

СТУДИЈА ВОДЕЋИХ ЕВРОПСКИХ ЕКСПЕРАТА ЗА ЈЕЗИЧКЕ ТЕХНОЛОГИЈЕ УПОЗОРАВА ДА ВЕЋИНА ЕВРОПСКИХ ЈЕЗИКА НЕЋЕ ОПСТАТИ У ДИГИТАЛНОМ ДОБУ

Горг Рем и Ханс Ускорајт, уредници МЕТА-НЕТ серије белих књига „Европски језици у дигиталном добу“

с енглеског превео Иван Обрадовић

Најновија студија водећих европских експерата за језичке технологије упозорава да већини европских језика прети дигитално изумирање. Процењујући ниво подршке коју језичке технологије пружају за 30 од приближно 80 европских језика, експерти закључују да је дигитална подршка за 21 од тих 30 језика „непостојећа“ или у најбољем случају „слаба“. Студија је урађена у оквиру МЕТА-НЕТ-а, европске мреже изврности коју чине 60 истраживачких институција из 34 земље.

Студија, коју је припремило више од 200 експерата и која је документована кроз 30 томова Серије белих књига МЕТА-НЕТ-а (која је на

располагању и на вебу и у штампаној форми), оцењивала је подршку језичких технологија за сваки од језика у четири области: аутоматско превођење, говорна интеракција, анализа текста и расположивост језичких ресурса. Експерти су укупно 21 од 30 језика (70%) сврстали у најнижу категорију „подршка слаба или непостојећа“ у најмање једној од поменутих области. Неколико језика, као што су исландски, летонски, литвански и малтешки оцењени су најнижом оценом у све четири области. Што се језика на другом крају спектра тиче, ни за један није закључено да постоји „одлична подршка“, а само за енглески језик је оцењено да је подршка

„добра“. За енглеским следе холандски, француски, немачки, италијански и шпански за које постоји „умерена подршка“. За језике као што су баскијски, бугарски, каталонски, грчки, мађарски и пољски постоји „фрагментарна подршка“, што и њих сврстава у скуп језика са високим степеном ризика. Овој групи језика припада и српски, као и сви други језици са простора бивше Југославије. Има више разлога због кога се српски нашао у овој опасној зони. Највећи број текстова, пре него што буде отиснут на папиру или објављен на вебу, настаје уз помоћ рачунара. Али за разлику од читаоца, који ће разумети написани или изговорени текст, рачунари најчешће ништа од тога „не знају“ и „не разумеју“ осим оних који се налазе у ретким истраживачким лабораторијама. Могућност да се једна колекција докумената аутоматски „паметно“ претражи, да се из ње извуку прецизне информације или да се појединачни документ резимира не постоји за наш језик. С друге стране, велики произвођачи текста, нпр. новинске и издавачке куће, само изузетно организују резултате своје производње на начин који би омогућио примену поменутих језичких технологија.

“Резултати наше студије су готово алармантни. Већина европских језика пати од озбиљног недостатка ресурса, а неки су у потпуности занемарени. У том смислу, многим од наших језика још увек није осигурана будућност”, каже професор Ханс Ускорајт, координатор МЕТА-НЕТ-а, научни директор DFKI (Немачки истраживачки центар за вештачку интелигенцију) и један од уредника поменуте студије. Један други уредник, др Георг Рем (DFKI), додаје: “Постоје драматичне разлике у подршци које дају језичке технологије између појединих европских језика и појединих технолошких области. Јаз између „великих“ и „малих“ језика се и даље продубљује. Треба предузети све што је у нашој могућности како

бисмо мале језике и језике којима недостају ресурси снабдели неопходним основним технологијама, јер су у противном ови језици осуђени на дигитално изумирање.”

Област језичких технологија производи софтвер којим се обрађује говорни и писани језик. Добро познати примери таквог софтвера су алати за проверу правописа и граматике, интерактивни лични помоћници на паметним телефонима (као што је Siri на iPhone-у), дијалогски системи у телефонији, системи за аутоматско превођење, машине за претрагу веба, као и синтетички гласови који се користе у системима за навигацију у аутомобилима. Данашњи језичко-технолошки системи се ослањају пре свега на статистичке методе које захтевају невероватно велике количине писаних и говорних података. За језике са релативно малим бројем говорника веома је тешко обезбедити такво обиље података. Сем тога, статистички језичко-технолошки системи имају инхерентна ограничења када је у питању квалитет, што се може видети, на пример, из често забавних неправилних превода које дају системи за машинско превођење на вебу.

Европа је успела да уклони готово све границе између појединих земаља. Међутим, једна граница још увек постоји, и чини се непробојном: то је невидљива граница коју представљају језичке баријере и која омета слободан проток знања и информација. Она такође наноси штету и дугорочном циљу успостављања јединственог дигиталног тржишта, јер омета слободан проток добара, производа и услуга. Док са једне стране језичке технологије имају потенцијал да уклоне језичке баријере кроз системе за машинско превођење, резултати студије МЕТА-НЕТ-а јасно показују како многи европски језици за то још увек нису спремни. Постоје значајне технолошке празнине које треба попунити, а које су настале због усредсређености већег

дела истраживања и развоја на енглески језик, недовољног ангажовања и финансијских ресурса, као и недостатака јасне истраживачке и технолошке визије.

Потребна је координирана, свеобухватна европска иницијатива како би се произвеле технологије које недостају, и како би се обезбедио трансфер ових технологија према већини језика. Постоје снажни разлози да се овом огромном изазову приђе кроз иницијативу у коју ће бити укључена Европска унија, њене државе чланице као и придружене државе, али и представници привреде. Неки од тих разлога су високо финансијско „per capita“ оптерећење за мале језичке заједнице, потреба да се обезбеди трансфер технологија између језика, међусобна неповезаност постојећих ресурса, алата и услуга, и чињеница да се језичке границе често не поклапају са политичким. Европа мора предузети иницијативу како би припремила своје језике за дигитално доба. Језици представљају драгоцену компоненту нашег културног наслеђа и као такви заслужују да им се обезбеди сигурна будућност. Савет Европе 26. септембра обележава Европски дан језика, дан којим се одаје признање значају неговања и развоја богатог језичког и културног наслеђа нашег континента. Рад МЕТА-НЕТ-а представља оштро упозорење на изазове и могућности са којима се суочава наше језичко наслеђе у информационом добу.

Језичке технологије: позадина

Језичке технологије нам већ сада пружају подршку у свакодневним задацима као што је састављање електронске поште или куповина карата. Добробити језичких технологија користимо када претражујемо или преводимо веб странице, када употребљавамо алате за проверу правописа и граматике у програмима за обраду текста, када управљамо системом за разоноду у нашем аутомобилу или нашим мо-

билним телефоном помоћу говорних наредби, када добијамо препоруке у продавници књига на вебу, или кад следимо инструкције које нам даје мобилна апликација за навигацију. У скорој будућности бићемо у могућности да се обраћамо рачунарским програмима, као и машинама и кућним апаратима, укључујући и дуго очекиване услужне роботе који ће ускоро ући у наше домове и наше радне просторије. Где год да се налазимо, када нам буде потребна информација, једноставно ћемо за њу запитати, а када нам буде потребна помоћ, затражићемо је гласом. Уклањање комуникационе баријере између људи и технологије ће променити свет.

Опште је мишљење да језичке технологије данас представљају једну од кључних области развоја информационих технологија. Велике међународне корпорације као што су Google, Microsoft, IBM, и Nuance су уложиле значајна средства у ову област. Стотине малих и средњих предузећа у Европи се специјализовало у одређеним применама језичких технологија и пружању услуга у овој области. Језичке технологије омогућавају људима да сарађују, уче, обављају пословне трансакције и размењују знања превазилазећи језичке границе независно од својих информатичких вештина.

Серија белих књига МЕТА-НЕТ-а

Серија белих књига МЕТА-НЕТ-а „Језици у европском информационом друштву“ извештава о стању језичких технологија у 30 европских језика и образлаже шта су најургентнији ризици и изазови. Серија покрива све званичне језике земаља чланица ЕУ и неколико других језика који се говоре у Европи. Мада је и раније било неколико вредних и свеобухватних студија о одређеним аспектима језика и технологије, до сада није било опште разумљивог зборника који представља основне постигнуте резултате и изазове који

стоје пред сваким од језика са становишта технолошке подршке вишејезичној Европи. Серија белих књига МЕТА-НЕТ-а попуњава ову празнину. МЕТА-НЕТ сада може да покаже зашто се већина језика налази пред озбиљним проблемима и да укаже на најопасније празнине. Више од 200 аутора и сарадника помогло је у припреми белих књига о језицима.

Беле књиге су написане за следеће европске језике: баскијски, бугарски, каталонски, хрватски, чешки, дански, холандски, енглески, естонски, фински, француски, галицијски, немачки, грчки, мађарски, исландски, ирски, италијански, летонски, литвански, малтешки, норвешки (бокмал и нинорск), пољски, португалски, румунски, српски, словачки, словеначки, шпански и шведски. Бела књига о сваком од ових језика написана је на језику на који се односи, а обухвата и превод целокупног текста на енглески.

О МЕТА-НЕТ-у и МЕТИ

МЕТА-НЕТ је мрежа изврности коју чине 60 истраживачких центара из 34 земље, и посвећена је изградњи технолошке основе за вишејезично европско информационо друштво. Европска комисија пружа финансијску подршку МЕТА-НЕТ-у кроз укупно четири пројекта.

МЕТА-НЕТ гради вишејезични европски технолошки савез МЕТА (Multilingual Europe Technology Alliance). Више од 600 организација из 55 земаља, укључујући истраживачке центре, универзитете, мала и средња предузећа, као и неколико великих корпорација су се већ придружили овом отвореном технолошком савезу.

Контакт:

Prof.Dr. Hans Uszkoreit
Dr. Georg Rehm
META-NET Office c/o DFKI Berlin
Alt-Moabit 91c
10559 Berlin, Germany
Телефон: +49 (0)30 23895-1833
Е-пошта: office@meta-net.eu

Додатне информације:

МЕТА-НЕТ: <http://www.meta-net.eu>
Беле књиге: <http://www.meta-net.eu/whitepaper>