

## Приказ београдске радионице *European Language Grid*

**РАД ПРИМЉБЕН:** 19. април 2022.  
**РАД ПРИХВАЋЕН:** 16. мај 2022.

Бојана Башарагин  
bojana.basaragin@ivi.ac.rs  
*Истраживачко-развојни  
институт за вештачку  
интелигенцију Србије*

11. марта 2022. године одржан је први у низу дисеминационих догађаја у оквиру ЕУ пројекта European Language Grid (ELG)<sup>1</sup>. У питању је била онлајн ELG радионица у организацији Филолошког факултета у Београду и Друштва за језичке ресурсе и технологије JePTeX. Овај догађај је уједно најавио и скори, успешан завршетак пројекта средином 2022. године.

ELG пројекат (2019-2022) настао је с идејом да приближи академску заједницу и индустрију у домену језичких технологија и оснажи позицију Европе у односу на друге континенте у овом погледу. Резултат пројекта је јединствена ELG платформа на којој су сабрани ресурси и алати за обраду европских језика, како би постали видљивији и доступни како академским заједницама, тако и пословним партнерима, невладиним организацијама и државном сектору. ELG као један од својих циљева наводи и помоћ у решавању питања дигиталне језичке угрожености пружањем подршке европским језицима који нису довољно подржани кроз језичке технологије. Према подацима из јануара 2022. године, на платформи постоји преко 10.000 сервиса, алата, скупова података, ресурса и језичких модела за 87 језика, а тај број се увећава из месеца у месец. Након завршетка пројекта планирано је да платформа настави да живи, омогућавајући приступ како некомерцијалним, тако и комерцијалним језичким технологијама.

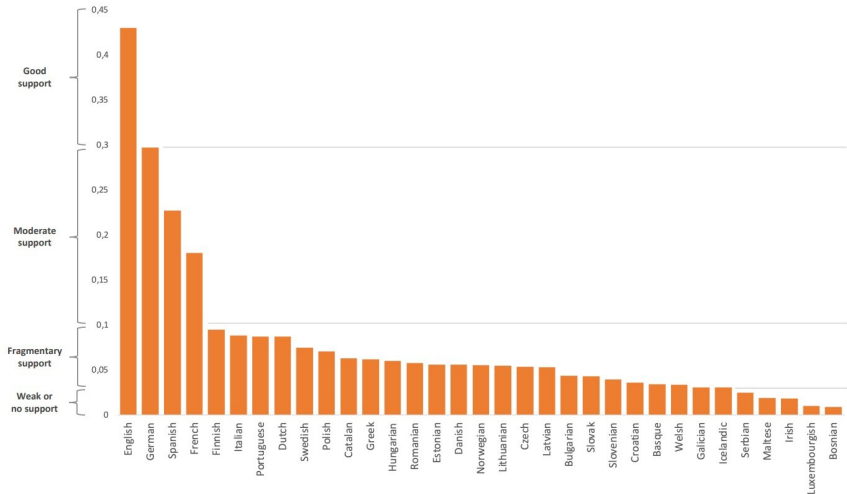
Радионица је трајала три сата и одржана је онлајн, преко платформе Zoom, а снимак радионице се може наћи на YouTube каналу Друштва JePTeX<sup>2</sup>. Званични језик радионице био је енглески. Након поздравне речи коју је упутила организаторка, проф. Цветана Крстев, одржана је уводна презентација координатора пројекта проф. Георга Рема. Проф. Рем је направио детаљан преглед пројекта ELG, његових резултата,

---

1. Пројекат European Language Grid (ELG).

2. YouTube канал друштва JePTeX.

као и перспектива за даљи раст и развој платформе. Представио је и проблем језичке неједнакости у заступљености језичких технологија за већину европских језика, изузимајући шпански, немачки и француски (Слика 1). Посебно се осврнуо на прилике српског језика, који према извештају пројекта European Language Equality (ELE) за српски језик<sup>3</sup> и даље спада у језике са слабом покривеношћу језичким технологијама. На питање шта види као следећи корак у развоју ресурса и алата за српски језик, навео је израду опсежног језичког модела који би се затим могао применити на различите задатке обраде и генерисања српског језика.



Слика 1. Покривеност европских језика језичким технологијама

Затим је уследила практична и детаљна демонстрација употребе платформе, коју је приказала Пени Лабропулу, виши научни сарадник грчког Института за обраду језика и говора. Присуство страних предавача и гостију дало је повода да се за потребе радионице искористи резултат скоро завршеног европског пројекта ELITR<sup>4</sup>. У оквиру овог

3. ELE извештај за српски језик.

4. Пројекат European Live Translator (ELITR).

пројекта развијен је аутоматски систем за превођење и титловање састанака и конференцијских представљања, као и аутоматског хватања белешки. Аутоматски запис свих говора на радионици и њихов превод на српски (или неки други европски) језик у реалном времену могли су се пратити у одвојеном прозору.

Након уводних презентација уследили су примери из домена језичких технологија за српски језик. У духу пројекта, радионица је замишљена тако да обједини примере успешне примене језичких технологија у индустрији и најновија достигнућа у научном домену. Тако су прве две презентације дошле из две велике српске фирме – телекомуникационог оператера Yettel и ИТ групе ComTrade. Ђорђе Хирш, вођа тима за управљање подацима у Yettel Serbia, представио је методу ненадгледане анализе сентимента коју у фирми Yettel примењују на базу гласовних жалби упућених корисничком сервису како би побољшали своје услуге. Душан Јовичић, програмер, приказао је регионално искуство фирме ComTrade у прављењу разговорних агената (четботова).

Примери из индустрије били су праћени представљањем ресурса и алата за српски језик развијених у сарадњи Београдског универзитета (БУ) и Друштва ЈеРТех. Најпре је проф. Ранка Станковић говорила о корпусу савременог српског језика СрпКор, који је крајем 2021. године добио нову верзију – СрпКор2021. У односу на претходну верзију, овај корпус је обогаћен новим текстовима из различитих домена, међу којима се истиче корпус литерарних текстова из 19. и с почетка 20. века ELTeC, настао у оквиру COST акције Distant Reading<sup>5</sup>. Овај корпус, састављен од преко 100 књижевних дела обележених врстама речи, уједно је био тема и последњег представљања на радионици. Корпус ELTeC су представиле проф. Цветана Крстев и Милица Иконић Нешић, као потпуно слободан ресурс који се може користити за претраге, анализу и тренирање модела за српски језик. Овај корпус се тренутно може преузети са GitHubа пројекта. СрпКор2021, па тако и ELTeC корпус, могу се претраживати преко система NoSketch, доступног за регистроване кориснике преко веб-странице ЈеРТеха<sup>6</sup>.

Биљана Рујевић је дала практичну демонстрацију платформе *Leximirka*, направљене за обједињено допуњавање, одржавање и коришћење електронских речника, корпуса, и других лексичких ресурса и алата развијених у сарадњи БУ и ЈеРТеха. Платформи је, уз неопходну регистрацију, такође могуће приступити са веб-странице ЈеРТеха.

---

5. COST акција Distant Reading for European Literary History.

6. Веб-страница друштва ЈеРТех.

Михаило Шкорић представио је ВЕаST, новоразвијени POS тагер за српски језик који представља побољшање у перформансама у односу на тагере у својој основи: *treetagger*, *spaCy* и *stanza*. Овај тагер, који је уједно искоришћен за обележавање корпуса текстова ELTeC, доступан је за преузимање преко *GitHuba*. Бранислава Шандрих Тодоровић говорила је о веб порталу *NER & Beyond*, развијеном у оквиру COST акције *Distant Reading*. Портал је направљен да омогући обележавање именованих ентитета различитим доступним алатима, али и мапирање између анотација као производа тих алата. Ова платформа је доступна за употребу преко веб-странице *JePTexa*.

Радионица ELG била је право освежење. С једне стране, пружила је прилику да се сазна за пројекат ELG, као важан корак ка обједињавању и већој видљивости алата и ресурса за српски језик. С друге стране, учесници су стекли увид и у неке од последњих резултата истраживања у домену језичких технологија за српски језик и њихове примене. Динамична структура у оквиру које су представљана сервиса, ресурса и алата трајала по 15 минута, омогућила је да се за кратко време добије преглед великог броја пројеката, уз довољно информација о сваком од њих. Радујем се следећој прилици да присуствујем овако садржајном и добро организованом догађају, уз уверење да сваки овакав напор доприноси побољшању ситуације језичких технологија за српски језик.