако да на скупу новинских текстова на ком је рађена евалуација аутори наводе (Krstev *et al.*, 2014) да Ф-мера препознавања је 0:96 на нивоу типова речи (лема) и 0:92 за токене.

Физичка аотација текста видљива машини, у облику XML-а, приказана је наслици 5.28.

Ležište uglja Drmno se nalazi u istočnom delu Kostolačkog ugljenog basena i zahvata površinu od `<measure.approx>oko <measure.exact>52 km2</measure.exact></measure.approx>`. Prostire se istočno od reke `<top.hyd>Mlave</top.hyd>`, a granicu mu čine: Reka `<top.hyd>Dunav</top.hyd>` na severu, Boževačka greda na istoku, linija `<top.gr>Bradarac </top.gr>` – Sirakovačka dolina na jugu i reka `<top.hyd>Mlava</top.hyd>` na zapadu. Ležište uglja Drmno zahvata atar sela `<top.gr>Drmno </top.gr>` po kome je dobilo ime, ali takođe i deo atara susednih sela `<top.gr>Bradarca</top.gr>` i `<top.gr>Kličevca</top.gr>`. Površinski kop "Drmno" se nalazi severoistočno od sela Drmno, slika 1.1.1. Ležište Drmno sa Požarevcem i Kostolcom povezuju asfaltirani putevi – od `<measure.approx>oko <measure.exact>18 km</measure.exact> </measure.approx>` do `<top.gr>Požarevca </top.gr>` i `<measure.approx>oko <measure.exact>4 km</measure.exact> </measure.approx>` do `<top.gr>Kostolca</top.gr>`.

Između `<amount.exact>dva grebena</amount.exact>` – Požarevačke i Boževačke grede, prostire se ravnica `<top.gr>Stig</top.gr>`, koja preko kose Klepečke, ka severu, prelazi u aluvion `<top.hyd>Dunava</top.hyd>`. U severnom delu `<top.gr>Stiga</top.gr>` se prostire samo ležište Drmno. Važni objekti na terenu iznad ležišta uglja su selo Drmno i `<org>Termoelektrana <top.gr>Kostolac</top.gr> B" </org>`. Na samom kopu u otkopanom delu ležišta je formirano unutrašnje odlagalište otkrivke. U graničnim arealima ležišta se nalaze:

- na zapadu: `<top.hyd>Mlava</top.hyd>` i Požarevačka greda sa manastirom Rukumija;
- na severu: `<top.hyd> Mali Dunav</top.hyd>` ("Dunavac"), močvarno-šumski predeo Hrastovače i `<top.hyd>Dunav</top.hyd>`;
- na istoku: selo `<top.gr>Kličevac</top.gr>` i
- na jugu: spoljno odlagalište otkrivke i selo `<top.gr>Bradarac</top.gr>`.

[...]

Kako je u predhodnom tekstu napomenuto, do `<time.date.abs>kraja 2018. godine</time.date.abs>`, kapacitet na uglju će ostati na nivou od `<measure.approx>oko <measure.exact>9 miliona tona</measure.exact></measure.approx><time.set>godišnje</time.set>`. Puštanje u rad novog bloka B3 je planirano za `<time.date.abs>2019. godini</time.date.abs>`, tako da će se kapacitet kopa Drmno povećati na `<measure.exact>11,3×106 t</measure.exact>` uglja u `<time.date.rel>ovoj godini</time.date.rel>`, što je ujedno i najveći projektovani kapacitet. U periodu nakon `<time.date.abs>2024. godine</time.date.abs>` eksploatacije kapacitet kopa na uglju će se stalno smanjivati.

Слика 5.28. XML приказ текста са анотираним именованим ентитетима

На слици 5.29 је приказан поглед на текст видљив кориснику, где су именовани ентитети обојени према типу (организација, локација, количина и временски израз).

Ležište uglja **Drmno** se nalazi u istočnom delu Kostolačkog ugljenog basena i zahvata površinu od **oko 52 km²**. Prostire se istočno od reke **Mlave**, a granicu mu čine: Reka **Dunav** na severu, Boževačka greda na istoku, linija **Bradarac** – Sirakovačka dolina na jugu i reka **Mlava** na zapadu. Ležište uglja Drmno zahvata atar sela **Drmno** po kome je dobilo ime, ali takođe i deo atara susednih sela Bradarca i Kličevca. Površinski kop "Drmno" se nalazi severoistočno od sela Drmno, slika 1.1.1. Ležište Drmno sa Požarevcem i Kostolcom povezuju asfaltirani putevi – od **oko 18 km** do Požarevca i **oko 4 km** do Kostolca.

Između **dva grebena** – Požarevačke i Boževačke grede, prostire se ravnica **Stig**, koja preko kose Klepečke, ka severu, prelazi u aluvion **Dunava**. U severnom delu **Stiga** se prostire samo ležište Drmno. Važni objekti na terenu iznad ležišta uglja su selo **Drmno** i **Termoelektrana "Kostolac B"**. Na samom kopu u otkopanom delu ležišta je formirano unutrašnje odlagalište otkrivke. U graničnim arealima ležišta se nalaze:

- na zapadu: **Mlava** i Požarevačka greda sa manastirom Rukumija;
- na severu: **Mali Dunav** ("Dunavac"), močvarno-šumski predeo Hrastovače i **Dunav**;
- na istoku: selo **Kličevac** i
- na jugu: spoljno odlagalište otkrivke i selo **Bradarac**.

[...]

Kako je u predhodnom tekstu napomenuto, do **kraja 2018. godine**, kapacitet na uglju će ostati na nivou od **oko 9 miliona tona** godišnje. Puštanje u rad novog bloka B3 je planirano za **2019. godini**, tako da će se kapacitet kopa Drmno povećati na **11,3×10⁶ t** uglja u **ovoj godini**, što je ujedno i najveći projektovani kapacitet. U periodu **nakon 2024. godine** eksploatacije kapacitet kopa na uglju će se stalno smanjivati.

Слика 5.29. Приказ текста са анотираним именованим ентитетима

Из аотираног текста је могуће једноставним трансформацијама извући само обележене сегменте и приказати табеларно, графички или у виду кратких сажетака. За питање "О коме се ради?" се везују делови текста обележени етикетом *<pers>* (прецизније *<persName.full>* за пуно име и презиме, *<persName.first>* за име и *<persName.last>* за презиме); за питање "О чему се ради?" се везују делови текста обележени етикетом *<organization>*; за питање "На ком месту/где?" се везују делови текста обележени етикетом *<top>*; за питање "Ког дана?" се везују делови текста обележени етикетом *<date>* или *<time.date.abs>* за апсолутно одређен датум, итд. У обрађеном документу би се тако издвојили:

О чему се ради?	Термоелектрана "Костолац Б"
На ком месту/где?	реке: Дунав, Млава,...
На ком месту/где?	Пожаревац, Дрмно, Костолац,...
Ког дана?	2018. године, до краја године,...

6. ЕВАЛУАЦИЈА МОДЕЛА РАЗВИЈЕНОГ СИСТЕМА

6.1. ПРЕТРАЖИВАЊЕ ДОКУМЕНАТА У ДИГИТАЛНОЈ БИБЛИОТЕЦИ

У дигиталној библиотеци RОmeка@RGF претраживање појмова и фраза могуће је на два начина:

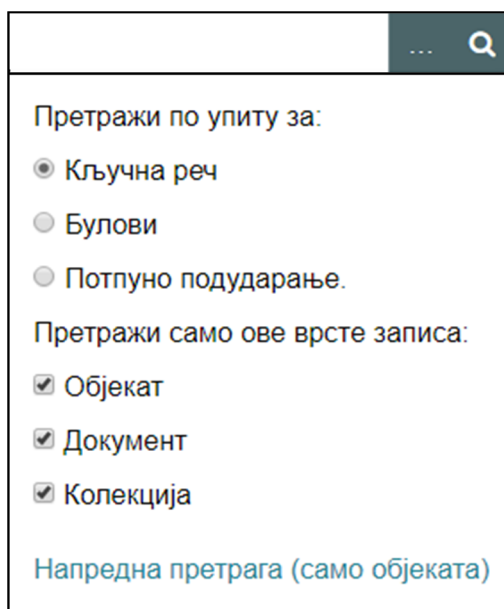
- напредно претраживање (енг. *Advanced Search*) по кључним речима и
- претраживање целокупног садржаја (енг. *Site-wide Search*) дигиталне библиотеке.

Напредно претраживање је фокусирано искључиво на тражење кључних речи унутар: *Dublin Core* метаподатака, колекција, етикета, корисника, јавне доступности објеката и геолокације. Главни недостатак овог начина претраживања је немогућност претраживања самих докумената.

За управљање рударском пројектном документацијом неупоредиво је значајније претраживање целокупног садржаја јер исправља главни недостатак претходног начина и омогућава претраживање свих побројаних елемената дигиталне библиотеке укључујући и документе, што је у овом случају било посебно важно.

Овај начин претраживања укључује три опције претраживања: кључна реч, буловски оператори (и/или тј. *and/or*) и потпуно подударање (слика 6.1). Претраживање кључних речи враћа резултате сортиране по релевантности. На пример, ако се претражује сложени појам "управљање квалитетом угља" резултат ће најпре бити листа докумената која садржи тражену кључни реч, а потом ће се приказати и записи који садрже појмове "управљање", "квалитетом" и "угља". Буловски оператори омогућавају прецизније упите али не враћају резултате према релевантности. Упити се уносе комбиновањем речи и логичких оператора: +, -, * и ". Знак + означава да реч, испред које је уметнут, мора бити присутна у сваком враћеном резултату, насупрот знака - који значи да се реч не сме наћи у резултату претраживања. Звездича, односно оператор скраћивања, као резултат даје речи које се подударају са речју која претходи

звездици. Изрази уметнути између знакова навода као резултат претраживања дају израз који дословно одговара упиту. Потпуно подударање враћа записе који садрже барем једно подударање с редоследом унетих речи.



Слика 6.1. Претраживање докумената коришћењем стандардног упита

На први поглед, рекло би се, да је овакав начин претраживања задовољава потребе корисника дигиталне библиотеке. Међутим, при тестирању овог начина претраживања констатован је проблем везан за српски језик. Претраживање докумената на српском језику је сложено због веома богатог морфолошког система. При постављању упита, променљиве речи се по правилу наводе само у свом канонском облику (номинатив једнине за именице, инфинитив за глагол) што резултује високим одзивом и ниском прецизношћу. При томе се заборавља на чињеницу да релевантна документа могу садржавати било који од флективних облика променљиве речи. Проблем постаје још сложенији ако се у обзир узму и деривациона својства (на пример, промена рода или релациони и присвојни придев) или сложене речи (Stanković, 2009).

Овај проблем је у дигиталној библиотеци ROmeka@RGF решен имплементирањем експанзије упита. За експанзију упита коришћени су веб сервиси (Krstev *et al.*, 2008a, Stanković *et al.*, 2012), и морфолошки е-речници за српски језик

(Krstev, 2008, Krstev *et al.*, 2008a, Stanković *et al.*, 2016) развијени у оквиру Групе за језичке технологије на Математичком факултету Универзитета у Београду (Vitas *et al.*, 2003). Поред корпуса српског језика, као и вишејезичних паралелних корпуса, од посебног су значаја систем морфолошких речника српског језика, као и семантичка мрежа за српски језик – *Serbian WordNet* (SWN, српски ворднет) (Krstev *et al.*, 2004), развијени у оквиру међународног пројекта Балканет (Tufiş, 2004).

Проблеми у претраживању текстуалних ресурса се могу класификовати у две категорије:

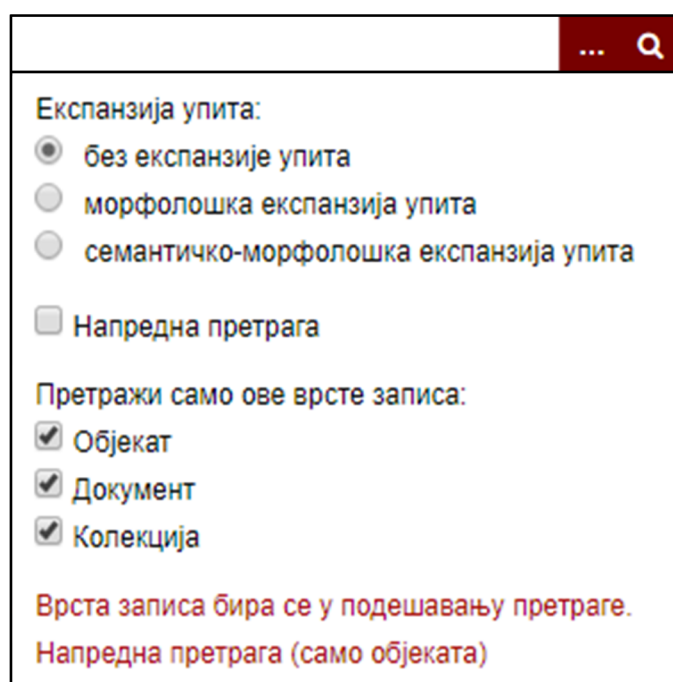
- општи, који не зависе од језика и
- проблеми који су специфични за поједини језик или групу језика.

Проблем при претраживању текстова на српском језику представљају различите кодне шеме као и постојање два равноправна алфабета (ћириличног и латиничног). То је случај са дигиталном библиотеком RОmeка@RGF, с обзиром да су документа увезена о оригиналној форми, на оба писма. Претраживање садржаја само по једном писму је могуће у случају претраживања без проширења упита, док проширење упита аутоматски подразумева оба писма. Претраживање докумената на српском језику је сложен процес због веома богатог морфолошког система. Упитима се најчешће траже речи у свом канонском облику (номинатив једнине за именице, инфинитив за глагол) (Lazić *et al.*, 2016). Међутим, документа могу садржати било који флективни облик променљиве речи. Овај проблем постаје сложенији ако се у обзир узму и сложене речи и синоними (Stanković, 2009).

Елементи дигиталне библиотеке које је могуће укључити у процес претраживања су: метаподаци, документи, етикете, извештаји, изложбе, веб странице. Изворно, претраживања се врше преко кључних речи, булових оператора и потпуним подударањем. Дигитална библиотека RОmeка@RGF унапређена је имплементирањем проширених упита. Коришћени су веб сервиси (Stanković *et al.*, 2012) и морфолошки електронски речници за српски језик (Krstev *et al.*, 2008a, Stanković *et al.*, 2016):

- за морфолошко проширење упита:
http://hlt.rgf.bg.ac.rs/vebran/api/delafs/ključna_reč
- за семантичко и морфолошко проширење упита:
http://hlt.rgf.bg.ac.rs/vebran/api/sinonimi/ključna_reč

Постојеће опције претраживања замењене су новим: без проширења упита, са морфолошким проширењем упита и са семантичко-морфолошким проширењем упита (слика 6.2). Претраживање без експанзије упита одговара већ описаном претраживању упита буловим операторима.



Слика 6.2. Претраживања докумената коришћење експанзије упита

Друга опција је морфолошка експанзија упита која користи веб сервис: http://hlt.rgf.bg.ac.rs/Vebran/api/Delafs/key_word³⁹ (Krstev, 2008, Stanković, 2009). Овим сервисом, упит се проширује на све флективне облике тражене кључне речи и то на оба писма, ћирилице и латинице. На пример, морфолошка експанзија кључне речи "хомогенизација", проширује упит на 14 облика тражене кључне речи:

³⁹ Human Language Technologies, <http://hlt.rgf.bg.ac.rs/>

homogenizacija, homogenizacijama, homogenizacije, homogenizaciji, homogenizacijo, homogenizacijom, homogenizaciju, хомогенизација, хомогенизацијама, хомогенизације, хомогенизацији, хомогенизацијо, хомогенизацијом, хомогенизацију.

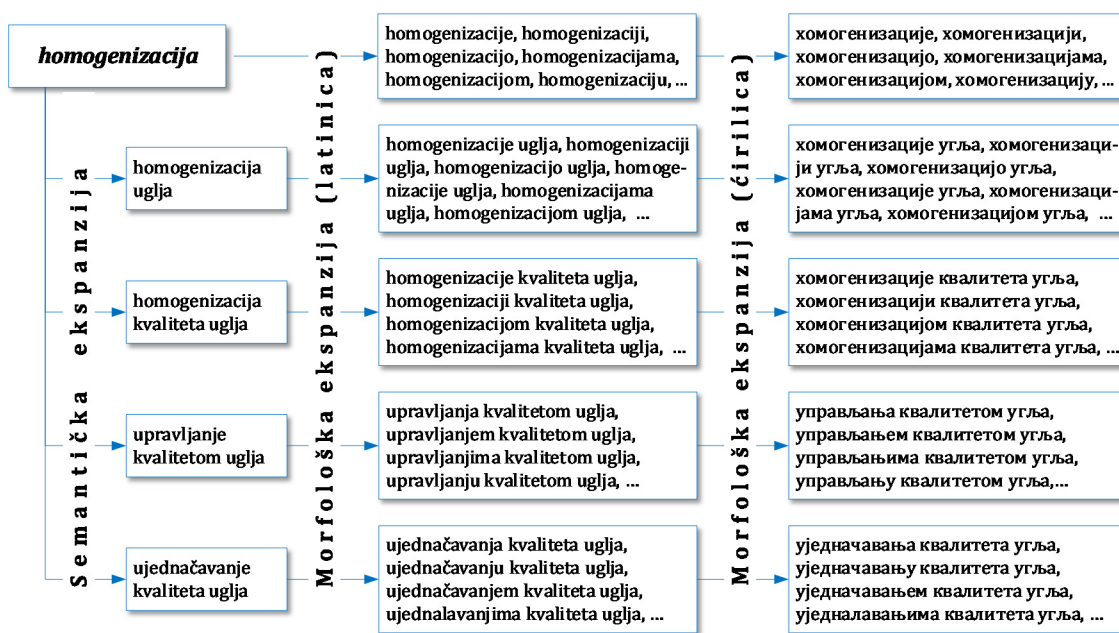
Трећа опција је семантичко-морфолошка експанзија. Она обједињује претходно описану морфолошку експанзију са семантичком експанзијом упита, која користи веб сервис http://hlt.rgf.bg.ac.rs/Vebran/api/sinonimi/key_word⁴⁰ (Krstev, 2008, Stanković, 2009). У овом случају као резултат враћају се сви флективни облици кључне речи и њени синоними и лексичке метонимије, као и њихови флективни облици, на оба писма, ћириличном и латиничном. На пример, морфолошко-семантичком експанзијом упита кључне речи "хомогенизација", поред задате кључне речи у претраживање укључују и синоними: "хомогенизација угља", "хомогенизација квалитета угља", "управљање квалитетом угља", "уједначавање квалитета угља". На овај начин, у претраживање се укључује чак 64 облика задате кључне речи, што значајно повећава одзив на упит:

homogenizacija, homogenizacijama, homogenizacije, homogenizaciji, homogenizacijo, homogenizacijom, homogenizaciju, ujednačavanje kvaliteta uglja, homogenizacija kvaliteta uglja, homogenizacijama kvaliteta uglja, homogenizacije kvaliteta uglja, homogenizaciji kvaliteta uglja, homogenizacijo kvaliteta uglja, homogenizacijom kvaliteta uglja, homogenizaciju kvaliteta uglja, homogenizacija uglja, homogenizacijama uglja, homogenizacije uglja, homogenizaciji uglja, homogenizacijo uglja, homogenizacijom uglja, homogenizaciju uglja, upravljanja kvalitetom uglja, upravljanje kvalitetom uglja, upravljanjem kvalitetom uglja, upravljanjima kvalitetom uglja, upravljanju kvalitetom uglja, upravljanja kvalitetom, upravljanje kvalitetom, upravljanjem kvalitetom, upravljanjima kvalitetom, upravljanju kvalitetom, хомогенизација, хомогенизацијама, хомогенизације, хомогенизацији, хомогенизацијо, хомогенизацијом, хомогенизацију, уједначавање квалитета угља, хомогенизација квалитета угља, хомогенизацијама квалитета угља, хомогенизације квалитета угља, хомогенизацији квалитета угља, хомогенизацијо квалитета угља, хомогенизацијом квалитета угља, хомогенизацију квалитета угља, хомогенизација угља, хомогенизацијама угља, хомогенизације угља, хомогенизацији угља, хомогенизацијо угља, хомогенизацијом угља, хомогенизацију угља, хомогенизацијама угља, хомогенизацијом угља, хомогенизацију угља,

⁴⁰ Human Language Technologies, <http://hlt.rgf.bg.ac.rs/>

угља, управљања квалитетом угља, управљање квалитетом угља, управљањем квалитетом угља, управљањима квалитетом угља, управљању квалитетом угља, управљања квалитетом, управљање квалитетом, управљањем квалитетом, управљањима квалитетом, управљању квалитетом.

Графички приказ проширивања упита за задату кључну реч "homogenizacija" дат је на слици 6.3.



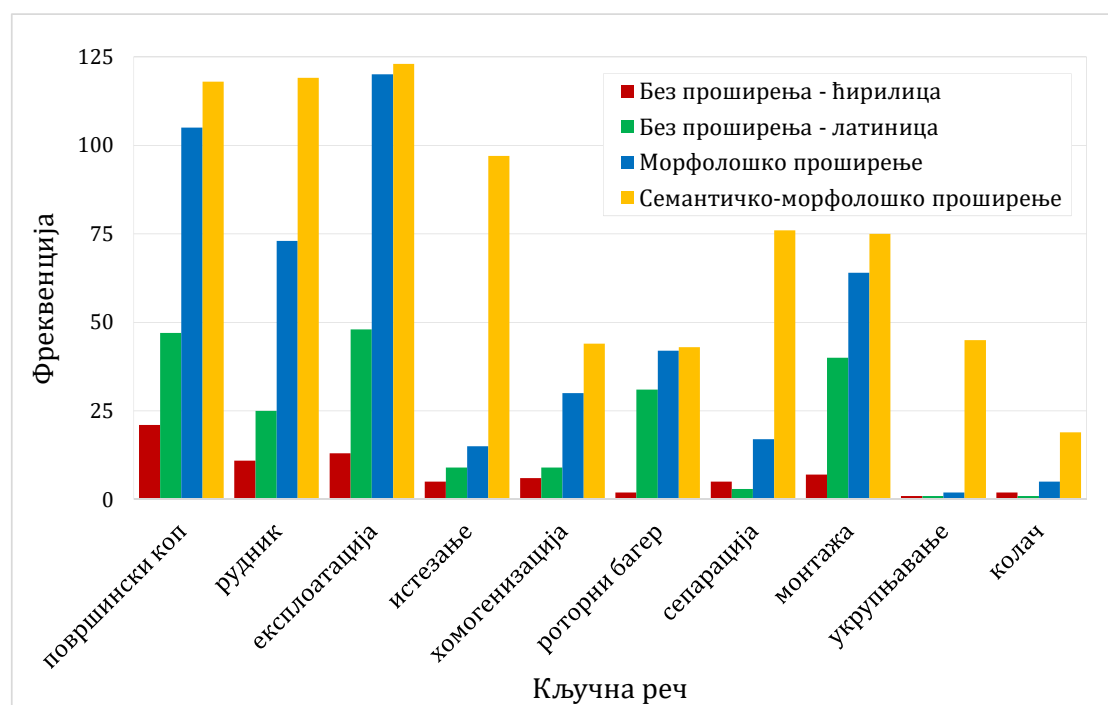
Слика 6.3. Пример проширења упита за кључну реч "хомогенизација"

Карактеристични резултати претраживања једночланих и вишечланих кључних речи у дигиталној библиотеци ROmeka@RGF приказани су у табели 6.1. Бројеви у табели приказују понуђена докумената, док су у загради наведени они који су у фази евалуације оцењени као нерелевантни. Графички приказ добијених резултата дат је на слици 6.4.

На основу добијених резултата претраживања урађен је прорачун прецизности, одзива и F-мере, који је приказан у табели 6.2. Графички приказ прецизности, одзива и F-мере дати су на слици 6.5.

Табела 6.1. Резултати претраживања у дигиталној библиотеци R0meKa@RGF

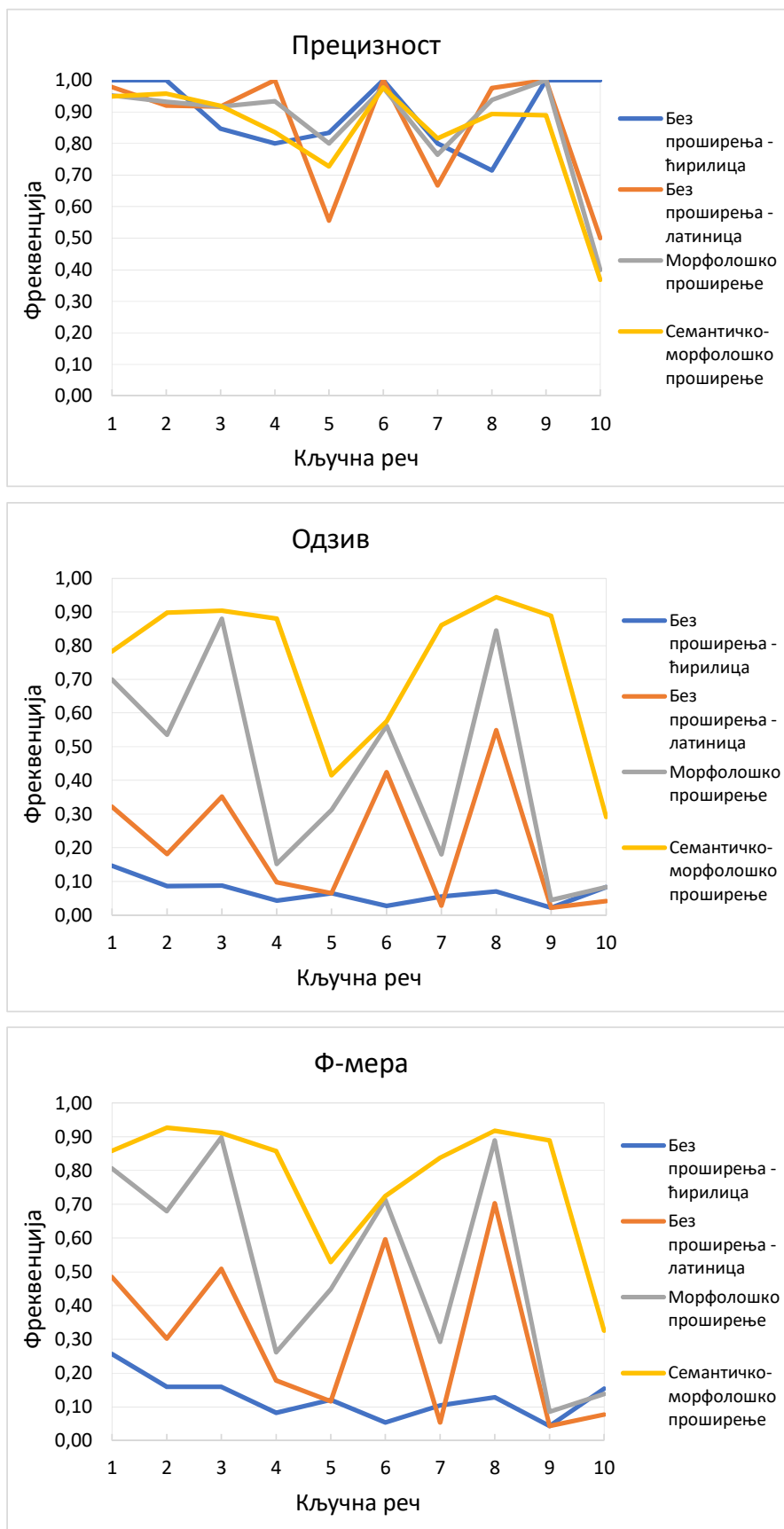
Р. бр.	Кључна реч	Без проширења		Са проширењем		Број релевантних докумената
		ћирилица	латиница	морфолошким	семантичко-морфолошким	
1.	површински коп	21 (0)	47 (1)	105 (5)	118 (6)	143
2.	рудник	11 (0)	25 (2)	73 (5)	119 (5)	127
3.	експлоатација	13 (2)	48 (4)	120 (10)	123 (10)	125
4.	истезање	5 (1)	9 (0)	15 (1)	97 (16)	92
5.	хомогенизација	6 (1)	9 (4)	30 (6)	44 (12)	77
6.	роторни багер	2 (0)	31 (0)	42 (1)	43 (1)	73
7.	сепарација	5 (1)	3 (1)	17 (4)	76 (14)	72
8.	монтажа	7 (2)	40 (1)	64 (4)	75 (8)	71
9.	укрупњавање	1 (0)	1 (0)	2 (0)	45 (5)	45
10.	колач	2 (0)	1 (1)	5 (3)	19 (12)	24



Слика 6.4. Резултати претраживања једночланих и вишечланих термина у дигиталној библиотеци R0meKa@RGF

Табела 6.2. Прорачун вредности одзива, прецизности и Ф-мере

Р. бр.	Кључна реч	Ir	In	Nr	Прецизност	Одзив	Ф-мера	Укупан број релевантних докумената
Без проширења - ћирилица								
1.	површински коп	21	0	122	1,000	0,147	0,256	143
2.	рудник	11	0	116	1,000	0,087	0,159	127
3.	експлоатација	11	2	114	0,846	0,088	0,159	125
4.	истезање	4	1	88	0,800	0,043	0,082	92
5.	хомогенизација	5	1	72	0,833	0,065	0,120	77
6.	роторни багер	2	0	71	1,000	0,027	0,053	73
7.	сепарација	4	1	68	0,800	0,056	0,104	72
8.	монтажа	5	2	66	0,714	0,070	0,128	71
9.	укрупњавање	1	0	44	1,000	0,022	0,043	45
10.	колач	2	0	22	1,000	0,083	0,154	24
Без проширења - латиница								
1.	површински коп	46	1	97	0,979	0,322	0,484	143
2.	рудник	23	2	104	0,920	0,181	0,303	127
3.	експлоатација	44	4	81	0,917	0,352	0,509	125
4.	истезање	9	0	83	1,000	0,098	0,178	92
5.	хомогенизација	5	4	72	0,556	0,065	0,116	77
6.	роторни багер	31	0	42	1,000	0,425	0,596	73
7.	сепарација	2	1	70	0,667	0,028	0,053	72
8.	монтажа	39	1	32	0,975	0,549	0,703	71
9.	укрупњавање	1	0	44	1,000	0,022	0,043	45
10.	колач	1	1	23	0,500	0,042	0,077	24
Морфолошко проширење								
1.	површински коп	100	5	43	0,952	0,699	0,806	143
2.	рудник	68	5	59	0,932	0,535	0,680	127
3.	експлоатација	110	10	15	0,917	0,880	0,898	125
4.	истезање	14	1	78	0,933	0,152	0,262	92
5.	хомогенизација	24	6	53	0,800	0,312	0,449	77
6.	роторни багер	41	1	32	0,976	0,562	0,713	73
7.	сепарација	13	4	59	0,765	0,181	0,292	72
8.	монтажа	60	4	11	0,938	0,845	0,889	71
9.	укрупњавање	2	0	43	1,000	0,044	0,085	45
10.	колач	2	3	22	0,400	0,083	0,138	24
Семантичко-морфолошко проширење								
1.	површински коп	112	6	31	0,949	0,783	0,858	143
2.	рудник	114	5	13	0,958	0,898	0,927	127
3.	експлоатација	113	10	12	0,919	0,904	0,911	125
4.	истезање	81	16	11	0,835	0,880	0,857	92
5.	хомогенизација	32	12	45	0,727	0,416	0,529	77
6.	роторни багер	42	1	31	0,977	0,575	0,724	73
7.	сепарација	62	14	10	0,816	0,861	0,838	72
8.	монтажа	67	8	4	0,893	0,944	0,918	71
9.	укрупњавање	40	5	5	0,889	0,889	0,889	45
10.	колач	7	12	17	0,368	0,292	0,326	24
<p>Легенда: Ir - издвојени релевантни документи, In - издвојени нерелевантни документи, Nr - неиздвојени релевантни документи $P = Ir/(Ir+In)$, $O = Ir/(Ir+Nr)$, $F = (2xPxO)/(P+O)$.</p>								



Слика 6.5. Графички приказ прецизности, одзива и Ф-мере

На основу добијених вредности ових величина, може се закључити да је прецизност већине упита задовољавајућа, осим кључних речи "хомогенизација" и "колач", јер су обе речи семантички двосмислене. Одзив, а самим тим и Ф-мера, су врло ниски у случају упита без проширења, док су резултати претраживања са проширеним упитом значајно већи. Одзив за упит "хомогенизација" не селектује све релевантне документе зато што синоним "мешање угља" се у документима користи као синоним а не постоји у терминолошкој бази, те тако није ни укључен у проширени упит.

Посматрајући кључну реч "површински коп" уочава се да је код претраживања без проширења на ћириличном писму, када се тражи искључиво наведена кључна реч, Ф-мера ниска (0,256) због мало пронађених релевантних докумената (21). Код упита без проширења на латиничном писму, број релевантних докумената је већи (46), па и Ф-мера има већу вредност (0,484). Када се у претраживање укључи морфолошко проширење (сви флективни облици наведене кључне речи на оба писма) број релевантних докумената нагло расте (100) па је и Ф-мера знатно већа (0,806). Када се у претраживање укључи и семантичко проширење, број релевантних докумената је највећи (112) а Ф-мера има вредност 0,858.

6.2. ПРЕТРАЖИВАЊЕ КОРПУСА ЛОКАЛНИМ ГРАМАТИКАМА

Претраживање дигиталне библиотеке ограничено је, са једне стране на претраживање засновано на кључним речима, ретко подржано регуларним изразима⁴¹, а са друге стране претраживањем заснованим на семантичком и морфолошком проширењу упита. Међутим, често оваква претраживања не могу адекватно да изразе све информационе потребе корисника.

У табели 6.3 дат је пример резултата упита *površinski kop Tamnava Zapadno polje*. Осенчена поља представљају резултате која се добијају претрагом заснованом на кључним речима укључујући експанзију упита.

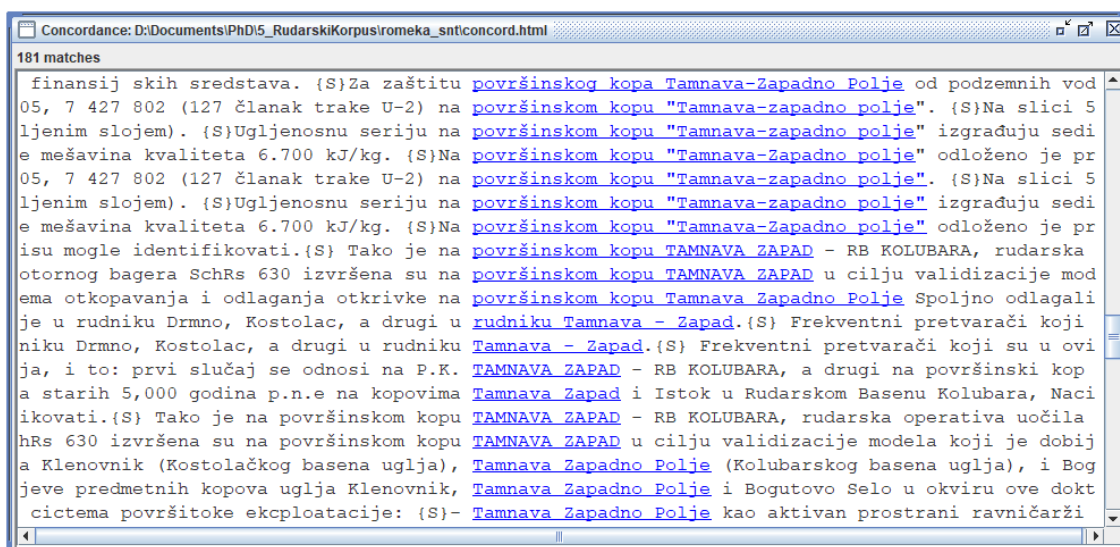
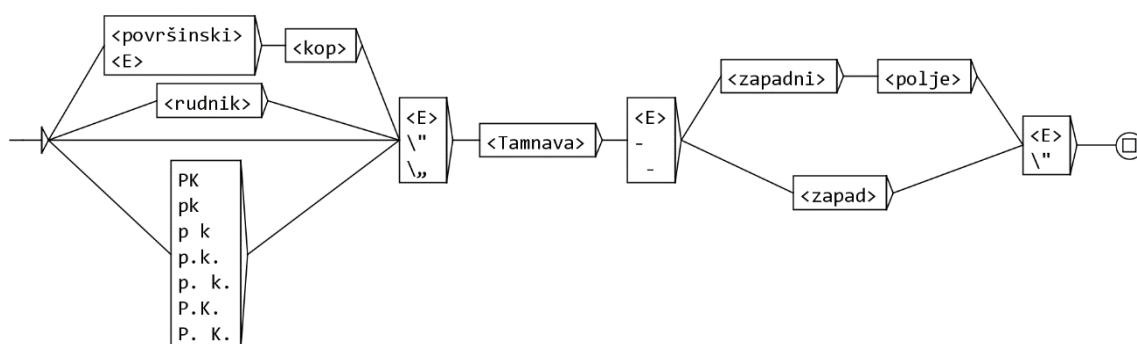
Табела 6.3. Препознати изрази са фреквенцијама добијени применом синтаксичког графа на рударски корпус

Препознати израз	Фрекв.	Препознати израз	Фрекв.
"Tamnava – Zapad"	1	Površinski kop "Tamnava-zapadno polje"	1
"Tamnava – Zapadno polje"	2	Površinski kop Tamnava Zapadno Polje	4
"Tamnava-Zapad"	1	površinski kop Tamnava Zapadno Polje	1
"Tamnava-zapadno polje"	3	Površinski kop Tamnava-Zapadno polje	1
"Tamnava-Zapadno polje"	2	površinskog kopa "Tamnava-zapadno polje"	1
"Tamnava-Zapadnom polju"	1	površinskog kopa Tamnava Zapadno Polje	5
kop "Tamnava-Zapadno polje"	1	površinskog kopa Tamnava Zapadno polje	2
kop Tamnava Zapadno Polje	1	površinskog kopa Tamnava-Zapadno Polje	2
kopa "Tamnava-Zapadno polje"	1	površinskom kopu "Tamnava-zapadno polje"	2
kopa "Tamnava-zapadno polje"	1	površinskom kopu "Tamnava-Zapadno polje"	1
kopa Tamnava Zapadno Polje	4	površinskom kopu TAMNAVA ZAPAD	2
kopa Tamnava-Zapad	1	površinskom kopu Tamnava Zapadno Polje	1
kopovima Tamnava Zapad	1	rudniku Tamnava - Zapad	1
kopu "Tamnava – Zapad"	1	Tamnava – Zapad	1
kopu "Tamnava-Zapadno polje"	1	Tamnava – Zapadno polje	1
kopu Tamnava – zapadno polje	1	Tamnava Zapadno Polje	8
P.K. TAMNAVA ZAPAD	1	Tamnava Zapadno polje	1
P.K. TAMNAVA-ZAPAD	3	Tamnava-Zapadno polje	1
		Tamnave Zapadnog polja	1

Из табеле се може уочити низак одзив јер је, од укупно 64 појављивања, пронађено свега 13 и то 5 за основне канонске облике - леме (поља освенчена светло-плаво бојом) и додатних 8 за флективне облике (поља освенчена сивом

⁴¹ Регуларни израз је формула којом се описује регуларни скуп ниски. Детаљније видети у (Vitas, 2006).

бојом). Разлог непроналажења лежи у различитим ортографским⁴² варијантама, а не флективним и семантичким варијантама. Овај проблем је могуће превазићи коришћењем синтаксичких графова у окружењу Unitex (Raumier, 2016). На слици 6.6 приказан је синтаксички граф за препознавање свих ортографских варијанти које задовољавају полазну информациону потребу, док су у табели 6.3 наведени одговарајући препознати изрази (са фреквенцијама), добијени применом графа на рударски корпус.



Слика 6.6. Синтаксички граф који препознаје ортографске варијанте назива *Tamnava Zapadno polje* (горе) са извученим конкорданцама (доле)

Граф представљен на слици 6.6 садржи лексичке изразе за препознавање различитих ортографских варијанти које се у појављују тексту, укључујући и званичне документе. Уместо *površinski kop* често се користи скраћеница *P.K.* (са

⁴² ортографски – који се односи на писање, писмени, који одговара правопису; правописни <http://www.vokabular.org/?lang=sr-lat&search=ORTOGRAFSKI>

или без тачака), али и термин *rudnik*, док се понекад атрибут *površinski* изоставља испред именице *kor*. Уз то, или се користе и различити знаци наводника (" , " , " , „) или се пак они потпуно изостављају. Понекад се у имену опционо користи и цртица (-, -, -). Такође, у истом значењу се наводе *Tamnava*, *Tamnava Zapad*, *Tamnava Zapadno Polje* и слично. Више токена у једном чвору графа представља алтернацију, тј. на тој позицији се може појавити један, било који од наведених токена, Етикета <E> у чвору графа означава празну реч, односно описује да произвољан токен тог чвора има опционо појављивање, односно да се ниједан од токена тог чвора не мора појавити на тој позицији. Лексеме (леме) уоквирене угластим заградама ("<>") препознају све одговарајуће флективне облике наведене леме, на пример:

<кор> замењује израз:

кор+кора+кори+кором+корови+корова+корове+коровима.

Представљени граф препознаје 64 појављивања ортографских и лексичких варијанти упита *Tamnava Zapadno polje*, од којих 37 има различите облике.

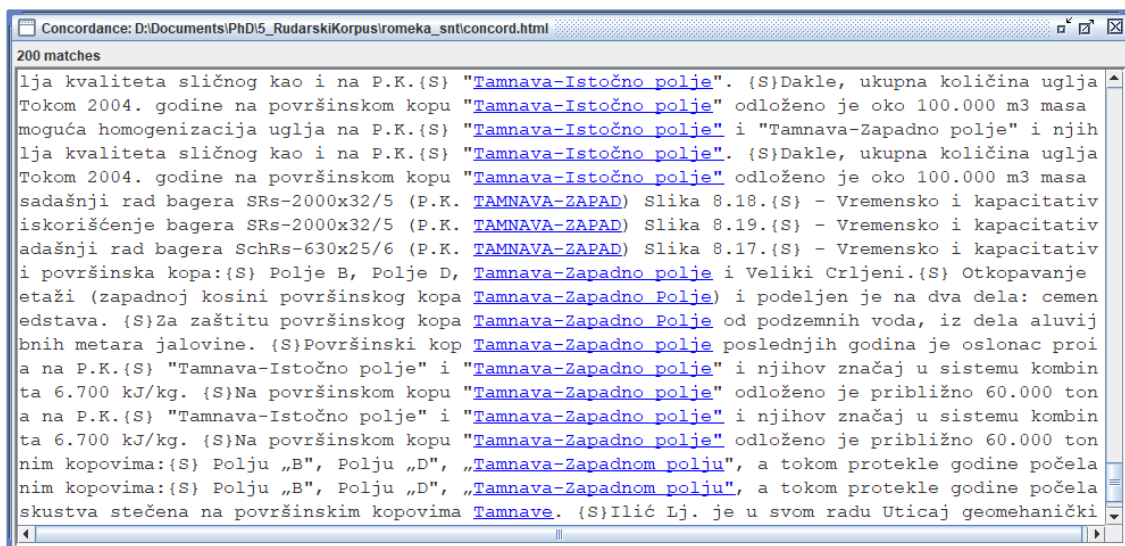
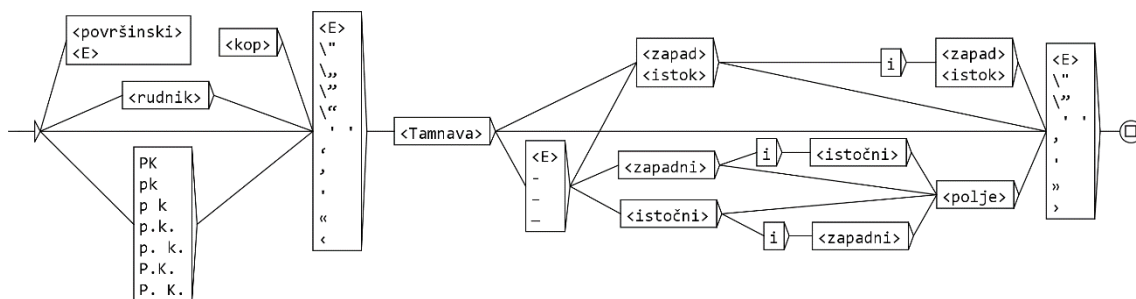
Синтаксички граф приказан на слици 6.7 уопштава претходни граф, па поред ортографске варијанте назива *Tamnava Zapadno polje*, препознаје и ортографске варијанте назива *Tamnava Istočno polje*.

Коришћењем лексичких маски са семантичким маркерима различити објекти у тексту се додатно могу идентификовати (Paumier, 2016). Тако на пример лексичка маска:

<(A>+<E>><N+DOM=Mining+DOM=Surface+Instrum>

препознаје термине (и било који њихов флективни облик) који се односе на рударску механизацију или опрему која се користи на површинском копу: транспортна трака, дреглајн, роторни багер итд.

Систем за екстракцију терминологије, описан у (Stanković *et al.*, 2016), може се унапредити обележавањем кандидата чије компоненте су из области рударства, тако да рецимо термин "багер" постане окидач за екстракцију термина роторни багер, кашика багера, ротор багера, багер утоварач, хидраулички багер итд.

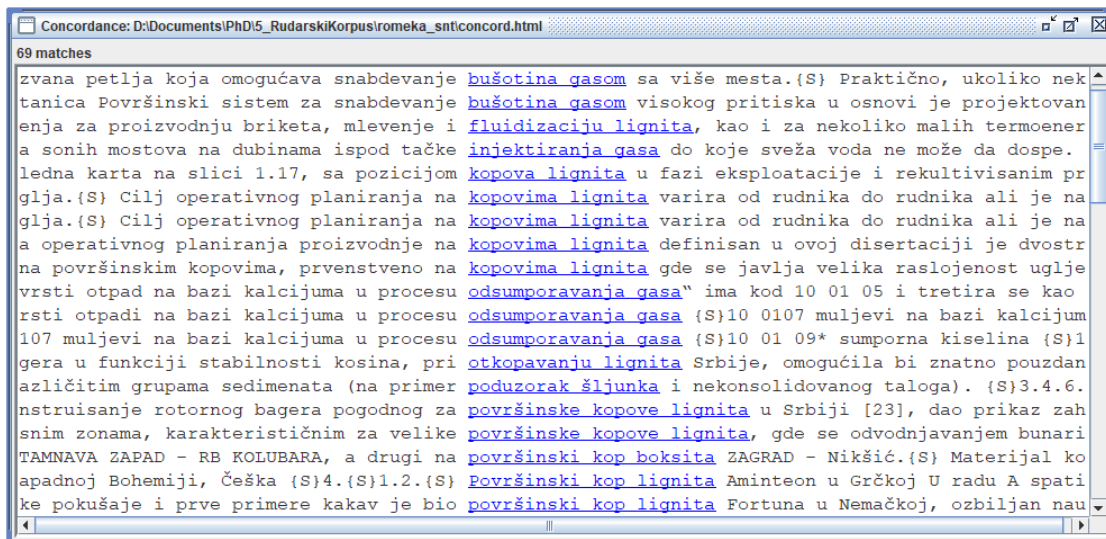
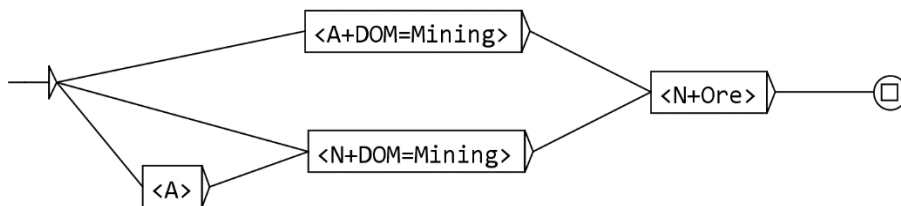


Слика 6.7. Синтаксички граф који у рударском корпусу препознаје ортографске варијанте назива тамнавских поља (горе), са извученим конкорданцама (доле)

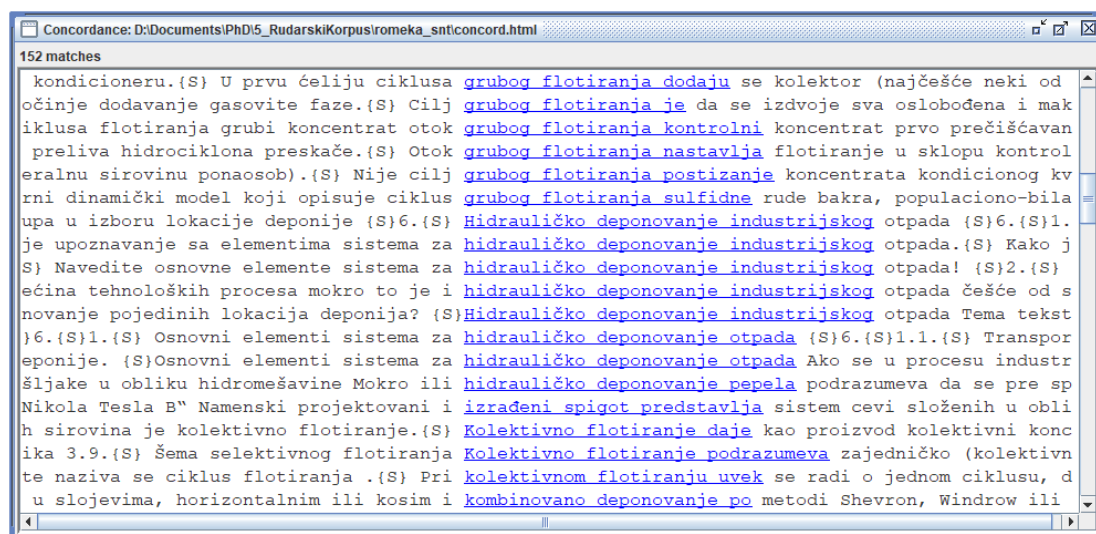
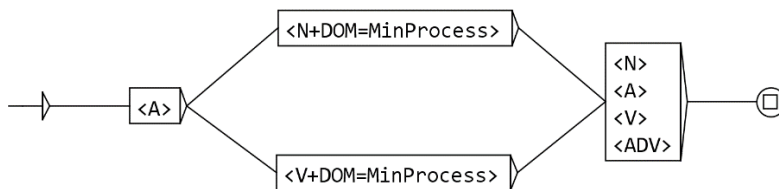
Слика 6.8 приказује граф који препознаје и приказује контекст за именицу обележену семантичким маркером "+Ore" (у значењу: нека врста руде), при чему њен леви контекст може бити придев (A) обележен доменским маркером "+DOM=Mining" (домен рударства) или пак именица такође из домена рударства, којој опционо претходи придев.

Резултат упита над рударским корпусом даје 80 појављивања од којих су неки "fluidizaciju lignita", "injektiranja gasa", "kopova lignita", "odsumporavanje gasa", "poduzorak šljunka", "površinski kop lignita", "površinski kop boksita". Пример добијених конкорданци приказан је слици 6.8.

Синтаксички граф приказан на слици 6.9. приказује екстракцију појмова из рударског корпуса који се односе на припрему минералних сировина. Граф препознаје фразе у којима именици (<N>) или глаголу (<V>) из домена припреме минералних сировина претходи придев (<A>), а за којим следи нека именица (<N>), придев (<A>), глагол (<V>) или прилог (<ADV>).



Слика 6.8. Синтаксички граф који у рударском корпусу препознаје назив руде (*горе*), са извученим конкорданцама (*доле*)



Слика 6.9. Граф и резултат његове примене за екстракцију примера обрасца из домена припреме минералних сировина

6.3. ПРЕТРАЖИВАЊЕ КОРПУСА РЕГУЛАРНИМ ИЗРАЗИМА

У дигиталним библиотекама, типичан резултат претраге је листа метаподатака о документима, а ако је потребно извлачење сегмената појединачних докумената, морају се применити додатне технике које нису интегрални део стандардних развојних система за дигиталне библиотеке. Конкорданце су један од често коришћених начина проучавања текста у лингвистици. На пример: упоређивање различитих употреба исте речи, анализирање кључних речи, анализирање фреквенција речи, проналажење и анализирање фраза, идиома и терминологије. Технике креирања конкорданци се широко користе у националним корпусима, али постоје и самосталне апликације које генеришу конкорданце и врше статистичку анализу над њима.

Већина конкорданци произведених помоћу софтверских алата приказане су у KWIC (енг. *keyword in context*) и KWICn (енг. *keyword in center*) форматима, у којима је наведена кључна реч јасно означена масним (bold) текстом у конзистентном положају, у ограниченом текстуалном контексту, неколико речи пре кључне речи и неколико речи које следе. Овај формат је изузетно користан због тога што се кључна реч лако идентификује заједно са њеним контекстом.

Информациона потреба (*površinski kop Tamnava Zapadno polje*) представљена у графу 6.6 може се формулисати и преко упита у CQPweb-овом претраживачком пољу, регуларним изразом у следећем формату:

$$\begin{aligned} &(((\{površinski\})? \{kop\}) | \{rudnik\} | pk | (p (.)? k (.)?))? \\ &([\", \", \", „])? \{Tamnava\} ((-|-|-)? (((\{zapadni\} \{polje\}) | \{zapad\}))) \\ &([\", \", \", „])? \end{aligned}$$

У овој нотацији витичасте заграде { } означавају леме, обле заграде () се користе за груписање, ? се користи за само један произвољан карактер, (... | ... | ...) одговара једној од алтернатива одвојених са "|", док угласте заграде [] одговарају једаном од алтернативних знакова наведених у заградама.

Систем CQRweb враћа конкорданце са линком ка метаподацима о документу и линком ка ширем контексту препознате фразе. Уз то, систем омогућава прегледање фреквенција не само погодака већ и колокација (слика 6.10).

Your query "((({površinski})? {kop}) | {rudnik} | pk | (p (.)? k (.)?))? ([",","'"])? {Tamnava} ((-|-|-)? (({zapadni} {polje})|{zapad}))([",","'"])?" returned 64 matches in 10 different texts (in 3,542,016 words [172 texts]; frequency: 18.07 instances per million words), ordered randomly [0.001 seconds - retrieved from cache]

Show Page: 1 KWIC View Show in corpus order New query Go!

No	Filename	Solution 11 to 20	Page 2 / 7
11	monogr0121	Tamnava - istočno polje " Slika 5 . 4 . Tehnološki profil površinskog kopa " Tamnava - zapadno polje " Površinski kop " Tamnava - zapadno polje " prvobitno je bio namenjen	
12	doktor0099	Očigledno je na ovom primeru Bogutovog sela , kao i u slučaju Tamnave Zapadnog polja , presudan uticaj na izbor pošumljavanja imala apsolutno najniža cena od 1	
13	doktor0099	Metadata for text <i>monogr0121</i>	je na površinskom kopu Tamnava Zapadno Polje Spoljno bašnjeg odlagališta
14	monogr0121	Text identification code: monogr0121	Displaying extended context for query match in text <i>monogr0121</i> File info for text monogr0121 Go! Show tags Tehnološki profil površinskog kopa " Tamnava - zapadno polje " Površinski kop " Tamnava - zapadno polje " prvobitno je bio namenjen za potrebe snabdevanja ugljem buduće termoelektrane " Kolubara B " , instalisane snage u prvoj fazi 2 350 MW . Kako ova elektrana još uvek nije izgrađena , kop se trenutno koristi kao dopuna kapacitetu površinskog kopa " Tamnava - istočno polje " . Eksploatacija uglja na ovom kopu počela je 1995 . godine , a projektovani razvoj će se odvijati kroz dve faze : - prva faza obuhvata period do kraja eksploatacionog veka površinskog kopa " Tamnava - istočno polje " .
15	monogr0121	Naslov dokumenta: Upravljanje kvalitetom uglja	
16	monogr0121	Godina nastanka dokumenta: 2007	
17	monogr0121	Autor(i) dokumenta: Ignjatović Dragan Knežević Nikola Stanković Ranka	
18	monogr0121	Poreklo (tip) dokumenta: monografije	
19	monogr0121	Lista ključnih reči dokumenta: uglj ,kvalitet ,homogeniza kop ,Tamnava	
20	monogr0121	Jezik dokumenta: srpski	
21	monogr0121	No. words in text: 49673	
22	doktor0099	Prikaz i analiza rešenja Prikaz rešenja i Zapadno polje postavlja se formiranjem kvantitacione matrice odlicavanja predstavljenog u tabeli 6 . 9	

Your query "((({površinski})? {kop}) | {rudnik} | pk | (p (.)? k (.)?))? ([",","'"])? {Tamnava} ((-|-|-)? (({zapadni} {polje})|{zapad}))([",","'"])?" returned 64 matches in 10 different texts (in 3,542,016 words [172 texts]; frequency: 18.07 instances per million words)

Showing frequency breakdown of words in this query, at the query node; there are 31 different types and 64 tokens at this concordance position.

Breakdown position: Node Frequency breakdown of words only Go!

No.	Search result	No. of occurrences	Percent
1	Tamnava Zapadno Polje	12	18.75%
2	površinskog kopa Tamnava Zapadno Polje	7	10.94%
3	Površinski kop Tamnava Zapadno Polje	5	7.81%
4	" Tamnava - Zapadno polje "	5	7.81%
5	P . K . TAMNAVA - ZAPAD	3	4.69%
6	površinskom kopu " Tamnava - Zapadno polje "	3	4.69%
7	površinskom kopu TAMNAVA ZAPAD	2	3.13%
8	površinskog kopa Tamnava - Zapadno Polje	2	3.13%
9	kopa " Tamnava - zapadno polje "	2	3.13%
10	" Tamnava - Zapadno polje "	2	3.13%

Слика 6.10. Пример конкорданци (горе) генерисаних у CQRweb систему са фреквенцијама (доле)

6.4. ТЕСТИРАЊЕ УПИТА НАД ОНТОЛОГИЈОМ

Као што је већ речено у поглављу 5.5.2.4, за тестирање онтологије *RuDokOnto*, проверу њене доследности и логичке конзистентности, као и за постављање SPARQL упита над онтологијом, коришћен је програмски додатак Ontop 1.18.1 интегрисан у Protégé.

Основну структуру SPARQL упита чине (слика 6.11):

- **PREFIX** – дефинише "простор имена" (енг. *name space*);
- **SELECT** – дефинише податке које упит треба да врати;
- **FROM** – идентификује податке над којима ће се упит извршити;
- **WHERE** – дефинише део RDF графа на који се упит односи;
- **FILTER** – дефинише различите типове ограничења;
- **ORDER** – дефинише сортирање добијених података.

The screenshot shows the 'ontop query editor' window. The query text is as follows:

```

PREFIX rdo: <http://www.semanticweb.org/RuDokOnto#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?Rudnik ?PoslovniSubjekt ?RadniStatus ?TipRudnika ?Basen
FROM <http://www.semanticweb.org/RuDokOnto#>
WHERE {{{?PoslovniSubjekt rdo:upravljaRudnikom ?Rudnik}
        {?Rudnik rdo:imaStatus ?RadniStatus}
        {?Rudnik rdo:jeTipRudnika ?TipRudnika}
        {?Rudnik rdo:pripadaBasenu ?Basen}}
FILTER (?PoslovniSubjekt=rdo:EPS)}

ORDER BY (?RadniStatus) (?Rudnik) (?PoslovniSubjekt)

```

Below the query, the execution time is 0.188 sec. The 'Show' options are 'All' (checked) and 'Short IRI'. There are buttons for 'Attach Prefixes', 'Execute', and 'Save Changes'.

Basen	PoslovniSubjekt	Rudnik	RadniStatus	TipRudnika
KolubarskiUgljeniBasen	EPS	PoljeC	aktivanRudnik	povrsinskiKop
KolubarskiUgljeniBasen	EPS	PoljeD	aktivanRudnik	povrsinskiKop
KostolackiUglejniBasen	EPS	PoljeE	aktivanRudnik	povrsinskiKop
KolubarskiUgljeniBasen	EPS	PoljeG	aktivanRudnik	povrsinskiKop
KolubarskiUgljeniBasen	EPS	TamnavaZapad	aktivanRudnik	povrsinskiKop
KolubarskiUgljeniBasen	EPS	JuznoPolje	neaktivanRudnik	povrsinskiKop
KolubarskiUgljeniBasen	EPS	PoljeB	neaktivanRudnik	povrsinskiKop
KolubarskiUgljeniBasen	EPS	Tamnavalstok	neaktivanRudnik	povrsinskiKop
KolubarskiUgljeniBasen	EPS	VelikiCrljeni	neaktivanRudnik	povrsinskiKop
KolubarskiUgljeniBasen	EPS	Radljevo	rudnikURazvoju	povrsinskiKop

Слика 6.11. Структура SPARQL упита над онтологијом *RuDokOnto*

Упитом над онтологијом *RuDokOnto*, приказаним на слици 6.11, се траже: рудник, пословни субјект који управља рудником, радни статус рудника (активан, неактивни или рудник у развоју), тип експлоатације (површинска, подземна или подводна експлоатација) и басен којем рудник припада. Ограничење је да пословни субјект који управља рудник буде ЈП Електропривреда Србије.

Као што је раније поменуто, један од проблема обраде текста на српском језику је богата морфологија и слободан ред речи у реченици. Да би се све наведене ортографске варијанте и флективни облици интегрисали у један ентитет, неопходно је увести релацију која указује на лексичку варијанту форму *nazivLexVar* између субјекта *TamnavaZapad* и свих ових литерала (објеката), који стварају RDF тројке. Екстраховани изрази се користе за лексикализацију онтологије, где онтологије хватају знање, али не успевају да ухвате структуру и употребу термина у изражавању и упућивање на ово знање на природном језику. Лексикализација је препозната као један од важних задатака у области Семантичког веба. Мотивација је да се обезбеди лексички слој који представља унутрашњу језичку структуру онтолошких етикета (литерала објеката), који се даље могу користити за локализацију онтологије и вербализацију, онтолошко учење итд. (Buitelaar *et al.*, 2011). Пример (извод) лексикализације онтологије је приказан на слици 6.12.

Пример SPARQL упита који тражи званичан назив и ортографске варијанте рудника, приказан је на слици 6.13. Упитом је дефинисано ограничење само на површински коп Тамнава-Западно поље.

Пример на слици 6.14 представља SPARQL упит којим се тражи сва инвестиционо-техничка документација по фазама пројекта, уз ограничење да се тражи само документација која се односи на површински коп Дрмно.

```

<!-- http://www.semanticweb.org/RuDokOnto#TamnavaZapad -->
<owl:NamedIndividual rdf:about="http://www.semanticweb.org/RuDokOnto#TamnavaZapad">
  <rdf:type rdf:resource="http://www.semanticweb.org/RuDokOnto#Rudnik"/>
  <imaMinSirovinu rdf:resource="http://www.semanticweb.org/RuDokOnto#lignit"/>
  <imaStatus rdf:resource="http://www.semanticweb.org/RuDokOnto#aktivanRudnik"/>
  <jeAktivan rdf:resource="http://www.semanticweb.org/RuDokOnto#aktivanRudnik"/>
  <jePovršinskiKop rdf:resource="http://www.semanticweb.org/RuDokOnto#površinskiKop"/>
  <jeTipRudnika rdf:resource="http://www.semanticweb.org/RuDokOnto#površinskiKop"/>
  <pripadaBasenu rdf:resource="http://www.semanticweb.org/RuDokOnto#KolubarskiUgljeniBasen"/>
  <naziv>Површински коп Тамнава-Западно поље</naziv>
  <nazivLexVar>P.K. TAMNAVA ZAPAD</nazivLexVar>
  <nazivLexVar>P.K. TAMNAVA-ZAPAD</nazivLexVar>
  <nazivLexVar>Površinski kop Tamnava Zapadno Polje</nazivLexVar>
  <nazivLexVar>Površinski kop Tamnava-Zapadno polje</nazivLexVar>
  <nazivLexVar>Površinski kop Tamnava-zapadno polje</nazivLexVar>
  <nazivLexVar>Tamnava - Zapad</nazivLexVar>
  <nazivLexVar>Tamnava Zapadno Polje</nazivLexVar>
  <nazivLexVar>Tamnava Zapadno polje</nazivLexVar>
  [...]
  <nazivLexVar>површинског копа Тамнава Западно поље<nazivLexVar>
  <nazivLexVar>површинског копа Тамнава-Западно Поље<nazivLexVar>
  <nazivLexVar>површинског копа Тамнава-западно поље<nazivLexVar>
  <nazivLexVar>површинском копу ТАМНАВА ЗАПАД<nazivLexVar>
  <nazivLexVar>површинском копу Тамнава Западно Поље<nazivLexVar>
  <nazivLexVar>површинском копу Тамнава-Западно поље<nazivLexVar>
  <nazivLexVar>површинском копу Тамнава-западно поље<nazivLexVar>
  <nazivLexVar>руднику Тамнава - Запад<nazivLexVar>
</owl:NamedIndividual>

```

Слика 6.12. Онтолошке лексикализације

ontop query editor: Query Editor

```

PREFIX rdo: <http://www.semanticweb.org/RuDokOnto#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?Rudnik ?naziv ?nazivLexVar
WHERE {{?Rudnik rdo:naziv ?naziv}
       {?Rudnik rdo:nazivLexVar ?nazivLexVar}}
FILTER (?Rudnik = rdo:TamnavaZapad)}

ORDER BY (?nazivLexVar)

```

Execution time: 0.031 sec - Numb... Show: All Short IRI Attach Prefixes Execute Save Changes

Rudnik	naziv	nazivLexVar
TamnavaZapad	"Површински коп Тамнава-Западно поље"	"P.K. TAMNAVA ZAPAD"
TamnavaZapad	"Површински коп Тамнава-Западно поље"	"P.K. TAMNAVA-ZAPAD"
TamnavaZapad	"Површински коп Тамнава-Западно поље"	"Površinski kop Tamnava Zapadno Polje"
TamnavaZapad	"Површински коп Тамнава-Западно поље"	"Površinski kop Tamnava-Zapadno polje"
TamnavaZapad	"Површински коп Тамнава-Западно поље"	"Površinski kop Tamnava-zapadno polje"
TamnavaZapad	"Површински коп Тамнава-Западно поље"	"Tamnava - Zapad"
TamnavaZapad	"Површински коп Тамнава-Западно поље"	"Tamnava Zapadno Polje"
TamnavaZapad	"Површински коп Тамнава-Западно поље"	"Tamnava Zapadno polje"
TamnavaZapad	"Површински коп Тамнава-Западно поље"	"Tamnava – Zapad"
TamnavaZapad	"Површински коп Тамнава-Западно поље"	"Tamnava – Zapadno polje"
TamnavaZapad	"Површински коп Тамнава-Западно поље"	"Tamnava-Zapad"
TamnavaZapad	"Површински коп Тамнава-Западно поље"	"Tamnava-Zapadno polje"
TamnavaZapad	"Површински коп Тамнава-Западно поље"	"Tamnava-zapadno polje"

Слика 6.13. Пример упита који препознаје све ортографске варијанте назива примерка *TamnavaZapad*

ontop query editor: Query Editor

```

PREFIX rdo: <http://www.semanticweb.org/RuDokOnto#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?Rudnik (?InvTehDokumentacija as ?ID) ?naslov (?Faza as ?Faza_projekta)

WHERE {
  {{?InvTehDokumentacija rdo:naslov ?naslov}
  {{?InvTehDokumentacija rdo:seOdnosiNaRudnik ?Rudnik}
  {{?InvTehDokumentacija rdo:pripadaFazi ?Faza}
FILTER (?Rudnik=rdo:Drmno)}

ORDER BY (?ID)
    
```

Execution time: 0.0 sec - Number ... Show: All Short IRI Attach Prefixes Execute Save Changes

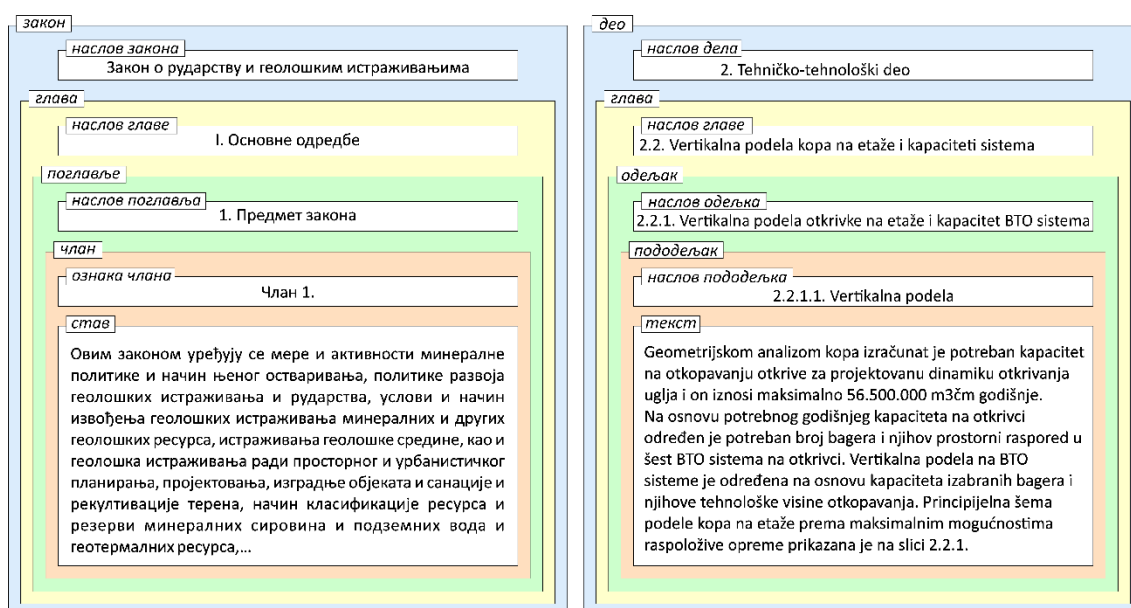
Rudnik	ID	naslov	Faza_projekta
Drmno	667	"Студија оправданости са идејним пројектом обезбеђивања потребни...	EvaluacijaProjekta
Drmno	668	"Студија оправданости са идејним пројектом увођења система за опе...	EvaluacijaProjekta
Drmno	670	"Техноекономска анализа израде екрана на површинском копу Дрмно"	Projektovanje
Drmno	672	"Упрошћени рударски пројекат одлагања откривке на унутрашњем о...	Projektovanje
Drmno	674	"Пројектни задатак за израду Главног рударског пројекта површинско...	Projektovanje
Drmno	675	"Књига I.1: Основна концепција"	Projektovanje
Drmno	676	"Књига I.2: Основна концепција"	Projektovanje
Drmno	677	"Пројектни задатак за израду Техничког пројекта уклапања нове опре...	Projektovanje
Drmno	678	"Књига II.1: Технички пројекат уклапања нове опреме у постојеће ста...	Projektovanje
Drmno	679	"Пројектни задатак за израду Техничког пројекта откопавања, транспо...	Projektovanje
Drmno	680	"Књига II.2: Технички пројекат откопавања, транспорта и одлагања отк...	Projektovanje
Drmno	681	"Пројектни задатак за израду Техничког пројекта откопавања и трансп...	Projektovanje

Слика 6.14. Пример упита који препознаје инвестиционо-техничку документацију по фазама пројекта

6.5. ОБЕЛЕЖАВАЊЕ ДОКУМЕНТА ПРЕМА СМЕРНИЦАМА ТЕI P5

Према (ISO 9001:2015) документ се дефинише као средство за преношење информација и комуницирање. Одликују га две формалне структуре: *логичка* – која описује садржај документа, и *графичка* – која описује организацију текста (на пример стране, пасуси, редови,...). Логичка и графичка структура не утичу једна на другу (Eickel, 1990).

Основна логичка структура документа биће приказана на примерима закона ("Закон о рударству и геолошким истраживањима") (Vasiljević, 2015), правилника ("Правилник о садржини студије изводљивости експлоатације лежишта минералних сировина"), пројектног задатка и документа који припадају рударској инвестиционо-техничкој документацији ("Студија оправданости са идејним пројектом обезбеђивања потребних количина угља за рад постојећих термоелектрана у "ТЕ-КО" Костолац и новог блока БЗ (350MW)"), слика 6.16.



Слика 6.16. Нивои обележавања текста закона и студије

Са слике 6.16 (лево) се може уочити да је законска регулатива обележена по нивоима: *закон*, *глава*, *поглавље* и *члан* (Vasiljević, 2015). Пројектна документација обележена је по нивоима: *део*, *глава*, *одељак* и *пододељак* (слика 6.16, десно).

Рударски корпус који сачињава 172 документа (законска регулатива и пројектна документација, литература и др.), као што је већ напоменуто у одељку 5.2.2, морфосинтаксички је аотиран у смислу да је сваком токену аутоматски придружена информација о одговарајућој врсти речи и леми. За потребе тестирања структурног аотирања, припремљени су текстови за мањи корпус, при чему је изабрано пет карактеристичних докумената. Из колекције законске регулативе обрађени су: *Закон о рударству и геолошким истраживањима*, *Правилник о садржини рударских пројеката* и *Правилник о садржини студије изводљивости експлоатације лежишта минералних сировина*, а из колекције пројектне документације обрађени су: *Пројектни задатак* и *Студија оправданости са идејним пројектом обезбеђивања потребних количина угља за рад постојећих термоелектрана у "ТЕ-КО" Костолац и новог блока В3 (350MW)*.

Са циљем екстракције информација из дигиталних објеката, који су део дигиталне библиотеке ROneka@RGF, одлучено је да се објекти, који су у текстуалном облику, похране и у XML формату, у складу са Смерницама TEI P5 (TEI-Consortium, 2018), који представља *de facto* стандард за аотацију произвољних типова докумената, укључујући правне текстове и текстове пројектне документације.

Аотација у складу са Смерницама TEI P5 треба да омогући повезивање делова текста у пројектној документацији који упућују на чланове закона и реферисане законске регулативе и/или повезивање са другом пројектном документацијом, како би се касније у претраживању и екстракцији информација сегменти текста лоцирали на одговарајући начин.

Слика 6.17 приказује аотирање законске регулативе у складу са Смерницама TEI P5, и то: *Закон о рударству и геолошким истраживањима* и *Правилника о садржини студије изводљивости експлоатације лежишта минералних сировина*.

Слика 6.18 приказује аотирање рударске пројектне документације у складу са Смерницама TEI P5, и то: *Пројектног задатка* и *Студије оправданости са идејним пројектом обезбеђивања потребних количина угља за рад постојећих термоелектрана у "ТЕ-КО" Костолац и новог блока В3 (350MW)*.

Закон о рударству и геолошким истраживањима

I. Основне одредбе

1. Предмет закона

Члан 1.

Овим законом уређују се мере и активности минералне политике и начин њеног остварења, услови и начин извођења геолошких истраживања минералних и других геолошких ресурса, истраживања геолошке средине и геолошка истраживања ради просторног и урбанистичког планирања, пројекат изградње објеката и санације и рекултивације терена, начин класификације ресурса и рекултивације минералних сировина и подземних вода и геотермалних ресурса, експлоатација рударских објеката, постројења, машина и уређаја, извођење рударских радова, управљање рударским отпадом, поступци санације и рекултивације напуштених рударских објеката надзор над спровођењем овог закона.

Члан 2.

Геолошки завод Србије, као посебна организација, са својством правног лица обавља геолошка истраживања и друга геолошка истраживања, као и послове примењених геолошких истраживања од важности за Републику Србију, у складу са овим законом.

2. Појмови

Члан 3.

Појмови употребљени у овом закону имају следеће значење:

- 1) геолошка средина је део земљине коре коју чине: земљиште са зонама аера минералног прихрањивања биљака, стене, површинске и подземне воде, минерални и други геолошки ресурси;
- 2) геолошки ресурси обухватају: простор са својим геолошким, амбијенталним и другим карактеристикама, минералне ресурсе, ресурсе подземних вода и геотермалне ресурсе;

Правилник о садржини студије изводљивости експлоатације лежишта минералних сировина

Правилник је објављен у "Службеном гласнику РС", бр. 108/2006 од 1.12.2006. године.

Члан 1.

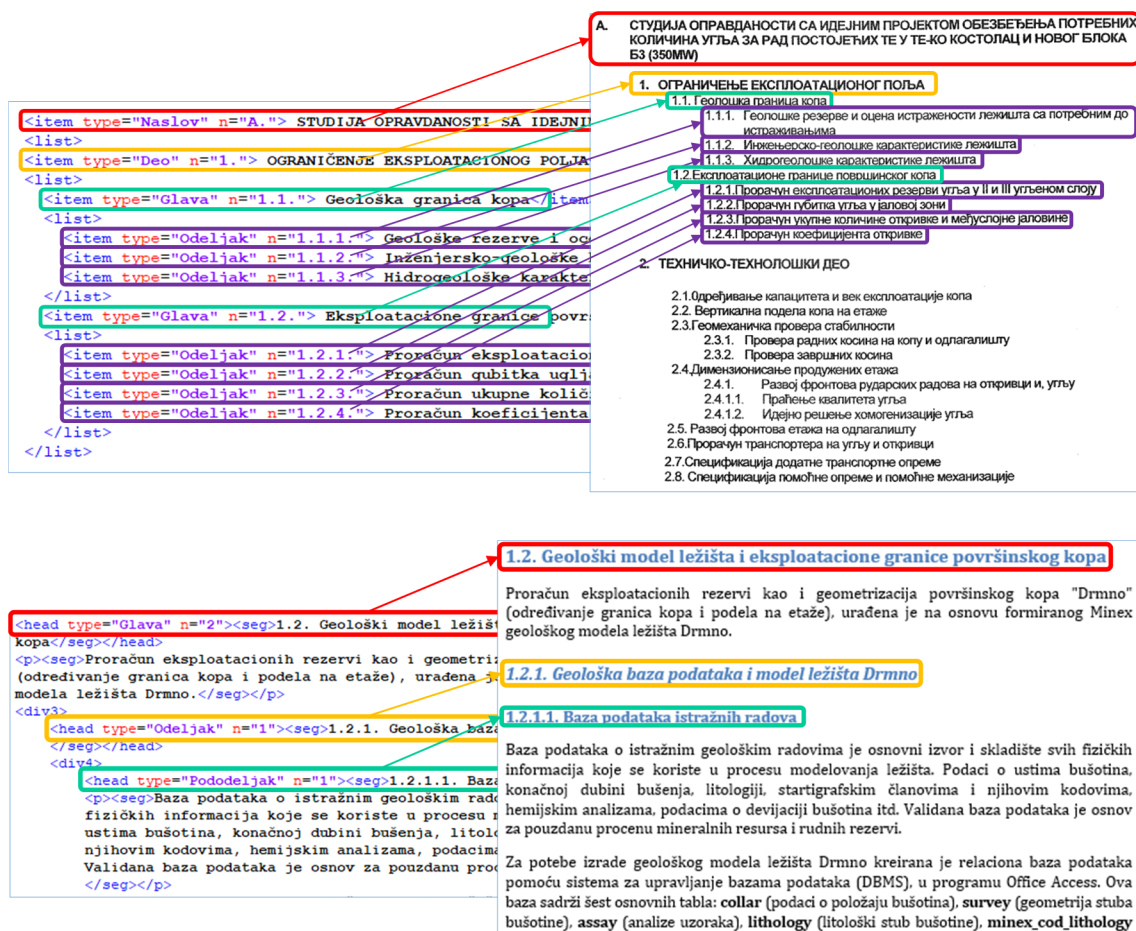
Овим правилником ближе се уређује садржина студије изводљивости експлоатације лежишта минералних сировина (у даљем тексту: студија).

Члан 2.

Студија садржи:

- 1) приказ услова и начина експлоатације и припреме минералних сировина;
- 2) приказ пласмана минералних сировина;
- 3) приказ мера заштите и санације животне средине и утицаја експлоатације на друштвену заједницу;
- 4) приказ потребних новчаних средстава која ће бити ангажована и броја запослених.

Слика 6.17. Анотирање законске регулативе у складу са Смерницама ТЕИ Р5



Слика 6.18. Анотирање рударске пројектне документације у складу са Смерницама ТЕИ Р5

За анотацију структурних јединица (региона) у оквиру корпусног текста, у складу са Смерницама ТЕИ Р5, коришћена је етикета `<div>`. Сlike 6.17 и 6.18 илуструју да се у документима сусрећемо са угњежденим структурним јединицама текста (једна структурна јединица текста унутар друге), што би се могло анотирати на следећи начин:

```
<div><div><div><div>[...]</div></div></div></div>
```

Међутим, приликом индексирања CWB-верзије корпуса угњеждени `<div>` региони би аутоматски били игнорисани (Evert, 2016a). Смернице ТЕИ Р5 нуде као решење да се региони унапред дефинисане дубине анотирају елементима за нумерисане јединице текста: `div1– div7`, при чему цифра на крају имена елемента означава дубину региона, па. би претходни пример изгледао овако:

```
<div1><div2><div3><div4>[...]</div4></div3></div2></div1>
```

За анотирање наслова и поднаслова коришћене су етикете `<head>` са одговарајућим атрибутима `type` и `n`. Вредности атрибута `type` су кључне за дистинкцију различитих делова документа (табела 6.1)).

Табела 6.1. Могуће вредности атрибута `type` етикете `<head>`

Врста документа	Вредности атрибута <code>type</code>
законска регулатива	glavni, glava, poglavlje или clan
пројектна документација	deo, glava, odeljak или pododeljak

Вредност атрибута `n` представља нумерацију (под)наслова. Нумерација поднаслова ће садржати нумерацију (над)наслова претходног нивоа и редни број самог поднаслова у односу на остале поднаслове на истом нивоу.

Пасуси у оквиру корпусног текста су анотирани етикетом `<p>`, а за анотацију листа коришћене су етикете `<list>` и `<item>`.

Изложени начин обележавања региона, када су вредности атрибута у питању, је само једно од могућих решења, чије даље имплементације ће свакако бити предмет даљих истраживања.

Анотација документа етикетама `<div>`, `<head>` и `<p>` приказана је у наредном прегледу:

```
<div1>
  <head type="glavni">Zakon o rudarstvu...</head>
  <div2>
    <head type="glava" n="I.">Osnovne odredbe</head>
    <div3>
      <head type="poglavlje" n="I.1.">Predmet zakona</head>
      <div4>
        <head type="član" n="1">Član 1.</head>
        <p>Ovim zakonom uređuju se mere...</p>
      </div4>
    </div3>
  </div2>
</div1>
```


Етикета <p> коришћена је као еквивалент тачке, подтачке и алинеје закона и подзаконских аката, односно пасуса у рударској пројектној документацији. За аотирање реченица коришћена је етикета <seg>. Начин аотирања пасуса и реченица дат је у наредном примеру:

```
<p><seg>Ležište uglja Drmno se nalazi u istočnom delu  
Kostolačkog ugljenog basena i zahvata površinu od oko 52 km2.  
</seg><seg> Prostire se istočno od reke Mlave, a granicu mu  
čine: Reka Dunav na severu, Boževačka greda na istoku, linija  
Bradarac - Sirakovačka dolina na jugu i reka Mlava na zapadu.  
</seg></p>
```

Имајући у виду представљање рударске пројектне документације, важно је очувању свих елемената текста, а посебно табела као важних извора информација. У овом истраживању, аотирање табела је рађено само на нивоу експериманта, али ће детаљније бити развијано у даљим истраживањима. У наредном прегледу илустрован је начин репрезентације/аотирања једног табеларног приказа у оквиру ТЕИ верзије пројектне документације. За сам корпус, због потребе унификације етикета за различите типове документације, припремљена је и верзија без аотације табеларних података.

Sistem	Bager	Moguća visina/dubina odlaganja, m	Projektovana dubina/visina etaže,m
VI BTO	A2RsB 2000	25/20	31 (16+15)
V BTO	A2RsB 2000	25/20	31 (16+15)
IV BTO	A2RsB 7200	25/20	31 (16+15)
III BTO	A2RsB 7200	25/20	31 (16+15)
II BTO	A2RsB 5500	18/15	30 (15+15)
I BTO	A2RsB 3500	17/15	30 (15+15)
Ukupno		200	200

```
<table rows="8" cols="4">  
<head type="Tabela" n="2.2.5.">Raspored odlagača po BTO sistemima</head>  
<row><cell>Sistem</cell><cell>Bager</cell><cell>Moguća visina/dubina odlaganja, m</cell>  
<cell>Projektovana dubina/visina etaže, m</cell></row>  
<row><cell>VI BTO</cell><cell>A2Rs B2000</cell><cell>25/20</cell><cell>31(16+15)</cell></row>  
<row><cell>V BTO</cell><cell>A2RsB 2000</cell><cell>25/20</cell><cell>31(16+15)</cell></row>  
<row><cell>IV BTO</cell><cell>A2RsB 7200</cell><cell>25/20</cell><cell>31(16+15)</cell></row>  
<row><cell>III BTO</cell><cell>A2RsB 7200</cell><cell>25/20</cell><cell>31(16+15)</cell></row>  
<row><cell>II BTO</cell><cell>A2RsB 5500</cell><cell>18/15</cell><cell>30(15+15)</cell></row>  
<row><cell>I BTO</cell><cell>A2RsB 3500</cell><cell>17/15</cell><cell>30 (15+15)</cell></row>  
<row><cell>Ukupno</cell><cell></cell><cell>200</cell><cell>200</cell></row>  
</table>
```

Пасус припремљеног документа са структурним апликацијама приказан је на примеру студије изводљивости:

```
<head type="Odeljak" n="1"><seg>1.1.1. Geološke rezerve i  
ocena istraženosti ležišta</seg></head>  
<p><seg>Prve istražne bušotine na prostoru Kostolačkog basena  
rađene su osamdesetih godina XIX veka, radi utvrđivanja  
krovine ugljenog sloja.</seg><seg> Istraživanja na prostoru  
ležišta "Drmno" izvođena su sa manjim ili većim prekidima do  
danas u više faza i to:</seg></p>  
<list type="unordered">  
  <item><seg>I faza: od 1949. do 1963. god.</seg></item>  
  <item><seg>II faza: od 1965. do 2002. god.</seg></item>  
  <item><seg>III faza: od 2002. do 2007. god.</seg></item>  
  <item><seg>IV faza: od 2008. do 2013. god.</seg></item>  
</list>
```

Транформација полазног текста у текст са структурним и морфолошким XML апликацијама подразумева додељивање врсте речи и леме оним токенима који су из текста, док етикете остају у изворном облику. Дакле, после обележавања структурним етикетама (<head>, <p>, <seg>, <list>, <item>, ...) врши се вертикализација текста. Резултат вертикализације је текст представљен као скуп колона, при чему прву колону чини комплетан корпусни текст (у једном реду прве колоне се налази тачно један токен корпусног текста), док се у осталим колонама налазе информације придружене токenu, пре свега, морфосинтаксичке информације: врста речи и лема одговарајућег токена (слика 6.19). Тако се рецимо уз реч "законa" додаје "N" као ознака да је именица и лема "закон", као основни, канонски облик речи. Тиме се обезбеђује да претрага кључне речи касније обухвати све флективне облике те речи: закон, законa, закону,...

У општем случају, апликација токена може да садржи и потпуније граматичке категорије о роду, броју падежу код именица, код глагола се могу аотирати време, вид, лице, и др., али за потребе овог рада то није било од значаја. Друга врста апликације, која може бити од интереса за овај рад је придруживање семантичких маркера речима, што би омогућило претраживање по семантичким обрасцима, како је приказано у одељцима 6.2 и 6.3. Модел

развијен у овом раду подразумева другачији приступ, тако да је предвиђено комбиновано претраживање по обрасцима у CWB корпусу уз подршку веб сервиса за експанзију упита. Ово изискује да се у наредном периоду измени изворни кôд интерфејса (веб сучеља) CQPweb.

<pre> <div1> <head type='glavni'> Закон N закон о ПРЕП о рударству N рударство и CONJ и геолошким А геолошки истраживањима N истраживање </head> <div2> <head type='glavni' n='I.'> Основне А основни одредбе N одредба </head> <div3> <head type='poglavlje' n='1.'> Предмет N предмет закона N закон </head> <div4> <head type='clan' n='1.'> Члан N Члан </head> </pre>	<pre> <div2> <head type='Deo' n='1.'> <seg> OGRANIČENJE N ograničenje EKSPLOATACIONOG A EKSPLOATACIONOG POLJA N polje </seg></head> <div3> <head type='Glava' n='1.1.'> <seg> Geološka А геолошки granica N granica kopa N kop </seg></head> <div4><head type='Odelxak' n='1.1.1.'> <seg> Geološke А геолошки rezerve N rezerva i CONJi ocena N ocena istraženosti N istraženost ležišta N ležište sa PREPsa potrebnim А potreban doistraživanjima N doistraživanjima </seg></head></div4> </pre>
---	--

Слика 6.19. Вертикализоваан и аотиран сегмент текста закона (лево) и студије (десно)

Имајући у виду крајње кориснике којима није блиско постављање комплексних упита, модел треба да омогући да се задата кључна реч на одговарајући пример допуни својим синонимима: коп/површински коп/рудник...

Друга планирана могућност проширења упита се своди на замену семантичких маркера листом термина из речника, укључујући њихову флективну парадигму, форматирану у облику који очекује алат за претрагу корпуса. На пример, маркер +Mach може да се замени листом термина: роторни багер, багер ведричар, камион, дозер, цевополагач, одлагач, грејдер, комбајн, депонијска машина, хидромонитор, ...

Имајући у виду обимност рударске документације у реалним системима која настаје током дугог века експлоатације, нарочито је важно обезбедити брзо проналажење докумената и ефикасну екстракцију релевантних сегмената текста у документу који може да има стотине страна. Публиковање линка ка метаподацима уз сегменте екстрахованог текста и линка ка месту у документу одакле се може видети шири контекст око екстрахованог текста се обезбеђује одговарајућим обележавањем докумената XML етикетама.

Приликом кодирања и индексирања корпуса помоћу CWB-алата неопходно је регистровати оне XML-елементе и XML-атрибуте који ће бити коришћени у будућим упитима као структурни атрибути корпуса (Evert, 2016a). Нерегистровани XML-елементи и атрибути се приликом кодирања третирају као обичан текст, а не као структурна анотација. Када се у упиту користи структурни атрибут корпуса који представља XML-атрибут, на њега се реферише комбиновањем имена одговарајућег XML-елемента који садржи тај XML-атрибут и имена самог XML-атрибута (Evert, 2016b). Тако на пример за атрибуте *type* и *n* елемента *head* се у упитима користи нотација *head_type*, односно *head_n*. Другим речима, интерно се XML-атрибути третирају на исти начин као и XML-елементи, тј. као да је XML-елемент

```
<head type="clan" n="1."> Наслов </head>
```

представљен као

```
<head><head_type><head_n> Наслов </head_n></head_type></head>
```

при чему су "елементима" *head_type* и *head_n* придружене редом вредности "clan" и "1."

Ограничавање вредности XML-атрибута у упитима се може постићи на различите начине. Један од могућих задавања упита користи комбиновано име (елемент_атрибут) одговарајућег структурног атрибута као назив XML-елемента, при чему се морају навести и отворена и затворена етикета, а само ограничавање вредности атрибута се наводи у отвореној етикети:

```
<head_type="clan"> []+ </head_type> ; (6.1)
```

Резултат упита (6.1) су поднаслови обележени XML-елементом `head` чији атрибут `type` мора имати вредност `clan`, при чему је поднаслов низ једног или више (+) произвољних токена.

Алтернативни начин задавања упита еквивалентног упиту (6.1) користи могућност да се одређеном сравњеном токenu придружи ознака (енг. *label*) помоћу које се реферише на токен и задају жељена ограничења. У примеру (6.2) се користи ознака `a` као референца на први сравњени токен, док се десно од симбола `::` наводи услов који тај токен мора да задовољи (елемент `head` коме припада означени токен мора имати атрибут `type` са вредношћу `clan`):

```
<head> a:[ ] []* </head> :: a.head_type = "clan" ; (6.2)
```

Исти ефекат као у упиту (6.2) се постиже упитом (6.3):

```
/region[head,a] :: a.head_type="clan" ; (6.3)
```

где се ознака `a` придружује региону ограниченом етикетама `<head>` и `</head>`, дакле, XML-елементу `head` који има `type` са вредношћу `clan`.

Резултат добијен извршавањем упита (6.3) приказан је на слици 6.20.

No	Filename	Solution 151 to 200	Page 4 / 6
168	zakoni0002	je doneo nadležni organ autonomne pokrajine , žalba se podnosi ministru .	Član 78 . Odobrenje za eksploataciju iz člana 77 . ovog zakona sadrži : 1
169	zakoni0002	tehničkim propisima , kao i uslovima utvrđenim rešenjima drugih nadležnih organa .	Član 79 . Nosilac rešenja kojim se odobrava eksploatacija nemetalnih mineralnih sirovina
170	zakoni0002	je doneo nadležni organ autonomne pokrajine , žalba se podnosi ministru .	Član 80 . Ministarstvo , odnosno nadležni pokrajinski organ ukinuće odobrenje za eksplo
171	zakoni0002	je doneo nadležni organ autonomne pokrajine , žalba se podnosi ministru .	Član 81 . Odobrenje za eksploataciju iz člana 77 . stav 1 . ovog zakona
172	zakoni0002	je doneo nadležni organ autonomne pokrajine , žalba se podnosi ministru .	Član 82 . Nemetalne mineralne sirovine za dobijanje građevinskih materijala i nemetalni
173	zakoni0002	dokumentacija za izvođenje rudarskih radova 1 . Vrste investicione i tehničke dokumentacije	Član 83 . Eksploatacija rezervi i resursa mineralnih sirovina izvodi se prema investiciono
174	zakoni0002	tehničko - tehnološke i ekonomske opravdanosti eksploatacije i izvođenja rudarskih radova .	Član 84 . Investiciono - tehničkom dokumentacijom u smislu ovog zakona smatra se : 1
175	zakoni0002	na istraživanju čvrstih mineralnih sirovina ; 6) uprosćeni rudarski projekat .	Član 85 . Investiciono - tehnička dokumentacija mora biti usklađena sa : 1) odredbama
176	zakoni0002	propisa iz oblasti zaštite od požara koje definišu sadržinu tehničke dokumentacije .	Član 86 . Prethodna studija opravdanosti je dokument koji se izrađuje u toku geoloških ist
177	zakoni0002	o opravdanosti ulaganja u dodatne istražne radove i izradu studije izvodljivosti .	Član 87 . Studija izvodljivosti eksploatacije ležišta mineralnih sirovina sadrži prikaz uslov
178	zakoni0002	tokom i potrebnim novčanim sredstvima i brojem angažovanih i zaposlenih lica .	Član 88 . Za eksploataciono polje za koje postoji akt Vlade , odnosno za eksploataciju
179	zakoni0002	ovog člana predstavlja stručnu podlogu za izradu prostornog plana posebne namene .	Član 89 . Rudarski radovi izvode se prema glavnom rudarskom projektu , dopunskom rud
180	zakoni0002	projektu za izvođenje rudarskih radova pri geološkim istraživanjima čvrstih mineralnih sirovina .	Član 90 . Glavni rudarski projekat izrađuje se u skladu sa studijom izvodljivosti eksploata
181	zakoni0002	osnovu rezervi mineralnih sirovina , na prostoru zahvaćenom projektovanim rudarskim radovima .	Član 91 . Dopunski rudarski projekat se izrađuje za odstupanja od glavnog rudarskog proj
182	zakoni0002	celine i stacionarne rudarske objekte i tehnološko - ekonomsku ocenu projekta .	Član 92 . Tehnički rudarski projekti izrađuju se u skladu sa glavnim i dopunskim rudarski
183	zakoni0002	građevinskih materijala i sadrži i tehničko rešenje rekultivacije prostora zahvaćenog eksploatacijom .	Član 93 . Rudarski projekat na istraživanju čvrstih mineralnih sirovina predstavlja projekta
184	zakoni0002	održavanje podzemnih prostorija , objekata i instalacija po završetku istražnih radova .	Član 94 . Uprosćeni rudarski projekat izrađuje se za : 1) sva manja odstupanja
185	zakoni0002	otklanjanje posledica akcidentnih situacija u trajanju ne dužem od godinu dana .	Član 95 . Nosilac eksploatacije je dužan da izradi godišnji operativni plan , kao i
186	zakoni0002	i godišnjeg izveštaja o poslovanju iz stava 1 . ovog člana .	Član 96 . Glavni projektant i odgovorni projektanti koji su izradili rudarski projekat , pisan
187	zakoni0002	pričiniti drugoj strani , odnosno trećem licu . 2 . Tehnička kontrola	Član 97 . Tehnička kontrola se vrši za rudarske projekte iz člana 84 . stav
188	zakoni0002	Ministar bliže propisuje uslove i način vršenja tehničke kontrole rudarskih projekata .	Član 98 . Izveštaj i potvrdu o tehničkoj kontroli rudarskog projekta izdaje privredni subjel

Слика 6.20. Резултат добијен извршавањем CQP упита коришћењем структурних атрибута корпуса

Упућивање на законски документ, са приказом метаподатака, и на део документа (тражени члан закона), приказан је на слици 6.21.

The screenshot displays a search interface with the following components:

- Search Results:** A table with columns 'No' and 'Filename'. It shows results for 'zakoni0002' with a snippet: 'je doneo nadležni organ autonomne pokrajine, žalba se podnosi ministru...'. A red arrow points from this snippet to the metadata view.
- Metadata View:** A panel titled 'Metadata for text zakoni0002' with a table of document details:

Text identification code	zakoni0002
Naslov dokumenta	Zakon o rudarstvu i geološkim istraživanjima
Godina nastanka dokumenta	2015
Autor(i) dokumenta	Ministarstvo rudarstva i energetike
Poreklo (tip) dokumenta	zakoni
Lista ključnih reči dokumenta	zakon, rudarstvo, geologija, istraživanje objekata, sanacija, rekultivacija, resursi, sirovina, upravljanje rudarskim otpadom
Jezik dokumenta	srpski
No. words in text	38542
- Extended Context View:** A panel titled 'Displaying extended context for query match in text zakoni0002' showing a snippet of the document text: 'Investiciono - tehnička dokumentacija za izvođenje rudarskih radova 1. Vrste investicione i tehničke dokumentacije **Član 83.** Eksploatacija rezervi i resursa mineralnih sirovina izvodi se prema investiciono - tehničkoj dokumentaciji za izgradnju rudarskih objekata i / ili izvođenje rudarskih radova, za eksploataciju nemetalnih mineralnih sirovina za dobijanje građevinskih materijala i za eksploataciju mineralnih resursa za dobijanje prirodnih građevinskih materijala. Dokumentacija iz stava 1. ovog člana izrađuje se na osnovu rezultata istraživanja, odnosno elaborata o resursima i rezervama, razvrstanih u skladu sa propisima o klasifikaciji resursa i rezervi, izveštaja o mineralnim resursima u slučaju eksploatacije prirodnih građevinskih materijala i druge dokumentacije kojima se razrađuju i analiziraju tehnički, tehnološki i ekonomski uslovi'.

Слика 6.21. Упућивање на документ и део документа

Овако приказана брза екстракција сегмената текста у једном документу је од великог значаја, с обзиром на уобичајено велики обим рударске документације. Примера ради, главни рударски пројекат може имати више десетина књига. Ако је потребно проверити да ли је корпус књига комплетиран свим деловима главног или допунског рударског пројекта, или техничког пројекти и др., као и да ли су испуњани сви захтеви дефинисани пројектним задатком, физички је немогуће то урадити за кратко време без софтверске подршке алатима и ресурсима засновним на језичким реурсима и технологијама презентованим кроз развијени модел за управљање рударском пројектном документацијом.

7. ЗАКЉУЧАК И ПРЕПОРУКЕ ЗА ДАЉИ РАД

Пројектовање и извођење рударских радова, као и геолошка истраживања, укључују дугорочно ангажовање великог броја стручњака, који свој посао често обављају на различитим, понекад веома удаљеним локацијама. У таквим радним условима, брз и ефикасан приступ потребним документима и подацима је од највеће важности.

У докторској дисертацији "Развој модела за управљање рударском пројектном документацијом" описан је развој система за управљање рударском пројектном документацијом, заснован на језичким технологијама. Такође, дат је детаљан опис рударске пројектне документације, према важећој законској регулативи Републике Србије и важећим домаћим и међународним стандардима.

Развијен је модел за управљање рударском пројектном документацијом имплементиран у систем, чији су основни елементи: дигитална библиотека, рударски корпус, лексички и термилошки ресурси и онтологија. Сваки од елемената детаљно је теоријски и практично обрађен и евалуиран.

Дигитална библиотека ROneka@RGF је централни репозиторијум за складиштење рударске пројектне документације, законске регулативе из домена рударства, безбедности и заштите на раду. Похрањени дигитални документи разврстани су колекције према врсти.

Посебна пажња посвећена је развијању рударског корпуса, у коме значајно место заузима процес обележавања текстова. У дисертацији је приказан начин обележавања текстова, посебно за екстракцију сегмената текста, а на примеру рударства илустровано је како се таква структура може користити и за друге домене.

Разматран је проблем формалних описа законске регулативе и пројектне документације, а потом механизам њиховог срањивања, што чини језгро система за управљање рударском пројектном документацијом. Овај проблем

до сада није истраживан, а имајући у виду све већу количину електронски доступне документације, јавила се потреба за одговарајућим решењем.

По први пут у Србији, креирана је онтологија која покрива домен рударства и класификује рударску пројектну документацију чиме се омогућава екстракција знања.

Употреба система илустрована је примерима који показују претраживање по кључним речима, без или са експанзијом упита, претраживање на основу регуларних израза, претраживање корпуса базирано на локалним граматицама, а затим и праћење тока информација базираних на овој претрази и коначно претраживање помоћу лексичких маске коришћењем доменских и семантичких маркера.

Предности оваквог система су бројне. Приступ документима могућ је са било које локације у било ком тренутку, без обзира на уређај који се користи. Архивирање докумената је централизовано, чиме се обезбеђује да сви документи буду на једном месту, чиме они постају лако доступни, једноставни за прегледање, измену и размењивање. Према неким анализама, запослени утроше 1,8 сати дневно, односно 9,3 сата недељно за тражење информација па је и економски ефекат примене оваквог система очигледан. Одржавање докумената на једној локацији осигурава симултано ажурирање и спречава рад на више беспотребних копија докумената, чиме се чува интегритет и тачност података. Праћење самог садржаја докумената осигурава да сви садржаји буду објављени на јединствен начин у складу са наменом и врстом документа. Осим форме, могуће је и усклађивање докумената са прописима. Системи управљања документима обезбеђују висок степен заштите докумената. Она се постиже кроз ригорозне безбедносне мере и контролу приступа. У случају непредвиђених околности, обезбеђује се да подаци, који су од виталног значаја за пословање, буду сачувани. Најзначајнија предност ових система је лако проналажење и преузимање докумената.

На основу досадашњих истраживања, поред опште анализе проблема, научни допринос огледа се у креирању доменског рударског корпуса. Доступна

документација, која је обухватала рударску пројектну документацију, законску регулативу, стандарде и стручну литературу, послужила је за допуну терминолошких речника рударским појмовима.

Анализиране су постојеће онтологије из домена блиских рударству, пре свега геологије, те је развијена нова онтологија прилагођена рударском домену, која описује рударску документацију и пратећу законску регулативу.

Развијен је модел за анотацију рударских информација и изграђен је систем за индексирање, напредно претраживање и екстракцију информација из рударског домена.

Развијене су методе за успостављање веза између постављеног проблема (спецификације пројекта) и сличних пројеката, између текстова сродних и сличних пројеката. Тестирању и валидацији приступило се на конкретним проблемима.

Даља научна истраживања ићи ћу у правцу сталне допуне морфолошких и терминолошких речника специфичних за области рударства и екстракцију вишечланих термина, њихову лематизација, обележавање маркерима и укључење у продукциони скуп речника, како би тагирање било успешније. Истраживања ће се базирати на унапређењу и иновирању система за морфолошку анотацију текста засновану на морфолошким речницима за срски језик, затим на даљем развоју вишејезичких терминолошких, семантичких и електронских речника из области рударства, као и проширењу рударског корпуса. Посебна пажња биће посвећена аутоматизацији процеса припреме термина за корпус и креирање система за аутоматско обележавање именованих ентитета специфичних за домен рударства. Планирано је и даље развијање аотирање табела, као важних носилаца информација. За рударске инжењере посебно је важна измена начина задавања упита који би био прилагођен корисницима који не поседују специфична информатичка знања.

ЛИТЕРАТУРА

- Abbott Jr, David M. (2017), Quality Assurance and Quality Control in Sampling and Sample Analysis, Mineral Property Evaluation: Handbook for Feasibility Studies and Due Diligence, p. 13.
- Arms, William Y. (2000), Digital libraries, MIT Press, London, p. 287.
- Baeza-Yates, Ricardo & Ribeiro-Neto, Berthier (1999), Modern information retrieval, ACM press New York.
- Barker, Philip (1994), Electronic libraries-visions of the future, The electronic library, Vol. 12 No. 4, pp. 221-230.
- Berners-Lee, Tim (1999), Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor.
- Bhatt, Dilip (2003), Excellence Model and Knowledge Management Implications, 2000, Disponível na Internet via www. URL: <http://www.eknowledgecenter.com/articles/1010/1010.htm>. Arquivo capturado em, Vol. 21 No. 01.
- Blodgett, Chris, Crowell, Michael & Lahaise, Karen (2005), Mobilizing Knowledge: Status of KM in Defence, Bravo Defence 2005, Vol. 5 No. 2005, pp. 5-7.
- Borst, Willem Nico (1997), Construction of engineering ontologies for knowledge sharing and reuse, Centre for Telematics and Information Technology (CTIT), University of Twente, Enschede, p. 243.
- Brzaković, Miodrag & Straživuk, Duška (2016), Informacije i znanje kao ključni resursi za potrebe interaktivnog učenja, Primus-informatika, pravo, ekonomija, bankarstvo, Vol. 1 No. 1.
- Buitelaar, Paul, Cimiano, Philipp, McCrae, John, Montiel-Ponsoda, Elena & Declerck, Thierry (2011), Ontology lexicalisation: The lemon perspective, in Slodzian, M., *et al.* (eds), 9th International Conference on Terminology and Artificial Intelligence, Paris, France, INALCO, pp. 33-36.
- Bullock, Richard L (2011), Mineral property feasibility studies, SME Mining Handbook, Vol. 1.
- Bullock, Richard L (2017), Introduction to Mineral Property Feasibility Reporting, Mineral Property Evaluation: Handbook for Feasibility Studies and Due Diligence, p. 1.
- Bush, Vannevar (1945), As We May Think, The atlantic monthly, Vol. 176 No. 1, pp. 101-108.
- CBRR (2018), Comissão Brasileira de Recursos e Reservas [Online]. Available: <http://cbrr.org.br/> [Accessed 17.01.2018].
- CGI (2018), Commission for the Management and Application of Geoscience Information [Online]. Available: <http://www.cgi-iugs.org/> [Accessed 15.01.2018].

- Chambers, Les (2018), Software in practice [Online]. Available: http://www.chambers.com.au/glossary/knowledge_management.php [Accessed 04.03.2018].
- Chaouni, Mamoun (2015), 7 Powerful Advantages of Using a Document Management System [Online]. Business 2 Community. Available: <https://www.business2community.com/tech-gadgets/7-powerful-advantages-using-document-management-system-01148648> [Accessed 10.03.2018].
- CIM (2018), Canadian Institute of Mining, Metallurgy and Petroleum [Online]. Available: <https://www.cim.org/> [Accessed 17.01.2018].
- Cleveland, Gary (1998), Digital libraries: definitions, issues and challenges, IFLA, Universal dataflow and telecommunications core programme.
- Comisionminera (2018), Comisión calificadora de competencias en recursos y reservas mineras [Online]. Available: <http://www.comisionminera.cl/> [Accessed 17.01.2018].
- Corcho, Oscar, Fernández-López, Mariano & Gómez-Pérez, Asunción (2003), Methodologies, tools and languages for building ontologies. Where is their meeting point?, Data & knowledge engineering, Vol. 46 No. 1, pp. 41-64.
- CQP (2017), The IMS Open Corpus Workbench (CWB), CQPweb [Online]. Available: <http://cwb.sourceforge.net/cqpweb.php> [Accessed 21.12.2017].
- CRIRSCO (2018), Committee for Mineral Reserves International Reporting Standards [Online]. Available: <http://www.cirisco.com> [Accessed 17.01.2018].
- Dalkir, Kimiz (2005), Knowledge Management in Theory and Practice.
- Davenport, Thomas H & Prusak, Laurence (1998), Working knowledge: How organizations manage what they know, Harvard Business Press.
- DCMI (2017), Dublin Core Metadata Initiative [Online]. Available: <http://dublincore.org/> [Accessed 29.12.2017].
- Đokić, Dragan (2012), Model portala za inteligentno upravljanje elektronskim dokumentima, Doktorska disertacija, Универзитет у Београду, Факултет организационих наука,
- Đorđević, Bojana P. (2017), Izrada osnova formalne gramatike srpskog jezika upotrebom metagramatike, Doktorska disertacija, Univerzitet u Beogradu, Filološki fakultet,
- EarthResourceML. Available: <http://www.earthresourceml.org/>.
- EarthResourceML (2018), EarthResourceML [Online]. Available: <http://www.earthresourceml.org/> [Accessed 15.01.2018].
- Eickel, Jürgen (1990), Logical and layout structures of documents, Computer Physics Communications, Vol. 61 No. 1-2, pp. 201-208.
- Evans, M Max, Dalkir, Kimiz & Bidian, Catalin (2014), A Holistic View of the Knowledge Life Cycle: The Knowledge Management Cycle (KMC) Model, Electronic Journal of Knowledge Management, Vol. 12 No. 2.

- Evert, Stefan & The CWB Development Team (2016a), The IMS Open Corpus Workbench (CWB) - Corpus Encoding Tutorial - CWB Version 3.4.
- Evert, Stefan & The CWB Development Team (2016b), The IMS Open Corpus Workbench (CWB) - CQP Query Language Tutorial - CWB Version 3.4.
- Evert, Stefan & Hardie, Andrew (2011), Twenty-first Century Corpus Workbench: Updating a Query Architecture for the New Millennium, in Corpus Linguistics Conference 2011, Birmingham, University of Birmingham.
- Fenton, Paul (2014), 10 Benefits of Moving to an Electronic Document Management System (EDMS) [Online]. Montréal, Québec, Canada: Montrium. Available: <https://blog.montrium.com/blog/10-benefits-of-moving-to-electronic-document-management-for-life-science-companies> [Accessed 15.12.2017].
- GeolISS (2017), Geološki informacioni sistem Srbije [Online]. Available: <http://geoliss.mre.gov.rs/recnik/TerminoloskiResursi.aspx> [Accessed 04.11.2017].
- GeoSciML (2018), Geoscience Markup Language [Online]. Available: <http://www.geosciml.org/> [Accessed 15.01.2018].
- GILS (2018), Global Information Locator Service [Online]. Available: <http://www.gils.net/> [Accessed 02.03.2018].
- Girard, John & Girard, J (2015), Defining knowledge management: Toward an applied compendium, Online Journal of Applied Knowledge Management, Vol. 3 No. 1, pp. 1-20.
- Gruber, Thomas R (1993), A translation approach to portable ontology specifications, Knowledge acquisition, Vol. 5 No. 2, pp. 199-220.
- Gruninger, Michael, Bodenreider, Olivier, Olken, Frank, Obrst, Leo & Yim, Peter (2008), Ontology Summit 2007–Ontology, taxonomy, folksonomy: Understanding the distinctions, Applied Ontology, Vol. 3 No. 3, pp. 191-200.
- Guarino, Nicola (1998), Formal ontology and information systems, in Proceedings of FOIS, Vol. 98, pp. 81-97.
- Hodge, Gail (2001), Metadata made simpler, Niso Press, Bethesda, Maryland, USA.
- Horridge, Matthew, Knublauch, Holger, Rector, Alan, Stevens, Robert & Wroe, Chris (2004), A practical guide to building OWL ontologies using the Protégé-OWL plugin and CO-ODE tools edition 1.0, University of Manchester.
- ISO 9001:2015, Quality management systems. Requirements, International Organization for Standardization, Genève, Switzerland.
- ISO 15836-1:2017, Information and documentation - The Dublin Core metadata element set, International Organization for Standardization, Genève, Switzerland.
- JORC (2018), Joint Ore Reserves Committee [Online]. Available: <http://www.criresco.com> [Accessed 17.01.2018].
- Jurafsky, Dan & Martin, James H (2014), Speech and Language Processing, Pearson, London.

- King, William R (2009), Knowledge management and organizational learning, Springer.
- Kolonja, Ljiljana (2016), Sistem poslovne inteligencije za upravljanje zaštitom na radu u rudarskoj industriji, Doktorska disertacija, Univerzitet u Beogradu, Rudarsko-geološki fakultet, Beograd.
- Krstev, Cvetana (2002), Digitalne biblioteke-razgraničenje pojmova, Infoteka, Vol. 3 No. 1-2, pp. 3-12.
- Krstev, Cvetana (2008), Processing of Serbian - Automata, Text and Electronic Dictionaries, Faculty of philology, Belgrade.
- Krstev, Cvetana, Obradović, Ivan, Utvić, Miloš & Vitas, Duško (2014), A system for named entity recognition based on local grammars, Journal of Logic and Computation, Vol. 24 No. 2, pp. 473-489, Oxford Journals, doi:10.1093/logcom/exs079, first published online February 19, 2013.
- Krstev, Cvetana, Pavlovic-Lazetic, Gordana, Vitas, Duško & Obradovic, Ivan (2004), Using textual and lexical resources in developing Serbian wordnet, Romanian Journal of Information Science and Technology, Vol. 7 No. 1-2, pp. 147-161.
- Krstev, Cvetana, Pavlović-Lažetić, Gordana, Obradović, Ivan & Vitas, Duško (2003), Corpora issues in validation of Serbian WordNet, in Matoušek, V., *et al.* (eds), 6th International Conference on Text, Speech and Dialogue, Czech Republic, Springer, Berlin, Heidelberg, pp. 132-137.
- Krstev, Cvetana, Stankovic, Ranka, Obradovic, Ivan & Lazic, Biljana (2015), Terminology Acquisition and Description Using Lexical Resources and Local Grammars, in Poibeau, T., *et al.* (eds), 11th Conference on Terminology and Artificial Intelligence (TIA), Granada, Spain, pp. 81-89.
- Krstev, Cvetana, Stankovic, Ranka, Vitas, Dusko & Obradovic, Ivan (2008a), The Usage of Various Lexical Resources and Tools to Improve the Performance of Web Search Engines, in Sixth International Conference on Language Resources and Evaluation (LREC '08), Marrakech, Morocco, European Language Resources Association (ELRA).
- Krstev, Cvetana, Vitas, Duško & Pavlović-Lažetić, Gordana (2008b), Resources and Methods in the Morphosyntactic Processing of Serbo-Croatian, in Formal Description of Slavic Languages: The Fifth Conference, pp. 3-17.
- Krstev, Cvetana, Zečević, Anđelka, Vitas, Duško & Kyriacopoulou, Tita (2013), NERosetta for the Named Entity Multi-lingual Space, in Vetulani, Z., *et al.* (eds), 6th Language and Technology Conference, LTC 2013, Poznań, Poland, pp. 327-340.
- Kucsma, Jason, Reiss, Kevin & Sidman, Angela (2010), Using Omeka to Build Digital Collections: The METRO Case Study, D-Lib magazine, Vol. 16 No. 3/4.
- Lazić, Biljana, Seničić, Danica, Tomašević, Aleksandra & Zlatić, Bojan (2016), Terminological and Lexical Resources Used to Provide Open Multilingual Educational Resources, in The Seventh International Conference on eLearning (eLearning-2016), Belgrade, Serbia.

- Lee, Susanna (2014), Omeka Guidelines [Online]. Available: <http://susannalee.org/courses/?p=111> [Accessed 23. may 2017].
- Legg, Catherine (2007), Ontologies on the semantic Web, Annual review of information science and technology, Vol. 41 No. 1, pp. 407-451.
- LibraryMine (2017), InfoMine, LibraryMine [Online]. Available: <http://www.infomine.com/library/> [Accessed 15.12.2017].
- Lynch, Clifford (1997), Searching the internet, Scientific American, Vol. 276, pp. 52-56.
- Liotard, Jean-François (1993), Excerpts from The postmodern condition: A report on knowledge, A postmodern reader, pp. 71-90.
- Manning, Christopher, Raghavan, Prabhakar & Schütze, Hinrich (2008), Introduction to Information Retrieval, Cambridge University Press, New York, USA.
- Maurel, Denis, Friburger, Nathalie, Antoine, Jean-Yves, Eshkol, Iris & Nouvel, Damien (2011), Cascades de transducteurs autour de la reconnaissance des entités nommées, Traitement automatique des langues, Vol. 52 No. 1, pp. 69-96.
- McCarthy, Philip M & Jarvis, Scott (2010), MTL, vocd-D, and HD-D: A Validation Study of Sophisticated Approaches to Lexical Diversity Assessment, Behavior Research Methods, Vol. 42 No. 2, pp. 381-392.
- McGuinness, Deborah L (2005), Ontologies come of age, Spinning the semantic web: bringing the World Wide Web to its full potential, p. 171.
- McKenna, Frank (2008), A Knowledge Management System - A Discourse [Online]. Knowledgeone Corporation. Available: http://www.academia.edu/4764955/A_Knowledge_Management_System_A_Discourse_A_Knowledge_Management_System_-_A_Discourse [Accessed 23.07.2016].
- Mentes, Hilal Sevindik (2012), Design and Development of a Mineral Exploration Ontology, Doctoral Dissertation, ScholarWorks@ Georgia State University,
- METS (2017), Metadata Encoding and Transmission Standard [Online]. Available: <http://www.loc.gov/standards/mets/>.
- Milenković, Marko (2003), Dublin Core Metadata Initiative (DCMI), Review of the National Center for Digitization, No. 2, pp. 70-79.
- Mining, MPIGM - Mongolian Professional Institute of Geosciences and.
- Ministarstvo kulture i informisanja (2017), Smernice za digitalizaciju kulturnog nasleđa u Republici Srbiji, Beograd, Ministarstvo kulture i informisanja Republike Srbije.
- Mladenović, Miljana (2016), Informatički modeli u analizi osećanja zasnovani na jezičkim resursima, Doktorska disertacija, Univerzitet u Beogradu, Matematički fakultet, Beograd.
- MongoDB (2017), MongoDB [Online]. Available: <https://www.mongodb.com/> [Accessed 19.12.2017].
- Musen, Mark A. (2015), The protégé project: a look back and a look forward, AI matters, Vol. 1 No. 4, pp. 4-12.

- MySQL (2017), MySQL [Online]. Available: <https://www.mysql.com/> [Accessed 19.12.2017].
- Navigli, Roberto & Velardi, Paola (2004), Learning domain ontologies from document warehouses and dedicated web sites, *Computational Linguistics*, Vol. 30 No. 2, pp. 151-179.
- NCSA (2017), The National Center for Supercomputing Applications [Online]. Available: <http://www.ncsa.illinois.edu/> [Accessed 11.11.2017].
- Neches, Robert, Fikes, Richard E, Finin, Tim, Gruber, Thomas, Patil, Ramesh, Senator, Ted & Swartout, William R (1991), Enabling technology for knowledge sharing, *AI magazine*, Vol. 12 No. 3, p. 36.
- Nirenburg, Sergei & Raskin, Victor (2004), *Ontological semantics*, MIT Press.
- Nonaka, Ikujiro & Takeuchi, Hirotaka (1995), *The knowledge-creating company: How Japanese companies create the dynamics of innovation*, Oxford university press.
- NoSketch (2017), NoSketch Engine [Online]. Available: <https://nlp.fi.muni.cz/trac/noske> [Accessed 21.12.2017].
- Nürnberg, Peter J, Furuta, Richard, Leggett, John J, Marshall, Catherine C & Shipman III, Frank M (1995), Digital Libraries: Issues and Architectures, in *Digital Libraries*, pp. 147-153.
- Obradović, Ivan, Tomašević, Aleksandra, Stanković, Ranka & Lazić, Biljana (2017), Uvođenje domenskih i semantičkih markera za oblast rudarstva u srpske elektronske rečnike, in Dragičević, R., *et al.* (eds), *Naučni sastanak slavista u Vukove dane*, Vol. 46, pp. 147-158.
- OCLC (2017), Online Computer Library Center [Online]. Available: <https://www.oclc.org/en/home.html> [Accessed 01.10.2017].
- OGC (2017), Open Geospatial Consortium [Online]. Available: <http://www.opengeospatial.org/> [Accessed 21.12.2017].
- Omeka (2017), Omeka Classic User Manual [Online]. Available: https://omeka.org/codex/Managing_Users_2.0 [Accessed 23.06.2017].
- OWL (2017), Web Ontology Language [Online]. Available: <https://www.w3.org/TR/owl-features/> [Accessed 23.06.2017].
- Paumier, Sébastien (2016), UNITEX 3.1, Université Paris-Est Marne-la-Vallée.
- PERC (2018), Pan-European Reserves & Resources Reporting Committee [Online]. Available: <http://www.percstandard.eu/> [Accessed 17.01.2018].
- Petrušić, Darko (2016), *Semantičko modelovanje i ontološka integracija informacionih sistema Otvorene vlade*, Doktorska disertacija, Univerzitet u Novom Sadu, Fakultet tehničkih nauka, Novi Sad.
- Poursaba, Chris (2015), *The Top 7 Benefits of Document Management Systems* [Online]. Alexandria, Virginia, USA: Meridian Imaging Solutions. Available: <http://www.whymeridian.com/blog/the-top-7-Benefits-of-Document-Management-Systems> [Accessed 01.03.2018].

- PSUL (2017), Penn State University Libraries, Mining and Mineral Resources [Online]. Available: <http://guides.libraries.psu.edu/miningminresources> [Accessed 11.11.2017].
- Raskin, Rob (2006), Guide to sweet ontologies, NASA/Jet Propulsion Lab, Pasadena, CA, USA, <http://www.geospatialweb.com/sweet.jpl.nasa.gov/guide.doc>.
- RDF (2017), Resource Description Framework [Online]. Available: <https://www.w3.org/RDF/> [Accessed 15.08.2017].
- RDFS (2017), Resource Description Framework Schema [Online]. Available: <https://www.w3.org/TR/rdf-schema/> [Accessed 15.08.2017].
- Rosson, Mark (2017), 8 Advantages to a Document Management System [Online]. O'Neil Digital Solutions. Available: <https://www.oneildata.com/insight/blog-post/2017/8-advantages-to-a-document-management-system> [Accessed 28.02.2018].
- Saffady, William (1996), The document life cycle: a White Paper, Association for Information and Image Management International.
- Sağsan, Mustafa (2006), A new life cycle model for processing of knowledge management, in 2nd International Congress of Business, Management and Economics, pp. 15-18.
- SAMREC (2018), South African Code For The Reporting Of Exploration Results, Mineral Resources And Mineral Reserves [Online]. Available: <https://www.samcode.co.za/samcode-ssc/samrec> [Accessed 17.01.2018].
- Schmid, Helmut (1999), Improvements in Part-of-Speech Tagging with an Application to German, in Armstrong, S., *et al.* (Eds.) Natural Language Processing Using Very Large Corpora, Springer, Dordrecht.
- Schmid, Helmut (2013), Probabilistic Part-of-Speech Tagging Using Decision Trees, in Jones, D. B., *et al.* (Eds.) New methods in language processing, Routledge, p. 154.
- Sekine, Satoshi & Ranchhod, Elisabete (2009), Named entities: recognition, classification and use, John Benjamins Publishing.
- SharePoint (2018), Microsoft SharePoint [Online]. Available: <https://products.office.com/en-us/sharepoint/collaboration> [Accessed 19.02.2018].
- Shiri, Ali (2003), Digital library research: current developments and trends, Library Review, Vol. 52 No. 5, pp. 198-202.
- Shreve, Gregory (2002), Process for the document management and computer-assisted translation of documents utilizing document corpora constructed by intelligent agents, Google Patents.
- Službeni glasnik RS (27/97a), Pravilnik o sadržini dugoročnog programa eksploatacije ležišta mineralnih sirovina i godišnjih planova izvođenja rudarskih radova, Beograd, JP "Službeni glasnik".
- Službeni glasnik RS (27/97b), Pravilnik o sadržini rudarskih projekata, Beograd, JP "Službeni glasnik".
- Službeni glasnik RS (101/2015), Pravilnik o sadržini rudarskih projekata, Beograd, JP "Službeni glasnik".

- Službeni glasnik RS (108/2006), Pravilnik o sadržini studije izvodljivosti eksploatacije ležišta mineralnih, Beograd, JP "Službeni glasnik".
- Službeni glasnik RS (135/2004, 36/2009), Zakon o proceni uticaja na životnu sredinu, Beograd, JP "Službeni glasnik".
- SME (2018), Society for Mining, Metallurgy, and Exploration [Online]. Available: <https://www.smenet.org/> [Accessed 17.01.2018].
- Smith, Robin (2000), Aristotle's logic [Online]. Stanford Encyclopedia of Philosophy. Available: <http://plato.stanford.edu/entries/aristotle-logic/>.
- Sprague, Ralph H. Jr (1995), Electronic document management: Challenges and opportunities for information systems managers, MIS Quarterly, pp. 29-49.
- Stanković, Ranka (2009), Modeli ekspanzije upita nad tekstuelnim resursima, Doktorska disertacija, Univerzitet u Beogradu, Matematički fakultet, Beograd.
- Stanković, Ranka, Krstev, Cvetana, Lazić, Biljana & Škorić, Mihajlo (2018), Electronic Dictionaries – from File System to lemon Based Lexical Database, in 6th Workshop on Linked Data in Linguistic (LDL-2018), Towards Linguistic Data Science (in print).
- Stanković, Ranka, Krstev, Cvetana, Obradović, Ivan & Kitanović, Olivera (2017), Improving Document Retrieval in Large Domain Specific Textual Databases Using Lexical Resources, Transactions on Computational Collective Intelligence XXVI, Springer, pp. 162-185.
- Stanković, Ranka, Krstev, Cvetana, Obradović, Ivan, Lazić, Biljana & Trtovac, Aleksandra (2016), Rule-based Automatic Multi-word Term Extraction and Lemmatization, in Calzolari, N., *et al.* (eds), 10th International Conference on Language Resources and Evaluation, LREC 2016, Portorož, Slovenia, The European Language Resources Association (ELRA), Paris, France, pp. 507-514.
- Stanković, Ranka, Krstev, Cvetana, Obradović, Ivan, Trtovac, Aleksandra & Utvić, Miloš (2012), A Tool for Enhanced Search of Multilingual Digital Libraries of E-journals, in Eighth International Conference on Language Resources and Evaluation (LREC), Istanbul, Turkey, pp. 1710-1717.
- Stanković, Ranka, Trivić, Branislav, Kitanović, Olivera & Blagojević, Branislav (2011a), Razvoj geološkog terminološkog rečnika GeolISSTerm, Информотека, Vol. 12 No. 1, pp. 53-67.
- Stanković, Ranka, Trivić, Branislav, Kitanović, Olivera, Blagojević, Branislav & Nikolić, Velizar (2011b), The Development of the GeolISSTerm Terminological Dictionary, INFOtheca-Journal of Informatics & Librarianship, Vol. 12 No. 1, pp. 49-63.
- Stevens, Robert (2010), Mineral exploration and mining essentials, Pakawau GeoManagement Port Coquitlam, British Columbia, Canada.
- Stojanović, Cvjetko P. (2015), Model upravljanja investicionim projektima otbaranja površinskih kopova uglja, Doktorska disertacija, Univerzitet u Beogradu, Rudarsko-geološki fakultet, Beograd.
- Studer, Rudi, Benjamins, V. Richard & Fensel, Dieter (1998), Knowledge engineering: principles and methods, Data & knowledge engineering, Vol. 25 No. 1, pp. 161-197.

- Sugumaran, Vijayan & Storey, Veda C (2002), Ontologies for conceptual modeling: their creation, use, and management, *Data & knowledge engineering*, Vol. 42 No. 3, pp. 251-271.
- SWEET (2017), Semantic Web for Earth and Environmental Terminology [Online]. Available: <https://sweet.jpl.nasa.gov/> [Accessed 19.02.2017].
- TEI-Consortium (2018), TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 3.3.0. [Online]. Available: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/> [Accessed 06.04.2018].
- TEI (2017), Text Encoding Initiative [Online]. Available: <http://www.tei-c.org/Vault/P5/1.3.0/doc/tei-p5-doc/en/html/> [Accessed 04.05.2017].
- Tomašević, Aleksandra, Lazić, Biljana, Vorkapić, Dalibor, Škorić, Mihajlo & Kolonja, Ljiljana (2017), *Upotreba veb platforme Omeka za digitalne biblioteke iz domena rudarstva i geologije*, Infoteka.
- Trtovac, Aleksandra (2016), *Metadata Descriptors and Content Descriptors for Information Retrieval in Digital Libraries*, Doctoral Dissertation, University of Belgrade, Faculty of Philology, Belgrade.
- Tufiş, Dan (2004), Special Issue on BalkaNet Project, *Romanian Journal on Information Science and Technology*. Bucureşti: Publishing house of the Romanian Academy, Vol. 7 No. 1-2.
- UME (2018), *Underground mining engineering* [Online]. Available: <http://ume.rgf.bg.ac.rs/index.php/ume> [Accessed 30.03.2018].
- UNDL (2017), *United Nations Digital Library* [Online]. Available: <https://digitallibrary.un.org/> [Accessed 11.11.2017].
- Unitex (2017), *Unitex/GramLab* [Online]. Available: <http://unitexgramlab.org/> [Accessed 19.12.2017].
- Uschold, Mike & Gruninger, Michael (1996), *Ontologies: Principles, methods and applications*, *The knowledge engineering review*, Vol. 11 No. 02, pp. 93-136.
- Utvić, Miloš (2011), *Annotating the Corpus of Contemporary Serbian*, *INFOtheca - Journal of Informatics & Librarianship*, Vol. 12 No. 2 (December 2011), pp. 36a-47a.
- Utvić, Miloš (2014), *Izgradnja referentnog korpusa savremenog srpskog jezika*, Doktorska disertacija, Univerzitet u Beogradu, Filološki fakultet, Beograd.
- Uzelac, Milan (1998), *Uvod u filozofiju*, Vršac: Viša škola za obrazovanje vaspitača, I-Metaphysica generalis.
- Vasiljević, Nebojša (2015), *Automatska obrada pravnih tekstova na srpskom jeziku*, Doktorska disertacija, Univerzitet u Beogradu, Filološki fakultet, Beograd.
- Verma, Rajeev (2015), *Knowledge Management Process* [Online]. Available: <https://www.linkedin.com/pulse/rajeev-verma-knowledge-management-process-rajeev-verma?forceNoSplash=true> [Accessed 22.07.2016].
- Vickery, Brian C (1997), *Ontologies*, *Journal of information science*, Vol. 23 No. 4, pp. 277-286.

- Vitas, Duško, Krstev, C, Obradović, I, Popović, Lj & Pavlović-Lažetić, Gordana (2003), Processing Serbian Written Texts: An Overview of Resources and Basic Tools, in Piperidis, S., *et al.* (eds), Workshop on Balkan Language Resources and Tools in the Framework of the 1st Balkan Conference in Informatics - BCI 2003, Thessaloniki, Greece, Vol. 21, pp. 97-104.
- Vitas, Duško, Popović, Ljubomir, Krstev, Cvetana, Obradović, Ivan, Pavlović-Lažetić, Gordana & Stanojević, Mladen (2012), The Serbian Language in the Digital Age - Српски језик у дигитално доба, in Rehm, G., *et al.* (Eds.) META-NET White Paper Series, Springer, Heidelberg, New York, Dordrecht, London.
- Vujičić Stanković, Staša I. (2016), Ekstrakcija informacija vođena ontologijama (model za srpski jezik), Doktorska disertacija, Univerzitet u Beogradu, Matematički fakultet, Beograd.
- Wimalasuriya, Daya C & Dou, Dejing (2010), Ontology-based information extraction: An introduction and a survey of current approaches, Journal of Information Science, DOI 10.1177/0165551506nnnnnn.
- XML (2017), eXtensible Markup Language [Online]. Available: <https://www.w3.org/XML/> [Accessed 05.07.2017].
- XTM (2017), XML Topic Maps [Online]. Available: <http://www.topicmaps.org/xtm/> [Accessed 05.07.2017].
- Young, R., Institution, British Standards & KMS/1, British Standards Institution. Technical Committee (2003), Knowledge Management Vocabulary, PD 7500, BSI.
- Zack, Michael H (1999a), Developing a knowledge strategy, California management review, Vol. 41 No. 3, pp. 125-145.
- Zack, Michael H (1999b), Managing Codified Knowledge, Sloan Management Review, Vol. 40 No. 4, pp. 45-58.
- НАЭН (2018), Национальная ассоциация по экспертизе недр [Online]. Available: <http://naen.ru/> [Accessed 17.01.2018].

БИОГРАФИЈА

Александра Ђ. Томашевић, дипл. инж. руд., рођена је 17. августа 1962. године у Београду, Србија. Након завршене средње Грађевинско техничке школе, школске 1981/1982. године уписала је Рударско-геолошки факултет Универзитета у Београду, смер Израда подземних просторија. Студије је завршила са просечном оценом 8,31, а дипломски рад под називом "Идејно решење тунела у улици Тадеуша Кошћушког у Београду" одбранила је 1990. године, са оценом 10.

Од 1990. до 1994. године радила је као стипендиста Министарства просвете на Катедри за израду подземних просторија. Стални радни однос на Рударско-геолошком факултету засновала је 1994. године и радила је на радном месту стручног сарадника на Рударском одсеку. Од 1995. до 1998. године радила је као секретар пословодног органа Рударско-геолошког факултета Универзитета у Београду, а 1998. године прелази на радно место библиотекара. Од 2006. године ради на радном месту стручно-техничког секретара Рударског одсека.

Објавила је више научних и стручних радова. Учествовала је у више научно-истраживачких пројеката везаних за примену нових технологија у рударству, као и у изради студија и привредних пројеката која се односе на пројектовање информационих система. Као члан тима учествовала је у изради софтверских пакета у којима су примењене математичке методе за решавање проблема из рударства и заштите животне средине. Учествовала је у изради више информатичких пројеката у области рударства, од којих су најважнији: Систем за анализу сигурности и заштите на раду у рудницима (PROTECTOR) и Систем за управљање квалитетом угља (SUKU). Тренутно је ангажована на пројекту технолошког развоја "Унапређење технологије површинске експлоатације лигнита у циљу повећања енергетске ефикасности, сигурности и заштите на раду", који финансира Министарство просвете, науке и технолошког развоја Републике Србије.

Изјава о ауторству

Име и презиме аутора АЛЕКСАНДРА ТОМАШЕВИЋ

Број индекса P707/10

Изјављујем

да је докторска дисертација под насловом

РАЗВОЈ МОДЕЛА ЗА УПРАВЉАЊЕ РУДАРСКОМ ПРОЈЕКТНОМ

ДОКУМЕНТАЦИЈОМ

- резултат сопственог истраживачког рада;
- да дисертација у целини ни у деловима није била предложена за стицање друге дипломе према студијским програмима других високошколских установа;
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио/ла интелектуалну својину других лица.

Потпис аутора

У Београду, 22.03.2018.

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора АЛЕКСАНДРА ТОМАШЕВИЋ

Број индекса P707/10

Студијски програм РУДАРСКО ИНЖЕЊЕРСТВО

Наслов рада РАЗВОЈ МОДЕЛА ЗА УПРАВЉАЊЕ
РУДАРСКОМ ПРОЈЕКТНОМ ДОКУМЕНТАЦИЈОМ

Ментор Др БОЖО КОЛОЊА, редовни професор
Др РАНКА СТАНКОВИЋ, ванредни професор

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла ради похрањена у **Дигиталном репозиторијуму Универзитета у Београду**.

Дозвољавам да се објаве моји лични подаци везани за добијање академског назива 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.

У Београду, 22.03.2018.

Потпис аутора

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку "Светозар Марковић" да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

РАЗВОЈ МОДЕЛА ЗА УПРАВЉАЊЕ РУДАРСКОМ ПРОЈЕКТНОМ

ДОКУМЕНТАЦИЈОМ

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигиталном репозиторијуму Универзитета у Београду и доступну у отвореном приступу могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство (CC BY)
2. Ауторство – некомерцијално (CC BY-NC)
3. Ауторство – некомерцијално – без прерада (CC BY-NC-ND)
4. Ауторство – некомерцијално – делити под истим условима (CC BY-NC-SA)
5. Ауторство – без прерада (CC BY-ND)
6. Ауторство – делити под истим условима (CC BY-SA)

(Молимо да заокружите само једну од шест понуђених лиценци.

Кратак опис лиценци је саставни део ове изјаве).

У Београду, 22.03.2018.

Потпис аутора

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 прерада.** 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 прерада.** 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.