

Просветитељски дух у 21. веку: хипертекстуално оплемењени материјал Сабраних дела Доситеја Обрадовића

УДК 004.738.5:027.7(497.11)

УДК 004.6:821.163.410bradović D.

САЖЕТАК: Универзитетска библиотека „Светозар Марковић“ годинама остварује успешну сарадњу са Задужбином „Доситеј Обрадовић“ у Београду са идејом да се направи претражива дигитална колекција дела Доситеја Обрадовића и дела о Доситеју Обрадовићу како би различите циљне групе – ученици основних и средњих школа, студенти, професори и истраживачи, на једном месту, брзо и једноставно, пронашли неопходну литературу о животу, раду и делима Доситеја Обрадовића. Сабрана дела Доситеја Обрадовића, у издању Задужбине 2007/2008. године, обрађена су у складу са савременим технологијама за обраду текста и креирана је дигитална колекција доступна на порталу дигиталне библиотеке Универзитетске библиотеке.

КЉУЧНЕ РЕЧИ: Доситеј Обрадовић, дигиталне колекције, анотација садржаја, XML, TEI, метаподаци, именовани ентитети

РАД ПРИМЉЕН: 3. август 2020.

РАД ПРИХВАЋЕН: 27. октобар 2020.

Јелена Андоновски
andonovski@unilib.rs

Ивана Гавриливић
gavrilovic@ubsm.rs
Универзитетска библиотека
„Светозар Марковић“
Београд, Србија

1. Увод

Универзитетска библиотека „Светозар Марковић“ годинама остварује успешну сарадњу са Задужбином „Доситеј Обрадовић“¹

¹ Задужбина „Доситеј Обрадовић“

у Београду са идејом да се направи претражива дигитална колекција дела Доситеја Обрадовића и дела о Доситеју Обрадовићу како би различите циљне групе – ученици основних и средњих школа, студенти, професори и истраживачи, на једном месту, брзо и једноставно, пронашли неопходну литературу о животу, раду и делима Доситеја Обрадовића. Сарадња је започета 2018. године пројектом „Претражива Сабрана дела Доситеја Обрадовића: креирање система за приказ и обједињену претрагу дигитализованих материјала Задужбине Доситеј Обрадовић“, а затим настављена и 2019. кроз пројекат „Сабрана дела Доситеја Обрадовића за 21. век: мултимедијално оплемењени материјали Задужбине Доситеј Обрадовић“. Пројекте је финансирало Министарство културе и информисања Републике Србије, а за њихову реализацију коришћено је научно издање „Сабрана дела Доситеја Обрадовића у шест књига“. Доситејева дела била су популарна још за пишчева живота и доживела су многобројна издања (1870, 1902, 1911, 1932, 1953). Најпознатија и најпотпунија су Сабрана дела у три тома, које је 1961. године издала београдска „Просвета“ у едицији „Српски писци“ (приређивачи – Ђуро Гавела, Јелена Шаулић, Боривоје Маринковић) (Обрадовић, 1961). Научно издање „Сабрана дела Доситеја Обрадовића у шест књига“ припремљено је поводом обележавања 200 година од Доситејевог доласка у Србију (1807/2007). Влада Републике Србије је тим поводом годину 2007. прогласила за „Годину Доситеја“, а Задужбина „Доситеј Обрадовић“ покренула је пројекат објављивања дела Доситеја Обрадовића. Као резултат припремљено је научно издање у шест томова приређивача проф. др Мирјане Д. Стефановић, др Миодрага Матицког и проф. др Душана Иванића. На 1.643 странице, уз напомене и коментаре, речнике мање познатих речи, регистре имена и географских појмова, регистре скраћеница и библиографских јединица, илустрације и литературу о Доситеју Обрадовићу објављени су сви текстови за које је несумњиво утврђено да су Доситејеви, затим радови који се приписују овом писцу, чак и кад то није могуће са потпуном сигурношћу утврдити, као и досад позната целокупна његова преписка (Обрадовић, 2007, 5-6). „Сабрана дела Доситеја Обрадовића“ садрже следећих шест књига:

1. Прва књига – *Писмо Харалампцију и Живот и прикљученија* (прир. Мирјана Д. Стефановић);
2. Друга књига – *Басне и две приповетке, Истина и прелест, Пут у један дан* (прир. Миодраг Матицки);

3. Трећа књига – *Совјети здраваго разума, Слово поучително господина Георгија Јоакима Цоликофера, Етика* (прир. Душан Иванић);
4. Четврта књига – *Собраније разних наравоучителних вештеј, Мезимац* (прир. Д. Иванић);
5. Пета књига – *Списи из Далмације, Проповеди и беседе, Записи и белешке, Фрагменти из Србије* (прир. М. Матицки);
6. Шеста књига – *Песме, Писма, Документи* (прир. М. Д. Стефановић).

Задужбина „Доситеј Обрадовић“ је Сабрана дела у потпуности уступила Универзитетској библиотеци за реализацију поменутих пројеката. Резултат рада на овим пројектима јесте портал „Сабрана дела Доситеја Обрадовића кроз призму дигиталних технологија“.²

2. Обрада текстова

Да би једна оваква колекција постала доступна ширем кругу корисника, односно да би се од ње направила дигитална збирка претражива по различитим параметрима, неопходно је да материјал буде у одговарајућем електронском формату, обрађен применом савремених технологија за обраду текста. Ако материјал није у електронском формату, што је јако често, ради се дигитализација која подразумева скенирање материјала (Ikonov and Dobrova, 2008, 2), а затим се дигитална слика оптичким препознавањем карактера (*Optical Character Recognition* – OCR) конвертује у електронски читљив текст (Трговац, 2016, 105-107). Међутим, ако су одабрани текстови већ у електронском формату, они се као такви користе у даљој обради.³

Материјали које је Универзитетска библиотека добила били су већ у електронском формату, тачније у формату .docx, па процес скенирања и оптичког препознавања карактера није био потребан. Тиме је добијена значајна уштеда у времену. Обрада текста остварена је кроз неколико етапа, а као њихов резултат, материјал је припремљен у одговарајућем XML (Extensible Markup Language) формату⁴ (W3C, 2020a) у складу са

² Сабрана дела Доситеја Обрадовића кроз призму дигиталних технологија

³ Задужбина „Доситеј Обрадовић“ је Универзитетској библиотеци „Светозар Марковић“ уступила коначне верзије текстова Сабраних дела припремљене за штампу.

⁴ Extensible Markup Language

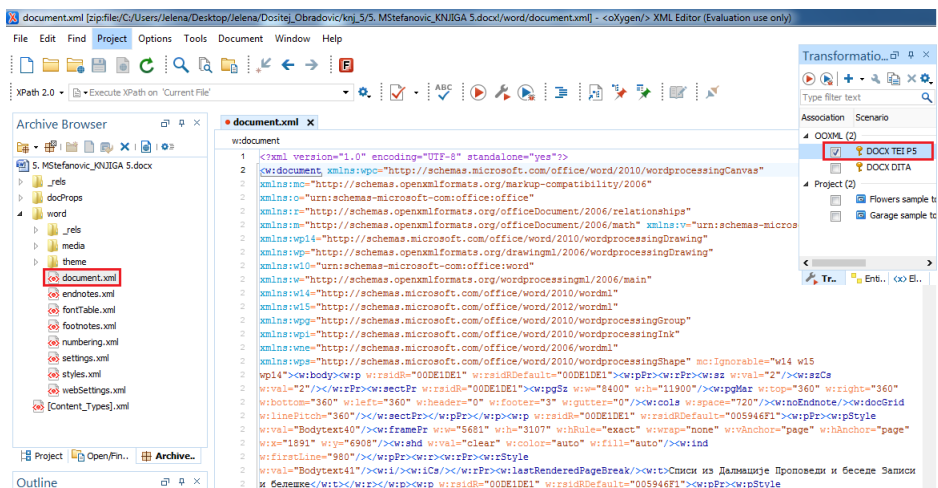
прописаним смерницама Иницијативе за кодирање текста (*Text Encoding Initiative* – TEI)⁵ којима се дефинишу општа правила за кодирање текста, као и део уграђеног заглавља који садржи метаподатке о делу, TEI заглавље (TEI Header) (TEI Consortium, 2020).

Поступак припреме XML докумената подразумевао је анотацију формалних и структурних карактеристика материјала и индексирање делова садржаја текста како би се касније омогућило њихово претраживање по различитим критеријумима, као и повезивање са релевантним ресурсима на интернету. Креирање XML докумената урађено је аутоматски. Материјал припремљен у .docx формату је већ компримована датотека за коју се подразумева да се чува у XML формату и распакује када се отвори у одговарајућем програму за уређивање XML датотека. У овом случају за обраду материјала коришћен је Oxygen XML Editor. Када се датотека .docx отвори у Oxygen XML Editor-у, она се распакује и за даљу обраду користи се генерички креирана датотека document.xml која се, затим, опцијом DOCX TEI P5 аутоматски трансформише у TEI/XML датотеку (слика 1). Овако произведене TEI/XML датотеке су даље уређиване како би се добила одговарајућа структура, што је објашњено у наредним одељцима.

3. Формална анотација

Анотација формалних карактеристика одабраног материјала подразумева доделу метаподатака којим се описују, објашњавају, лоцирају или на други начин чине лакшим проналажење, коришћење или управљање неким извором информација (Hodž, 2004, 157) са јасно дефинисаном семантиком и структуром (Berners-Lee, 1997). Метаподаци за дела Доситеја Обрадовића наведени су у заглављу произведене TEI/XML датотеке (teiHeader), и они су у суштини библиографски и административни (Трговац, 2016, 43-45). Библиографски метаподаци могу се поделити на оне који описују електронску верзију дела (наслов/title, аутор/author, одговорност/respStmnt и датум/date) и оне који описују његову аналогну верзију (аутор/author, наслов/title, место издавања/pubPlace, издавач/publisher, датум/date и податак о едицији/series). Поред библиографских метаподатака, у заглављу је дефинисана и лиценца (административни метаподатак) којом се регулишу права приступа и коришћења доступног садржаја (етикета

⁵ Text Encoding Initiative



Слика 1. Датотека document.xml за дело „Списи из Далмације, Проповеди и беседе, Записи и белешке, Фрагменти из Србије“ и опција DOCX TEI P5

ref у оквиру етикете publisher у делу publicationStmnt). На слици 2 дат је пример заглавља за дело „Писмо Харалампију и Живот и прикљученија“, док је на слици 3 приказан изглед метаподатака за исто дело на порталу „Сабрана дела Доситеја Обрадовића“.

4. Структурна анотација


Поред анотације формалних карактеристика, тј. израде метаподатака, свако дело је и структурно анотирано. Структурна анотација подразумева обележавање логичких структура текста (поглавља, наслова, пасуса, реченица, речи) одговарајућим XML етикетама. Она може бити урађена аутоматски или ручно. Структурна анотација Сабраних дела урађена је највећим делом аутоматски. Након генерисања TEI/XML датотеке аутоматски су биле означене следеће логичке структуре текста: тело текста, поглавља, наслови поглавља и пасуси. Тело текста и пасуси су адекватно анотирани одговарајућим XML етикетама, за тело текста <body> и за пасусе <p>. Поголавља и њихови наслови били су, такође, анотирани али XML етикетама

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Писмо Харалампију, Живот и прикљученија: електронско издање</title>
      <author ref="viaf:76370256">Доситеј Обрадовић</author>
      <respStmt>
        <resp>Израда електронског издања</resp>
        <name>Универзитетска библиотека "Светозар Марковић"</name>
      </respStmt>
    </titleStmt>
    <editionStmt>
      <edition><date>2019-10-15</date></edition>
    </editionStmt>
    <publicationStmt>
      <publisher>
        <ref target="http://creativecommons.org/licenses/by-nc-sa/4.0/">Објављено под лиценцом Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International</ref>
      </publisher>
    </publicationStmt>
    <sourceDesc>
      <bibl type="copyText">
        <author>Доситеј Обрадовић</author>
        <title ref="COBISS-SR.ID:144066572">Писмо Харалампију, Живот и прикљученија: електронско издање</title>
        <pubPlace>Београд</pubPlace>
        <publisher>Задужбина Доситеј Обрадовић</publisher>
        <date>2007</date>
        <series n="1">Сабрана дела Доситеја Обрадовића</series>
        <idno type="wiki">https://sr.wikipedia.org/wiki/Доситеј_Обрадовић</idno>
        <idno type="viaf">https://viaf.org/viaf/76370256/#Обрадовић_Доситеј_1739-1811</idno>
      </bibl>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

Слика 2. Заглавље за дело „Писмо Харалампију и Живот и прикљученија“ у TEI/XML формату

МЕТАПОДАЦИ

☰ 1 / 6 - Писмо Харалампију, Живот и прикљученија: електронско издање / Сабрана дела Доситеја Обрадовића

✎ Доситеј Обрадовић 

✎ Израда електронског издања: Универзитетска библиотека "Светозар Марковић"

☰ Задужбина Доситеј Обрадовић - 2007 - Београд

✎ Објављено под лиценцом Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International

Слика 3. Метаподаци за дело „Писмо Харалампију и Живот и прикљученија“ на порталу „Сабрана дела Доситеја Обрадовића“

за пасусе. Према TEI Schema-и поглавља и наслови треба да буду анотирани одговарајућим етикетама, `<div>` за поглавља и `<head>` за наслове поглавља, те је било неопходно урадити корекције. Корекције су урађене аутоматским путем, применом одговарајућих регуларних израза.

Пример регуларног израза:

```
<p rend=Heading #[0-9]+ ">  
<anchor xml:id="bookmark[0-9]+"/>([A-Ша-ш]+) ([a-шA-Ш]+)]+  
</p>"
```

Replace:

```
<head style=text-align:center;" rend="italic">&1</head>"
```

Уз етикете за поглавља, наслове и пасусе додавани су и неретко атрибути који указују на додатне карактеристике садржаја који се анотира како би електронска верзија текста остала верна аналогном оригиналу. На пример:

Означавање поглавља

Пример 1:

```
<div type="chapter">  
<head style="text-align:center;">1.</head>  
...</div>
```

Пример 2:

```
<div type="chapter">  
<head style="text-align:center;" rend="italic">Предисловије о  
баснах</head>  
...</div>
```

Пример 3:

```
<div type="chapter" n="8[9]">  
<head style="text-align:center;" rend="italic">8[9]</head>  
<head style="text-align:center;" rend="italic">Лав и зец</head>  
...</div>
```

У претходно наведеним примерима уз етикету `<div>` увек је додаван атрибут `type` који има вредност `"chapter"` како би се нагласило да је реч о поглављу (Примери 1, 2 и 3). Када је реч о насловима, они су у самом тексту постављени централно што је у XML запису анотирано вредношћу `"text-align:center;"` атрибута `style` (Пример 1 и 3). Поред тога, већина наслова је и искошено означена, те је поред вредности која анотира централну позицију наслова додавана и вредност `"italic"` атрибута `rend` (Пример 2 и 3).

Означавање пасуса

Пример 4:

<p>[Писмо <rs type="person">Харалампију</rs>]<note>Овај рад Доситеј Обрадовића нема свој насловни лист. Због тога приређивач, према садржају и према адресату, насловљава текст и тај наслов ставља у угласте заграде. Текст је био штампан курзивом, као што је уобичајено било за предговоре. Овде су пагиновани листови, а не странице, и то звездицама, што је, такође, био обичај обележавања за предговоре. У науци постоје различита мишљења о том да ли је овај рад Доситејев посебна библиографска јединица (Стефановић, 2007. И ту наведену литературу; в. и Библиографија, 1990)... </p>

Пример 5:

<p style="text-align:center;">ЉУБИМАЈА ЈУНОСТЕ,</p>

Пример 6:

<p rend="italic">Здравствуй! Христос воскрес... </p>

Када је реч о пасусима, етикета <p> је врло често била без атрибута (Пример 4) осим у случају када је требало истаћи централну позицију пасуса (атрибут *style="text-align:center;"*) (Пример 5) или ако је требало истаћи да је пасус искошено означен (атрибут *rend="italic"*) (Пример 6). Поред поглавља и пасуса у сваком тексту анотирана је и пагинација за шта је коришћена етикета <pb> са атрибутом *p* којим се аотира број стране. На пример, <pb n="15"/>. Пагинација је анотирана како би се направио веран упоредни преглед оригиналног материјала и материјала припремљеног у TEI/XML формату. Као и у примерима за поглавља и њихове наслове, етикетирање пагинације урађено је аутоматски применом регуларних израза. Поред анотације логичких структура у сваком тексту анотирани су и елементи текста на страном језику, али и елементи текста који су на одређени начин истакнути: подебљани, искошени, подвучени и слично. За елементе текста на страном језику коришћена је етикета <foreign> са атрибутом *xml:lang* чија је вредност одговарајући међународни двословни код за језик. На пример, <foreign xml:lang="it">„Farina de diavolo va presto in semola</foreign>. Истакнути елементи текста као што су подебљане, искошене или подвучене речи или синтагме анотирани су етикетом <hi> са атрибутом *rend* чија је могућа вредност *bold*, *italic*, *underline* итд. у зависности од тога како је текст истакнут. На пример, <hi rend="bold">Кажу да се чуде</hi>. Важно је напоменути да су истакнути елементи текста већ након процеса аутоматског генерисања TEI/XML датотеке били анотирани.

5. Анотација садржаја

5.1 Анотација ентитета описним напоменама

Свако дело у едицији „Сабрана дела Доситеја Обрадовића“ поред основног текста садржи и напомене које представљају значајан део за разумевање и сазнавање додатних информација о личностима, географским појмовима, временским периодима, насловима одређених дела, значајним синтагмама, метафорама и стиховима који се у делима појављују. Кроз напомене истраживачи ће сазнати зашто су Доситеју важне одређене личности које је познавао, колико дуго их је познавао, на који су начин утицале на Доситеја, зашто су му важна одређена места у којима је боравио и како је све то обележило његов живот и стваралаштво. Такође, захваљујући напоменама истраживачи могу лакше разумети значење појединих синтагми и метафора којима се Доситеј током писања служио, као и на шта указују одређени периоди који се у делима спомињу. Напомене су сложене на самом крају сваког текста и пописане хронолошки, оним редом како се ентитети који се њима описују у тексту појављују. Списак садржи оригинални облик ентитета из текста уз који стоји објашњење, а затим и број странице у тексту на којој се ентитет налази како би се истраживачи лакше снашли. Анотација ентитета описним напоменама подразумевала је додавање објашњења, односно напомена ентитетима у тексту како би истраживачи одмах на одређеном месту могли да погледају објашњење без померања кроз текст, што им умногоме олакшава и убрзава рад. Прецизније речено, након аотације ентитета описним напоменама, оне се могу погледати на месту где се ентитет налази у тексту без потребе да се објашњење тражи у списку напомена на крају целог дела. Поступак аотације ентитета описном напоменама урађен је ручно, а за то је коришћена етикета `<note>`. На слици 4 приказана је описна напомена за ентитет „Здравствуй! Христос воскрес“ у делу „Писмо Харалампију и Живот и прикуљченија“, док су на сликама 5 и 6 приказани примери аотације у TEI/XML датотеци и приказ на порталу „Сабрана дела Доситеја Обрадовића“.

5.2 Анотација именованих ентитета

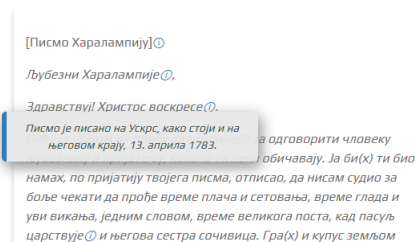
Поред аотирања ентитета описним напоменама, у тексту су аотирани и именовани ентитети – лична имена и геополитички

Христос воскресе (стр. 15) - Писмо је писано на Ускрс, како стоји и на његовом крају, 13. априла 1783.

Слика 4. Напомена за ентитет „Здравствуй! Христос воскресе“ на крају текста „Писмо Харалампију и Живот и прикљученија“

```
<p zend="italic">
Здравствуй! Христос
воскресе</p><note>Писмо је писано на Ускрс, како стоји и на његовом
крају, 13. априла 1783. </note.</p>
```

Слика 5. Анотација ентитета „Здравствуй! Христос воскресе“ напоменом у TEI/XML документу „Писмо Харалампију и Живот и прикљученија“



Слика 6. Приказ напомене за ентитет „Здравствуй! Христос воскресе“ на порталу „Сабрана дела Доситеја Обрадовића“

појмови. Препознавање именованих ентитета (*Named Entity Recognition* – *NER*) једна је од технологија за проналажење информација у самим документима (Chinchor and Robinson, 1997) и водећа тема у области обраде природних језика више од петнаест година (Vitas et al., 2007, 115-116) (Krstev et al., 2014). Технологија подразумева лоцирање специфичних појмова у тексту и њихово класификовање у одређене категорије (Андоновски, 2019, 124). Препознавање и класификација именованих ентитета изузетно су важни за многе научне дисциплине и области. Са једне стране, то је први од пет подзадатака екстракције информација која представља специјализовану област проналажења информација, док је са друге стране, у оквиру рачунарске лингвистике, то важно за обраду природних језика и аутоматско превођење (Utvić, 2008, 2). Поступак анотације именованих ентитета у „Сабраним делима“ урађен је ручно. За анотацију је коришћена етикета <rs>, а како би се прецизније одредило да ли је реч о личном имену или геополитичком појму, уз етикету <rs> додаван је атрибут *type* који може имати вредност *person* за анотацију личног имена или *place* за анотацију геополитичког појма. На слици 7 приказан је пример означавања именованих ентитета у четвртм делу из едиције „Собраније разних наравоучителних вештеј, Мезимац“. У примеру су означени именовани ентитети „Сократ“, „Софрониска“, „Фенарете“, „Атини“ и „Христа“⁶. Од тога ентитет „Атини“ спада у геополитичке појмове, док су остали ентитети лична имена. Означавањем именованих ентитета створени су предуслови за прецизније проналажење ових информација у тексту, али и за њихово увезивање са сродним ентитетима у релевантним ресурсима на мрежи као што су нормативне датотеке личних имена, нормативне датотеке географских имена, различите врсте онтологија и слично, а које су данас део шире семантичке мреже „Отворени повезани подаци“ (*Linked Open Data*– *LOD*) (Berners-Lee, 2006; Radulovic et al., 2015; W3C, 2020b; Blaney, 2020).

⁶ Синтагма „469 година пре Христа“, која се може видети на слици 7, временски је израз који се према дефиницији сматра именованим ентитетом. Међутим, како су у оквиру пројеката, који су у раду описани, од именованих ентитета означавани само лична имена и геополитички појмови, ова синтагма није аотирана као временски израз већ је само аотирано име „Христ“. Временски изрази биће аотирани у неком од следећих фаза када ће и ово бити исправљено.

```
<p><rs type="person">Сократ</rs>, син једног каменосеца именован  
<rs type="person">Софрониска</rs> и једне жене звалијем бабице, а  
имени <rs type="person">Фемарете</rs>, родио се је у  
<rs type="place">Атени</rs> 469 година пре  
<rs type="person">Христа</rs>. Испочетка научи занат  
својега оца и историја чини воспоминаније о трима  
ста|туама, које предававаху три грације и које су предивно  
41 соделате биле. Види се од прилодобивања која је овај  
преспавни муж у своји разговорни употребавао, да се он  
нивало није смирдио заната свога родитеља и родитељице.  
„Зачудо“, говораху он, „да један каменосецац</p>
```

Слика 7. Анотација именованих етитета у делу „Собраније разних наравоучителних вештеј, Мазимац“, TEI/XML датотека

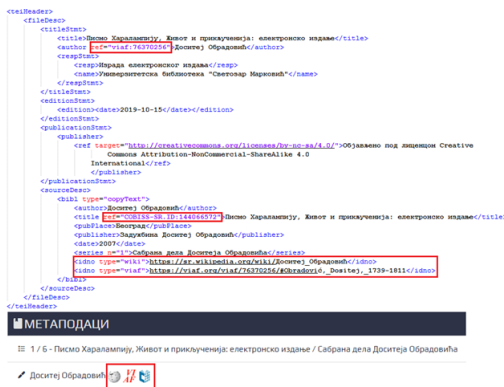
6. Увезивање метаподатака

Библиографски метаподаци наведени у заглављу сваког документа, име аутора и наслов, увезани су са записима у релевантним ресурсима на интернету. Име аутора, у овом случају Доситеј Обрадовић, повезано је са записом о Доситеју Обрадовићу у Међународној виртуелној нормативној датотеци (*Virtual International Authority File – VIAF*),⁷ која је истовремено и део мреже „Отворени повезани подаци“, и са страницом о Доситеју Обрадовићу на Википедији на српском језику. Поред повезивања имена аутора, и наслов дела повезан је са одговарајућим записом у електронском каталогу COBISS.SR.⁸ У поменутих базама сваки запис има свој идентификациони број који представља његов јединствени ID и који је искоришћен за увезивање. Идентификациони бројеви записа за Доситеја Обрадовића у бази VIAF и наслова у електронском каталогу COBISS.SR наведени су у заглављу TEI/XML документа као вредност атрибута *ref* који стоји уз етикете *<author>* и *<title>*, док је за запис на Википедији наведен линк у оквиру етикете *<idno>* са атрибутом *type* чија је вредност *“wiki”* (слика 8). На овај начин корисници могу у датом тренутку добити детаљније податке о аутору и наслову који проучавају у различитим ресурсима од којих су VIAF и COBISS.SR изузетно важни за библиотечку заједницу. База података VIAF је међународна нормативна датотека имена аутора, организација и наслова дела у коју се прикупљају национално усвојени термини из националних библиографских центара, док је COBISS.SR електронски каталог Републике Србије. Ове две базе производ су рада у библиотечко-информационој делатности те су као такве од великог значаја за ову истраживачку заједницу. Нормативне датотеке

⁷ [Virtual International Authority File](#)

⁸ [COBISS.SR](#)

и електронски каталози представљају значајније библиотечке ресурсе. Узевивањем поменутих библиографских метаподатака са овим базама указано је на значај и утицај тих ресурса на нове технолошке могућности које носи семантички веб, док је са друге стране важно да библиотеке увиде предност примене овакве технологије у области претраживања и проналажења информација. Богатство ресурса који у библиотекама већ постоје, али и нових који се стварају, одлично су полазиште за примену нових технолошких могућности које нуди семантички веб (Андоновски, 2019, 237-238). Са друге стране, базе знања, као што је на пример Википедија, пружају битне информације о различитим ентитетима и њихове странице могу постати основа за неко даље истраживање. Стога су одличан ресурс за увезивање са одабраним ентитетима.



Слика 8. Везе ка базама VIAF, COBISS.SR и Википедија

7. Закључак

Након вишегодишње успешне сарадње Задужбина „Доситеј Обрадовић“ и Универзитетска библиотека „Светозар Марковић“ наставиле су сарадњу и 2020. године пројектом *Критичка мисао о Доситеју Обрадовићу (анотација библиографских метаподатака и њихово увезивање са значајним библиотечки-информационим*

системима у мрежи повезаних отворених података). Циљ пројекта је да се настави развој портала „Сабрана дела Доситеја Обрадовића“ постављањем новог текстуалног корпуса који обухвата библиографску грађу о Доситеју Обрадовићу и његовим делима, али и да се унапреде технички услови за примену библиотечко-информационих технологија у погледу детаљне анотације библиографских метаподатака и њиховог увезивања са важним библиотечко-информационим ресурсима на интернету. На овај начин започиње стварање једног већег текстуалног корпуса о животу и раду Доситеја Обрадовића. У пројекту је посебно наглашена анотација библиографских метаподатака за сваку одабрану библиографску јединицу, а у складу са препорукама стандарда TEI, и њихово увезивање са релевантним библиотечко-информационим ресурсима као што су међународне нормативне датотеке и базе знања које су део мреже „Отворени повезани подаци“. Будући да је грађа о Доситеју Обрадовићу јако богата, како на српском језику тако и на другим језицима, за овај пројекат биће одабран само један део материјала који ће до краја године постати део постојећег портала. Такође, нови корпус материјала значајно ће допринети и садржају потколекције „Доситеј Обрадовић“ у оквиру колекције „Критика (18. и 19. век)“, на порталу „Српска књижевна критика“⁹ који се такође развија и одржава у Универзитетској библиотеци „Светозар Марковић“.

Литература

- Berners-Lee, Tim. “Metadata Architecture”, 1997, preuzeto 31. 7. 2020, <https://www.w3.org/DesignIssues/Metadata.html>
- Berners-Lee, Tim. “Linked Data”, 2006, preuzeto 31. 7. 2020, <http://www.w3.org/DesignIssues/LinkedData.html>
- Blaney, Jonathan. “Introduction to the principles of Linked Data”, 2020, preuzeto 8. 8. 2020, <https://programminghistorian.org/en/lessons/intro-to-linked-data>
- Chinchor, Nancy and Patricia Robinson. “MUC-7 named entity task definition”. У *Proceedings of the 7th Conference on Message Understanding*, Vol. 29, 1–21. 1997
- TEI Consortium. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, 2020, preuzeto 31. 7. 2020, <https://tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>

⁹ Српска књижевна критика. Потколекција „Доситеј Обрадовић“

- Hodž, Gejl. “Metapodaci na lak način”. *Glasnik Narodne biblioteke Srbije* br. 1 (2004): 157–180, preuzeto 31. 7. 2020, http://www.nb.rs/view_file.php?file_id=860
- Ikonomov, Nikola and Milena Dobрева. “The making of... digital book”. *NCD Review* no. 13 (2008): 1–8, preuzeto 8. 8. 2020, <http://elib.mi.sanu.ac.rs/files/journals/ncd/13/ncd13001.pdf>
- Vitas, Duško, Cvetana Krstev and Denis Maurel. “A note on the Semantic and Morphological Properties of Proper Names in the Prolex Project”. *Linguisticae Investigationes, Special issue on Named Entities: Recognition, Classification and Use* Vol. 30, no. 1 (2007): 115–134
- Krstev, Cvetana, Ivan Obradović, Miloš Utvić and Duško Vitas. “A system for named entity recognition based on local grammars”. *Journal for Logic Computation* Vol. 24, no. 2 (2014): 473–489, preuzeto 19. 10. 2020, <https://academic.oup.com/logcom/article-abstract/24/2/473/1008564?redirectedFrom=fulltext>
- Radulovic, Filip, María Poveda-Villalón, Daniel Vila-Suero, Víctor Rodríguez Doncel, Raúl García-Castro et al. “Guidelines for Linked Data generation and publication: An example inbuilding energy consumption”. *Automation in Construction* no. 57 (2015): 178–187
- Utvić, Miloš. “Konačni automati u regularnoj imenskoj derivaciji”. Mag. teza, Univerzitet u Beogradu, Matematički fakultet, 2008
- W3C. “Extensible Markup Language (XML)”, 2020a, preuzeto 31. 7. 2020, <https://www.w3.org/XML/>
- W3C. “Linked Data”, 2020b, preuzeto 31. 7. 2020, <https://www.w3.org/standards/semanticweb/data>
- Андоновски, Јелена. “Мрежа отворених података и језички ресурси у процеси+у изградње српско-немачког литерарног корпуса”. Докт. дисертација, Универзитет у Београду, Филолошки факултет, 2019
- Обрадовић, Доситеј. *Сабрана дела : 1811-1961. 1-3*. Београд: Просвета, 1961
- Обрадовић, Доситеј. *Писмо Харалампчију. Живот и прикљученија*. Београд: Задужбина „Доситеј Обрадовић“, 2007. Доступно на: URL <https://dositej.unilib.rs/>
- Трговац, Александра. “Дескриптори метаподатака и дескриптори садржаја у проналажењу информација у дигиталним библиотекама”. Докт. дисертација, Универзитет у Београду, Филолошки факултет, 2016