

Увод у Дигиталну хуманистику: радионице за имплементацију ‘удаљеног читања’ у истраживачкој пракси, Тршић – 4–8. 12. 2023.

Милица Рабреновић
milica.rabrenovic@isj.sanu.ac.rs
ORCID: 0000-0002-4441-4884
*Институт за српски језик
САНУ
Београд, Србија*

РАД ПРИМЉЕН: 17. април 2024.
РАД ПРИХВАЋЕН: 29. мај 2024.

1. Увод

Универзитетска библиотека „Светозар Марковић” и Друштво за језичке ресурсе и технологије – JePTex организовали су први семинар „Увод у Дигиталну хуманистику: радионице за имплементацију ‘удаљеног читања’ у истраживачкој пракси”, који је реализован у Тршићу од 4. 12. до 8. 12. 2023. године. Семинар је спроведен кроз сарадњу његових аутора и Министарства науке, технолошког развоја и иновација Републике Србије. Полазници семинара били су студенти основних, мастерских и докторских студија. Циљ овог семинара био је да се студенти филолошких и хуманистичких наука упознају с методама дигиталне хуманистике, припреме и обраде текста за примену техника ‘удаљеног читања’. Семинар је обухватао део с предавањима и практичан део, односно радионице.

Првог дана семинара одржана су два предавања, потом је у поподневним сатима уследила примена теоријског дела. Прво предавање „Увод у дигиталну хуманистику” одржао је др Василије Милновић. На почетку излагања, др Милновић се осврнуо на резултате њихових досадашњих пројеката, упознао је полазнике с пројектом у оквиру COST Action CA16204 *Distant Reading for European Literary History* (2018–2022) и истакао важност дигиталне хуманистике. Говорио је о значају развијања језичких технологија, с посебним освртом на тзв. "мале језике" (нпр. Исланд, Естонија или Израел где се улажу значајна средства у развоју тог процеса) и њиховој потенцијалној примени у

лексикографији, настави српског језика и књижевности, превођењу и другим областима.

Предавање проф. др Цветане Крстев било је посвећено корпусу *SrpELTeC*. Проф. др Крстев упознала је полазнике с процесом настајања овог корпуса и критеријумима које је било неопходно испунити да би српски роман у датом периоду постао део корпуса (равномерна заступљеност и мушких и женских аутора; равномерно покривање изабраног периода (1840–1920); разноврсност по питању дужине; позната и призната дела, али и непозната; разноврсност аутора). Било је речи о изазовима с којима су се сусретали током формирања српске потколекције, као и о процесу дигитализације првих издања (или најстаријих доступних). Процес дигитализације чине следећи кораци: скенирање, оптичко препознавање карактера, корекција, основна анотација, уношење метаподатака, аутоматска напредна анотација.

Полазници семинара упознати су на који начин се врши корекција и обележавање текста, као и како се уносе метаподаци, те су први дан радионице добили текст на којем су радили корекцију пратећи упутства за обележавање, а током осталих дана уносили су метаподатке и радили су на анотацији језичких ентитета у том тексту.

Другог дана је др Александра Тртовац, у свом предавању „Универзитетска библиотека ‘Светозар Марковић’ и удаљено читање”, говорила о повезаности библиотекарства и дигиталне хуманистике, као и о њиховој улози у COST акцији. Друго предавање др Тртовац посветила је пројекту Транскрибус и дигитализацији путем мобилног телефона. Полазници су тог дана на радионици направили налоге на Транскрибусу,¹ а потом су упознати и са апликацијом *DocScan* помоћу које је могуће учитати рукописне радове и добити транскрипт текста. Веома корисно било је упознавање с напредним претраживањем у Кобису.² Поред изборног претраживања, фокус је био усмерен и ка командном претраживању. Једна од корисних информација била је и да се у одељку *UNPAYWALL* налазе научни чланци са отвореним приступом, као и да сва имена и презимена једног аутора (корисно уколико имамо псеудоним, а желимо да излистамо сва дела једног аутора или ако је презиме ауторке мењано услед брачних разлога) можемо наћи у делу *CONOR.SR*.

1. [readcoop](#), приступљено 28. 6. 2024. године

2. [COBISS.SR](#), приступљено 28. 6. 2024.

Трећег дана проф. др Крстев упознала је полазнике с појмом именовани ентитети и начином како се они обележавају у тексту. Предавање проф. др Ранке Станковић надовезало се на претходно, било је речи о обради дигиталних текстова. Учесници радионице су радили на аутоматској анотацији делова текста именованим ентитетима (места, особе, организације, професије...), дате су и одређене напомене о напреднијим врстама анотације, потом је уследила контрола и евалуација аутоматски евалуираних података.

Истог дана на радионици учесници су упознати с википодацима о српским романима, те су уносили основне податке о тексту који су добили првог дана користећи алате *OpenRefine* и *QuickStatements*. Учесници су упознати и с претраживањем података језиком *SPARQL*.

Четврти дан започет је предавањем проф. др. Душка Витаса, говорено је о процесу настајања кулинарског корпуса. Проф. др Ранка Станковић наставила је тему о корпусима доступним на NOSKE платформи друштва ЈерТех.³ Учесници су имали прилику да се упознају с основама *CQL* језика и како да дају упите над корпусом *srpEL-TeC*. Радионица је била посвећена командним претрагама, почевши од једноставнијих упита, па све до сложенијих граматичких образаца и упита који укључују етикете именованих ентитета.

Последњег дана семинара учесници су слушали предавање „Иницијатива за развој отворених NLP/NLU ресурса и алата за српски језик” Слободана Марковића. Током предавања истакнути су изазови употребе вештачке интелигенције (ВИ), као и могућности и проблеми у вези с обрадом српског језика коришћењем ВИ. Проблем постоји јер модели нису суштински прилагођавани, њихова примена није практична у пословним решењима и могућности експресије су сужене. Решење је у томе да буде што више текста за обучавање. Идеално би било да је што више тога јавно доступно, под пермисивном лиценцом и да покрива оба изговора. Циљеви иницијативе за отворене NLP ресурсе јесу бољи резултати претраге и разумевање текста.

На овом семинару стечена су свеобухватна знања о методама дигиталне хуманистике и значају ове научне области. Овај семинар омогућио је боље увиде за будућа истраживања, посебно када је у питању обележавање грађе, али и рад са текстуалним корпусима, у чему је и највећа примена наученог током овог семинара. Оно што треба истаћи као најважнији резултат ових предавања и радионица јесте упознавање с колегама и развијање потенцијалне сарадње на будућим пројектима

3. noske.jerteh.rs, приступљено 28. 6. 2024.

и истраживањима. То и јесте главни разлог зашто бих као учесница топло препоручила семинар „Увод у Дигиталну хуманистику: радионице за имплементацију ‘удаљеног читања’ у истраживачкој пракси”.

Захвалница

Овај рад финансирао је Министарство просвете, науке и технолошког развоја Републике Србије према Уговору број 451-03-136/2025-03/200174, који је склопљен са Институтом за српски језик САНУ.