

Дигиталне библиотеке у рударству и геологији са посебним освртом на представљање сиве литературе

Биљана Лазић, Александра Томашевић, Михаило Шкорић



Дигитални репозиторијум Рударско-геолошког факултета Универзитета у Београду

[ДР РГФ]

Дигиталне библиотеке у рударству и геологији са посебним освртом на представљање сиве литературе | Биљана Лазић, Александра Томашевић, Михаило Шкорић | Научна конференција Библиоинфо — 55 година од покретања наставе библиотекарства на високошколском нивоу, Београд 18. мај 2017. | 2019 | |

10.18485/biblioinfo.2017.ch13

<http://dr.rgf.bg.ac.rs/s/repo/item/0004954>

Дигитални репозиторијум Рударско-геолошког факултета Универзитета у Београду омогућава приступ издањима Факултета и радовима запослених доступним у слободном приступу. - Претрага репозиторијума доступна је на www.dr.rgf.bg.ac.rs

The Digital repository of The University of Belgrade Faculty of Mining and Geology archives faculty publications available in open access, as well as the employees' publications. - The Repository is available at: www.dr.rgf.bg.ac.rs

МА Биљана Лазић¹

Др Александра Томашевић²

МА Михаило Шкорић³

Универзитет у Београду⁴

Рударско-геолошки факултет

ДИГИТАЛНЕ БИБЛИОТЕКЕ У РУДАРСТВУ И ГЕОЛОГИЈИ СА ПОСЕБНИМ ОСВРТОМ НА ПРЕДСТАВЉАЊЕ СИВЕ ЛИТЕРАТУРЕ

Сажетак

Имајући у виду потребу за проналажењем информација похрањених у различитим облицима документације која се генерише у областима рударства и геологије на Рударско-геолошком факултету Универзитета у Београду, отпочет је процес развоја дигиталне библиотеке *ROMeka@RGF*, на платформи за приказивање дигиталних колекција - Омека. Значајан део документације представља такозвана сива литература која је претежно заступљена у виду вишетомене документације. Први савладани изазов представљало је повезивање различитих вишетоменних делова пројектних извештаја у једну целину која би била лако доступна и претражива.

Кључне речи: дигиталне библиотеке, сива литература, Омека, језички ресурси, речници

1. Дигиталне библиотеке

Према једној од најцитиранијих дефиниција дигиталних библиотека, према Вилијаму Армсу (*William Arms*), дигиталне библиотеке су контролисане, системски организоване колекције информација

1 biljana.lazic@rgf.bg.ac.rs

2 aleksandra.tomasevic@rgf.bg.ac.rs

3 mihailo.skoric@rgf.bg.ac.rs

4 Рударско-геолошки факултет, Ђушина 7, 11000 Београд

са придруженим сервисима, ускладиштене у дигиталном формату, којима се приступа путем мреже (Arms 2000). Дигиталне библиотеке при високошколским установама представљају подршку учењу на даљину (Вранеш 2004: 222), као и подршку класичним студијским програмима. Иако јавне библиотеке у Србији предњаче са развојем дигиталних колекција, дигиталне библиотеке су развијене и у оквиру више високошколских установа и библиотека: на Филолошком факултету Универзитета у Београду⁵, у Универзитетској библиотеци „Светозар Марковић“⁶, на Филозофском факултету Универзитета у Новом Саду⁷ итд. Глобална искуства показују да дигиталне библиотеке које баштине материјале из области рударства и геологије настају при високошколским установама на којима се ове области изучавају кроз студијске програме. Пример такве библиотеке је *Ohio Geology Digital Library*⁸. Други мотив за креирање дигиталних библиотека из домена геологије јесте представљање пионирских и капиталних радова значајних за истраживаче. Примери библиотека које ово илуструју су: *Geology of Lehigh and Northampton Counties, Pennsylvania: Historical Works*⁹ и *Carnets géologiques de Philippe Glangeaud*.¹⁰

Инспирација за настанак дигиталне библиотеке на Рударско-геолошком факултету потекла је из потребе за обједињеним и брзим приступом референтној литератури од значаја за рударство и геологију. Рударство, сходно мултидисциплинарној природи, убрзано производи разнородне материјале као носаче информација. У њих се убрајају, картографска, нумеричка, графичка, и текстуелна грађа. Имајући ово у виду, појавила се потреба за систематичним прикупљањем и обрадом пројектне документације што није ништа ново имајући у виду да је уопште препознат значај адекватног чувања овакве литературе. Наиме, постоји читава научна област која се бави овим послом а то је систем пословне интелигенције. Захваљујући овој области препознат је значај правовременог и адекватног приступа свим подацима везаним за пројекат јер се у пракси неретко дешава да се, због разуђености технолошког система, кључни подаци загубе, што на дугорочном плану доводи у питање економске ефекте пројеката. Као резултат испуњења

5 <http://digbil.ananda.earhiva.com/Account/LogOn?ReturnUrl=%2f>

6 <http://www.unilib.rs/istorijske-novine/pretraga>

7 <http://digitalna.ff.uns.ac.rs/>

8 <https://library.osu.edu/>

9 <http://digital.lib.lehigh.edu/lvgeology/>

10 <http://bibliotheque.clermont-universite.fr/glangeaud/>

ових потреба, током рада на докторској дисертацији *Развој модела за управљање рударском пројектном документацијом* (Томашевић 2018), настала је дигитална библиотека *ROmeka@RGF*. Назив је настао од: *PO* (Рударски одсек), *Омека* (назив платформе за развој дигиталних колекција) и *РГФ* (Рударско-геолошки факултет).

2. Дигитална библиотека *ROmeka@RGF*

Дигитална библиотека *ROmeka@RGF* намењена је за складиштење, приказивање и управљање пројектном документацијом из области рударства, као и незаобилазних домена: безбедност и заштита на раду, заштита животне средине и процена ризика. Селекција докумената се базира на доступним дигиталним ресурсима и није рађено додатно скенирање докумената расположивих искључиво у папирном облику. Дигитална библиотека садржи 209 докумената, од чега су 172 документа стриктно из области рударства и она су послужила као подлога за истраживачки рад везан за екстракцију рударске терминологије, обележавање текста, екстракцију информација и др. (Томашевић и др. 2017). Окосницу дигиталне библиотеке чини 51 документ (пројектни задаци, студије и пројекти) који су део *“Главног рударског пројекта површинског копа “Дрмно” за капацитет од 9×10⁶ тона угља годишње”* који се радио на Рударско-геолошком факултету из Београда (Томашевић и др. 2018). Поред тога, дигитална библиотека садржи и целокупну актуелну законску регулативу Републике Србије из поменутих домена, као и стручну литературу која пројектантима и инжењерима омогућава проналажење потребних информација. Видљивост докумената на вебу је ограничена због ауторских права и поверљивости података.

Идеја је да ова врста литературе буде лако претражива и доступна, имајући у виду потребу за благовременом доступношћу информација приликом рада на рударским пројектима (Лазевић и др. Презентација).

На веб платформи *Омека* изворно претраживање докумената се врше преко кључних речи, булових оператора и потпуним подударавњем. У претраживање су укључени: метаподаци, документи, етикете, колекције итд. Претраживање у дигиталној библиотеци *ROmeka@RGF* унапређено је имплементирањем проширених упита. Коришћени су веб сервиси (Станковић и др. 2012) и морфолошки електронски речници за српски језик (Крстев и др. 2008, Станковић

и др. 2016) за морфолошко проширење упита (http://hlt.rgf.bg.ac.rs/vebran/api/delafs/ključna_reč) и семантичко-морфолошко проширење упита (http://hlt.rgf.bg.ac.rs/vebran/api/sinonimi/ključna_reč).

3. Сива литература

Значајан део документације похрањене у дигиталној библиотеци *ROMeka@RGF* представља такозвана сива литература. Она је у библиотечкој литератури дефинисана на различите начине. Једна од индивидуалних дефиниција је „непопуларни али драгоцени полуобјављени материјали са ограниченим бројем примерака који нису доступни путем комерцијалних књижарских канала какви су каталози издавача или књижаре“ (Okoroma 2012: 1).

Следећа „колективна“ и широко прихваћена дефиниција настала током Треће конференције о сивој литератури у Луксембургу 1997. године, тзв. „луксембуршка дефиниција“ дефинише сиву литературу као „производ свих нивоа управе, академских институција, пословних институција и индустрије у штампаном и електронском формату, али независно од комерцијалних издавача“ (Schöpfel и Farace 2010: 1). Она је 2004. године, током исте конференције у Њујорку приширена делом „неконтролисана од стране комерцијалних издавача, или где издаваштво није примарна делатност творца“ (2010: 1).

Можда је најсвеобухватнија и најсвежија тзв. Прашка дефиниција коју су 2010. године установили Фаранс и Шопфел. Ова дефиниција луксембуршкој додаје економски аспект као и аспект квалитета: „[Литература] ... која је заштићена законом о интелектуалној својини, довољног квалитета да би била прикупљена и чувана у оквиру библиотечких фондова и институционалних репозиторијума...“ (2010:5).

У сиву литературу се убрајају: технички извештаји, научни извештаји, пројектни извештаји, радни папири и зборници, билтени, тезе и дисертације, необјављена дела, материјали са конференција који нису лако доступни, белешке, технички цртежи, владине публикације, информативни материјали, итд. (Ћирковић 2018).

Оно што сиву литературу чини сивом јесте њена невидљивост и необавезност да буде описана библиографски. „Увођење нових библиографских стандарда са посебним пољима за сиву литературу био би један од начина да се увећа њихова видљивост и употребљивост.“ (2018: 85).

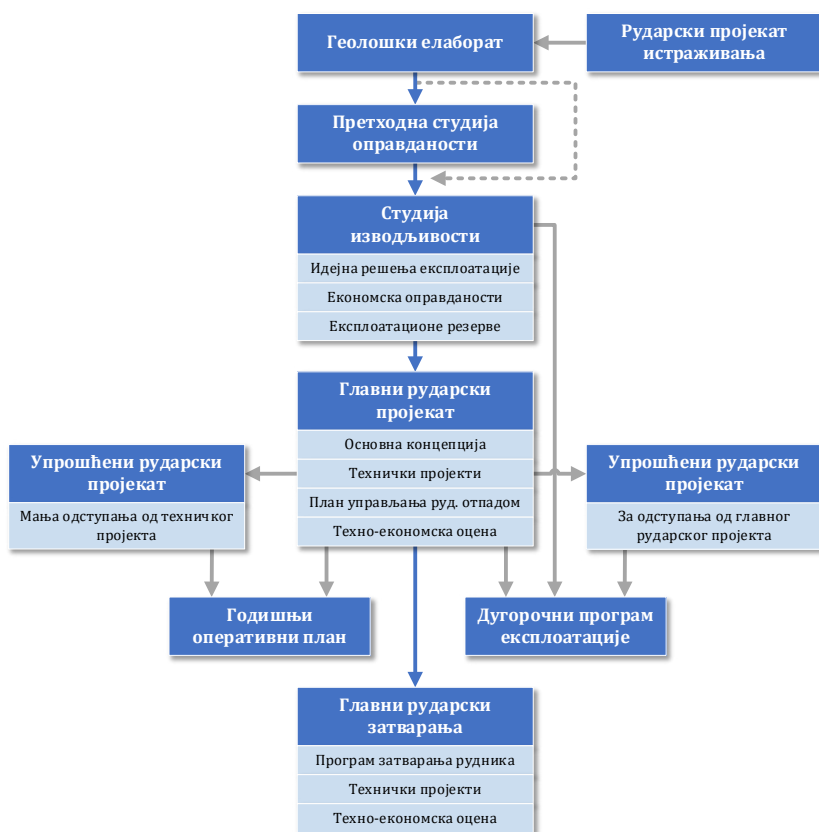
1.1 *ROMeka@RGF* и сива литература

Имајући у виду претходно изложене дефиниције долазимо до закључка да дигитална библиотека *ROMeka@RGF* располаже готово искључиво сивом литературом. Она је у претежно заступљена у виду вишетомене документације, најчешће мултидисциплинарних пројеката, законских прописа, правилника, као и текстова докторских дисертација. Први изазов представља повезивање различитих вишетоменних делова пројектних извештаја у једну целину која би била лако доступна и претражива. Још један од већих изазова јесте и природа документације пројеката који је симбиоза садржаја текстуалног, нумеричког, графичког и картографског типа.

Пројектна документација у рударству прати све фазе развоја рудника: анализу пројекта, истраживање лежишта, евалуацију пројекта, пројектовање и развој рударских радова, производњу и фазу затварања рудника (Томашевић 2018: 81-82). Век трајања ових фаза може бити од неколико година до више од сто година, стога је неопходно адекватно чување свих података. Инвестиционо-техничка документација у Републици Србији строго је прописана законом *Законом о рударству и геолошким истраживањима*¹¹ и прописује постојање: претходне студије оправданости, студије изводљивости експлоатације лежишта минералне сировине, дугорочног програма експлоатације, рударског пројекта и годишњег оперативног плана. Рударски извођачи пројекти могу бити: главни, допунски, технички, технички рударски пројекат за експлоатацију минералних ресурса за добијање природних грађевинских материјала, рударски пројекат на истраживању чврстих минералних сировина и упрошћени рударски пројекат (2018: 89). Како би се илустровала комплексност, на слици 1 дат је приказ структуре, редоследа израде и везе међу рударским пројектима. Поред наведене документације, присутна је и пратећа рударска документација која се може сврстати у групе: административне документације (услови, решења, дозволе, жалбе, итд.), додатних истражних радова (елаборати, цртежи, табеле, итд.), периодичних извештаја о раду рудника и појединих делова (извештаји надлежног особља о раду, стању опреме, постигнутим ефектима и капацитетима, итд.), извештаја о техничком одржавању (систематизација рокова и

11 http://romeke.rgf.rs/files/original/Zakoni/Zakon_o_rudarstvu_i_geoloskim_istrazivanjima/Zakon_o_rudarstvu_i_geoloskim_istrazivanjima.1.pdf

програма планираног одржавања, радови изведени као непланирано одржавање, итд.), извештаја из области безбедности и здравља на раду (акт о процени ризика, елаборат о испитивању услова радне средине, итд.) и мониторинг (извештаји и елаборати о стању животне средине, постигнутим резултатима, итд.) (2018: 99-100).



Слика 1 Структура, редослед израде и везе међу рударским пројектима (Томашевић 2018: 90)

Како би се омогућило представљање хијерархијски сложених релација међу деловима једног пројекта, развијена је онтологија *RuDokOnto* као онтологија OWL¹². Метаподаци објеката у дигиталној библиотеци *ROmeka@RGF*, представљени Даблинским језгром, повезани су са онтологијом *RuDokOnto* путем онтологије *protege-dc.owl*.¹³

На слици 2 представљена је визуелизација таксономије *Inv Teh Dokumentacija* која описује целокупну пројектну документацију која прати процес развоја појединачног рударског пројекта. Она представља део онтологије *RuDokOnto*. На слици се види да је пројектна документација подељена на: студије, планове и програме и пројектну документацију. Односно, подкласе *Studija*, *Projektna Dokumentacija*, *Plan Program* и *Izveštaj* су у релацији „is-a“ са класом *Inv Teh Dokumentacija*. Такође видимо да су подкласе *Prethodna Studija*, *Opravdanosti*, *Studija Izvodljivosti*, *Konceptualna Studija*, *Studija Uticaja Zatecenog Stanja* и *Studija Uticaja Na Zivu Sredinu* у релацији са подкласом *Studija*. Ова релација показује да су претходна студија оправданости, студија изводљивости, концептуална студија, студија утицаја затеченог стања и студија утицаја на животну средину врсте студије.

12 Web Ontology Language – OWL је језик за представљање онтологија.

13 Protege је програм за управљање онтологијама. Више на: <https://protege.stanford.edu/>



Слика 2 Визуелни приказ таксономије *InvTehDokumentacija*
(Томашевић 2018: 134)

Коришћење онтологије RuDokOnto је од значаја за процес екстракције информација. Резултат екстракције информација заснованог на онтологији су подаци издвојених из текста. Ови подаци се могу приказати помоћу језика за креирање онтологије (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.

4. Закључак

Сива литература након дигиталне револуције нема разлога да буде запостављена. Дигиталне платформе омогућавају релативно једноставан опис метаподатака и системе за складиштење докумената и брз приступ путем Интернета стога сива литература више не би требало да буде у другом плану. У прилог томе говори експанзија броја репозиторијума докторских дисертација и научних радова. У дигиталној библиотеци *ROMeka@RGF*, примењени ресурси и алати за српски језик, како би се корисницима на побољшан начин представила сива литература. Уз помоћ увођења онтологије која приказује однос елемената пројектне документације, кроз колекцију је могуће претраживати документе који припадају истом пројекту.

Извори и литература

- Arms, William Y. *Digital libraries*. Cambridge, Massachusetts, USA: MIT Press, 2000.
- Вранеш, Александра. *Високошколске библиотеке*. Београд: Конзорцијум TEMPUS пројекта UMI JEP 16059-2001, Универзитетска библиотека „Светозар Марковић“; Бањалука: Народна и универзитетска библиотека Републике Српске, 2004.
- Farance, Dominic и Joachim Schöpfel. „Introduction Grey Literature“. Dominic Farance, Joachim Schöpfel (ур.). *Grey Literature in Library and Information Studies*. Amsterdam: Walter de Gruyter, 2010. 1-9.
- Krstev, Cvetana, Ranka Stanković, Duško Vitas, Ivan Obradović. “The Usage of Various Lexical Resources and Tools to Improve the Performance of Web Search Engines”. *Sixth International Conference on Language Resources and Evaluation (LREC '08), Marrakech, Morocco*. Nicoletta Calzolari et al. (ur.). Marrakech : European Language Resources Association (ELRA), 2008. .
- Okoroma, Francisca. „Grey Literature Management at Southern Illinois University Edwardsville (SIUE) Library, USA: A Macarthur Foundation Grant Research Project“. *Library Philosophy and Practice*, 706 (2012) Веб. 10. 11. 2018.
- Stanković, Ranka, Cvetana Krstev, Ivan Obradović, Aleksandra Trtovac, Miloš Utvić. “A Tool for Enhanced Search of Multilingual Digital Libraries of E-journals”. *Eighth International Conference on Language Resources and Evaluation (LREC), Istanbul, Turkey*. Nicoletta Calzolari et al. (ur.). Istanbul : European Language Resources Association (ELRA), 2012. 1710-1717.
- Stanković, Ranka, Cvetana Krstev, Ivan Obradović, Biljana Lazić, Aleksandra Trtovac. “Rule-based Automatic Multi-word Term Extraction and Lemmatization”. *10th International Conference on Language Resources and Evaluation, LREC 2016, Portorož, Slovenia*. Calzolari, N., et al. (ur.). Paris :The European Language Resources Association (ELRA), 2016. 507-514.

- Томашевић, Александра, Биљана Лазић, Далибор Воркапић, Михаило Шкорић, Љиљана Колоња. „Употреба веб платформе Омека за дигиталне библиотеке из домена рударства“. *Инфотека* Год. 17, бр. 2 (2017): 27-51.
- Томашевић, Александра. *Развој модела за управљање рударском пројектном документацијом*. Београд: [А. Томашевић], 2018.
- Tomašević, Aleksandra, Ranka Stanković, Miloš Utvić, Ivan Obradović, Vožo Kolonja. „Managing mining project documentation using human language technology“. *The Electronic Library* Vol. 36 Issue: 6 (2018): 993-1009.
- Ђирковић, Сњежана. „Сива литература – камелеон информационих ресурса“. *Инфотека* Год. 18, бр. 1 (2018): 78-87.

Biljana Lazić
Aleksandra Tomašević
Mihailo Škorić

DIGITAL LIBRARIES IN MINING AND GEOLOGY WITH SPECIAL REFERENCE TO THE GRAY LITERATURE

Summary

Having in mind the need for information retrieval in different types of documentation which are produced in domains of mining and geology, we started development of a digital library at the Faculty of Mining and Geology University of Belgrade. Very significant part of this documentation consists on grey literature which is based on multivolume documentation generated during the multidisciplinary projects. There is recognized significance of real-time and appropriate access to the project data. It is common that projects are stopped due to information loss in long-time period and unstable environment. The first challenge is to connect different multivolume parts in one consistent unit that is easily accessible and searchable. One more challenge is the nature of the project documentation which is the mixture of textual, numerical, graphical and cartographical contents. In order to solve the above-mentioned problems we used a platform for digital objects representation. Special attention will be given to relational dictionaries which are designed to define document relations. We will also present some language resources for Serbian language which are used to improve information retrieval.

Keywords: digital libraries, grey literature, Omeka, language resources, dictionaries.