

## ПРИКАЗ ПРОЈЕКТА CESAR

**Тамаш Варади\***

Институт за лингвистичка истраживања, Мађарска академија наука

С енглеског превео: Зоран Ристовић

Кратки чланак који следи пружиће приказ недавно покренутог сарадничког напора финансираног из фондова Европске уније, да се језички ресурси и алати региона Централне и Југоисточне Европе побољшају и прошире, и да постану доступни у лингвистичкој инфраструктури отвореног типа.

Далекосежни циљеви овог пројекта ваљало би да се сагледају у контексту лингвистичких изазова које са собом носи глобализована дигитална комуникација модерног доба. Многи европски језици суочавају се са ризиком да постану жртве дигиталног

доба будући да нису у довољној мери представљени или ресурсно заступљени онлајн. Огромне могућности регионалног тржишта остају и даље неискоришћене услед језичких баријера. Уколико сада не предузмемо акцију, сама чињеница да говоре својим матерњим језиком постаће социјални и економски недостатак за многе европске грађане.

Иновативна мултилингвална језичка технологија јесте најсавременији посредник који може помоћи свим грађанима Европе да суделују у једнаком и правичном, инклузивном и економски успешном друштву

\*varadi@nytud.hu

знања и информација. Језичка технологија може бити замајак тренутне, јефтине комуникације и интеракције без напора која превазилази језичке баријере.

Од препознавања говора и аутоматског сумирања текста до ископавања из текста и машинског превођења језичка технологија нуди све шире перспективе. Сви ови изванредни алати и технологије раде уз помоћ података, а тих података треба да буде што више и они треба да буду интегрисани са лингвистичким подацима у форми анотација. Ова додатна вредност је оно што податке претвара у драгоцене ресурсе – језичке ресурсе.

Пројекат CESAR (енг. - Central and South-east European Resources – Ресурси Централне и Југоисточне Европе) има циљ да допринесе размени дигиталних ресурса широм Европе путем прикупљања ресурса и њиховог документовања, повезивања и надградње како би били у складу са општеприхваћеним стандардима и смерницама. Пројекат ће бити доступан у виду јасно профилисаног скупа језичких ресурса и алата на бугарском, мађарском, пољском, словачком, српском, и хрватском језику. Ти ресурси подразумеваће интероперабилне једнојезичне и вишејезичне мешусобно повезиве говорне и писане базе података, корпусе, речнике и мреже речи (wordnet), као и следеће алате: токенизаторе, лематизаторе, тагере и парсере.

Пројекат CESAR део је шире мреже изврности назване META-NET коју финансира Европска унија. Она се тренутно састоји од 44 члана, који представљају 31 европску земљу. META-NET сарађује са десетак других великих иницијатива, као што је CLARIN, који пружа подршку друштвеним наукама да у Европи успоставе ново поље деловања - Дигиталне хуманистичке науке. META-NET је посвећен остваривању технолошких основа за успостављање и

одржавање правог вишејезичног европског информационог друштва које

- омогућава комуникацију и сарадњу на више језика,

- обезбеђује једнак приступ информацијама и знању свим корисницима на њиховом матерњем језику,

- нуди напредне могућности умрежене информационе технологије свим грађанима по приступачним ценама.

Мисија META-NET-а је да допринесе да изградња вишејезичног европског дигиталног информационог простора постигне успех попут писане културе након Гутенберга. CESAR активно сарађује са другим партнерским пројектима у оквиру META-NET-а ради изградње конзистентног приступа, праксе и стандарда који осигуравају шири и лакши приступ и поновну употребу квалитетних језичких ресурса.

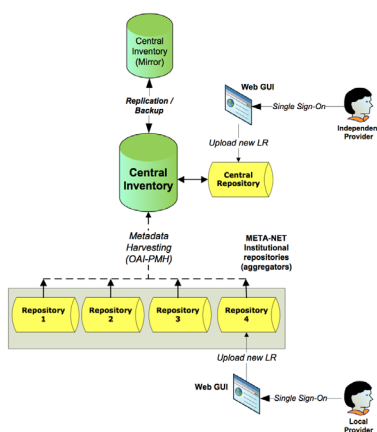
Пројекат CESAR тежи да подстакне комуникацију засновану на информационо-комуникационим технологијама, сарадњу и учешће на различитим језицима и тиме да допринесе креирању јединственог дигиталног тржишта широм Европе.

Један од основних циљева пројекта јесте да премости технолошки јаз који постоји између региона Централне и Југоисточне Европе и осталих делова Европе тако што тежи да попуни очигледне и значајне празнине које се тичу инфраструктурних језичких ресурса и алата. Информационо-комуникациона истраживања у овим земљама отпочела су са значајним заостатком у односу на земље Западне Европе. Језичка технологија развијала се у свим овим земљама независно, односно уз националну подршку како на академском нивоу тако и у оквиру приватног сектора. Али као последица тога, развијени ресурси често одсликавају изоловане околности свог настанка, и нису увек у складу са стандардима. Главни актери информационо-

комуникационих истраживања су у сваком случају сада спремни да обнове сарадњу међу кључним технолошким партнерима у региону, и да интегришу националне ресурсе на вишем нивоу како би их учинили доступнијим и употребљивијим, чинећи их доступним широј заједници језичких технологија и убрзавајући обезбеђивање вишејезичних онлајн сервиса. У том циљу, сви постојећи ресурси биће обједињени и унаоређени тако да буду у складу са широко прихваћеним стандардима и друштвеном праксом.

Кључни ресурси које покрива пројекат CESAR ће се повезати и постаће употребљивији коришћењем МЕТА-РАЗМЕНА (META-SHARE) репозиторијума. МЕТА-РАЗМЕНА тежи да изгради инфраструктуру за размену података отвореног типа. Циљна заједница корисника практично обухвата оне који имају удела у модерном дигиталном тржишту: свакодневне крајње кориснике, професионалне крајњекориснике (упривреди, администрацији, медијима, образовању, библиотекама, итд) као и стручњаке (истраживаче, индустријалце, доносиоце одлука, итд). То подразумева пажљиво испитивање потреба различитих типова корисника – од појединачних корисника до великих мултинационалних организација – како из перспективе тренутног стања тако и са становишта тенденција у непосредној будућности.

**META-SHARE**



Пројекат CESAR допринеће МЕТА-РАЗМЕНИ својим драгоценим ресурсима, што ће у крајњем исходу бити значајна компонента тржишта језичке технологије за истраживаче и развојне инжењере у области технологија природних језика, језичке стручњаке (преводиоце, експерте за локализацију софтвера и садржаја, итд), као и за велике индустријске играче, а нарочито мала и средња предузећа, обезбеђујући им комплетан развојни циклус технологије природних језика, од фазе истраживања до иновативних производа и услуга.

Један од главних циљева пројекта CESAR јесте да покрене на деловање националне и регионалне актере, јавна тела и агенције за финансирање пројеката, тако што ће организовати скупове и друга тематска догађања. Један такав догађај је конференција „МЕТА-Форум 2011 – Решења за вишејезичну Европу“, која је одржана у Будимпешти крајем јуна. То је званични догађај Мађарског председавања Саветом Европске уније. У оквиру конференције је одржано неколико сесија о европској индустрији језичких технологија и о вишејезичним језичким технологијама у европским институцијама, као и панел-дискусија о вишејезичном информатичком друштву, док су позвана предавања одржали представници Google-а и IBM-а. Више информација о овом догађају доступно је на [http://www.meta-net.eu/events/meta-forum-2011/index\\_html](http://www.meta-net.eu/events/meta-forum-2011/index_html).

Циљеви пројекта CESAR су усаглашени са МЕТА-NET-ом који треба да обезбеди материјал који описује националну сцену земаља-учесница. Тако је настао зборник белих књига МЕТА-NET-а, „Језици у европском информатичком друштву“, који даје увид у стање сваког европског језика у односу према језичким технологијама и објашњава највеће ризике и могућности.

Зборник покрива све званичне европске језике. Иако су до сада објављени многи значајни и свеобухватни научни прикази одређених аспеката појединачних језика и њима доступне језичке технологије, још увек не постоји опште прихваћен преглед који би сумирао сва главна сазнања и изазове за све језике заједно. Зборник белих књига МЕТА-NET-а има циљ да попуни ту празнину, а његова штампана верзија била је први пут подељена учесницима на МЕТА-Форуму у Будимпешти.

Пројекат је почео 1. фебруара 2011. и траје две године. Очекује се да ресурси буду објављени у три пакета. Први ће се појавити у децембру 2011, а након њега уследиће још један у јулу 2012. и коначно један јануара 2013.

Чланови конзорцијума пројекта су:

- Институт за лингвистичка истраживања, Мађарска академија наука (Мађарска)
- Технолошки универзитет у Будимпешти (Мађарска)
- Универзитет у Загребу, Филозофски факултет (Хрватска)
- Институт за рачунарство, Пољска академија наука (Пољска)
- Универзитет у Лоцу (Пољска)
- Математички факултет, Универзитет у Београду (Србија)
- Институт Михајло Пупин (Србија)
- Институт за бугарски језик (Бугарска)
- Институт за лингвистику Л. Штур (Словачка)

Координатор конзорцијума је Тамаш Варади из Института за лингвистичка истраживања Мађарске академије наука.

**Онлајн извори**

Основни подаци о пројекту

[http://ec.europa.eu/information\\_society/apps/projects/factsheet/index.cfm?project\\_ref=271022](http://ec.europa.eu/information_society/apps/projects/factsheet/index.cfm?project_ref=271022)

Веб страна пројекта

<http://www.meta-net.eu/projects/cesar/>

МЕТА-РАЗМЕНА

<http://www.meta-net.eu/meta-share>

