

## ГРУПНА РАСПОДЕЛА РАДА И ЊЕГОВА ПРИМЕНА

Јелена Митровић, jmitrovic@gmail.com, Универзитет у Београду, Филолошки факултет

### Апстракт

Циљ овог рада је да укаже на значај групне расподеле рада (енг. Crowdsourcing), те да пружи приказ неких успешних пројеката који функционишу на основу овог модела управљања који се из пословног света преселио у свет културе и науке. Спроведено је истраживање начина на који системи групне расподеле рада функционишу као и технологија на којима се заснивају. Утврђени су неки од услова које је потребно испунити да би један систем групне расподеле рада био успешан и дати су неки примери добре праксе, с посебним освртом на пројекте у области обраде природних језика (енг. Natural Language Processing – NLP) који су спроведени захваљујући овом новом начину расподеле рада.

### Кључне речи

групна расподела рада, микрозадаци, MTurk, обрада природних језика, wordnet

### 1. Увод

Са развојем веб 2.0 (енг. Web 2.0) технологија (Mugugesan 2007) омогућен је бржи развој и напредак парадигме дигиталне хуманистике као интердисциплинарне области истраживања која повезује хуманистичке науке са информационо-комуникационим технологијама (енг. ICT). За разлику од многих других интердисциплинарних области истраживања, дигитална хуманистика има веома добро познат почетак – 1949. годину,

када је отац Роберто Буса, италијански језуитски свештеник, одлучио да уз помоћ рачунара направи такозвани index verborum свих речи које су коришћене у делима Светог Томе Аквинског, што је око једанаест милиона речи средњовековног латинског. Данас је тај индекс познат као Index Thomisticus (Schreibman et.al. 2004). Термин веб 2.0 први пут је у чланку под називом “Fragmented future” јануара 1999. године употребила Дарси ДиНучи (Darcy Dinucci). Она је у

свом чланку нагласила да је тадашњи веб, који је још увек био статичан и заснивао се на пуком читавању страница у веб читаче, само ембрион веба који ће тек настати. Дарси ДиНучи је предвидела да ће веб постати нека врста транспортног механизма који ће омогућити интерактивност (Dinucci 1999). Веб 2.0 је, поред многих других иновација, донео и сајт Википедија (енг. Wikipedia), без кога данас не можемо да замислимо потрагу за информацијама, а који дозвољава размену и унапређивање знања на многим нивоима. Сајтови друштвених мрежа као што су Фејсбук (енг. Facebook) и Твитер (енг. Twitter), као и многобројни сајтови настали захваљујући блог технологији, отварају потпуно нове хоризонте сарадње, и ако се користе на прави начин, могу умногоме допринети развоју разних области хуманистике. Ови сајтови се често користе за пројекте групне расподеле рада (енг. crowdsourcing), а Википедија се сматра једним од првих таквих пројеката (Doan et.al. 2011). Уопштено говорећи, веб 2.0 је технолошка основа на којој се групна расподела рада развија и делује (Vuković and Bartolini 2010).

## **2. Неке дефиниције групне расподеле рада**

Данас постоје многе врсте групне расподеле рада, али и разне дефиниције овог појма. Термин групна расподела рада увео је Џеф Хауи (енг. Jeff Howe) и назвао га пословном праксом која дословно значи расподелити задатке (енг. термин који је употребљен је „outsource“) у неком пројекту групи људи (Howe 2006). Хауи је указао на значај овог начина расподеле рада и назвао га механизмом по коме се таленат и знање једне групе људи могу упарити са потребама других (Howe 2008). Аутор (Brabham 2008a) групну расподелу рада дефинише као модел дистрибуиран преко интернета, који се

користи за решавање проблема и производњу, а користе га профитне организације попут Threadless и iStock. Исти аутор у другом раду (Brabham 2008b) дефинише групну расподелу рада као стратешки модел који привлачи заинтересовану, мотивисану групу људи који су способни да пруже решења која квалитетом и квантитетом превазилазе решења која се могу добити традиционалним облицима пословања. Аутори (Buecheler et.al. 2010) описују групну расподелу рада као посебну врсту колективне интелигенције (енг. collective intelligence). У раду (Grier 2011) се групна расподела рада описује као начин да се интернет употреби за коришћење доприноса великог броја људи који другачије не би имали прилику да сарађују на истим пројектима. Други аутор (Kazai 2011) ову врсту сарадње сматра отвореним позивом групи људи да реше неки проблем или да изврше неки задатак за који је неопходна људска интелигенција, дакле рачунари га не могу обавити, обично у замену за малу новчану надокнаду, стицање статуса у друштву или ради забаве. Једна од најсвеобухватнијих дефиниција групне расподеле рада дата је у (Estellés-Arolas and González-Ladrón-de-Guevara 2012) – „Групна расподела рада је врста учествовања у некој активности у којој особа, институција, непрофитна организација или компанија нуде групи особа различитих нивоа знања, хетерогености и броја, преко флексибилног отвореног позива, да се добровољно посвете неком задатку. Посвећивање задатку, који је разнолике комплексности и модуларности и у коме група људи учествује својим знањем и/или искуством, новцем, радом, увек укључује узајамну корист. Кориснику ће бити задовољена нека потреба, било да је то економска, друштвена, лична потреба, или развијање сопствених вештина, док ће особа, институција или организација која нуди задатак, оно чиме су корисници допринели,

а чији облик ће зависити од врсте задатка, искористити у своју корист.”

### **3. Управљање пројектом групне расподеле рада**

При организовању пројеката групне расподеле рада морамо имати спремне одговоре на многа питања. Од одговора на та питања умногоме зависи успешно управљање пројектима које желимо да спроведемо ослањајући се на парадигму групне расподеле рада. У том смислу, нека од најзначајнијих питања биће размотрена у наредним сегментима овог рада.

#### **3.1. Ко чини групу људи који ће учествовати у пројекту?**

На основу до сада спроведених пројеката групне расподеле рада показало се да је за различите врсте ових пројеката неопходно ангажовати другачије структуриране групе људи, како би групна мудрост (енг. *Wisdom of the Crowd*) била што боље искоришћена. Групна мудрост односи се на појаву да у неком пројекту група која се састоји од великог броја учесника може бити много успешнија од неколико стручњака. На примеру сајта Википедија доказано је да велики број учесника који чине једну хетерогену групу, у смислу нивоа образовања и стручности, побољшава ниво квалитета чланака на Википедији (Arazy et.al 2006). У неким пројектима је боље користити допринос што хетерогеније групе особа (на пример, за пројекте обележавања слика или за пројекте који укључују прикупљање мишљења о новим производима), Ипак, у неким пројектима је неопходно упослити групу људи која је хомогена, то јест сличног нивоа образовања и стручности, како би резултати били бољи, на пример у случају уметничких пројеката, где је неопходно изнаћи решење за дизајн новог производа итд. (Estellés-Arolas and González-

Ladrón-de-Guevara 2012), као што је случај са креативним подухватима (на пример, у претходно поменутих пројектима Threadless и iStock).

#### **3.2. Како ће учесници обављати задатке?**

Допринос учесника у једном пројекту групне расподеле рада умногоме зависи од начина расподеле задатака. Када говоримо о начину расподеле задатака групи људи у једном пројекту групне расподеле рада, постоји подела на микрозадатке и макрозадатке. У случају микрозадатака, сваки учесник у пројекту групне расподеле рада испуњава више малих делова целокупног задатка, да би се на крају сва решења „склопила“ у коначно решење до кога је требало доћи у датом пројекту. Макрозадаци, с друге стране, односе се на пружање читавог задатка или проблема на увид групи људи, како би свако могао да одреди који и колики део ће решити и како ће пројекту допринети у складу са својим знањима и компетенцијама.

Показало се да се кроз микрозадатке задаци обављају брже и ефикасније него када би те исте задатке на уобичајен начин обављале запослене особе у оквиру обавеза које им налаже радно место (Chamberlain et.al. 2012). Микрозадаци су мали, добро дефинисани задаци које преко одређених платформи (као што су MTurk и Crowdfunder) обавља група људи. Проблеми који се могу решити преко микрозадатака су они проблеми који се лако могу поделити на велики број једноставних задатака, како би радници, то јест учесници у пројекту групне расподеле рада, могли лако да их реше, за релативно кратко време, без поседовања неких специфичних знања и вештина. Тако се микрозадаци могу користити за означавање и разврставање садржаја (Маји 2011), за проналажење специфичних садржаја на интернету, и уопште за задатке које људи с

лакоћом обављају, за разлику од рачунарских програма.

Да би се микрозадаци пружили на увид потенцијалним учесницима у пројекту неопходно је да буду добро дефинисани, дакле да учесницима буде јасно шта треба да ураде, те да микрозадаци буду на неки начин организовани у групе, било да је то по сличности или по припадности истом делу пројекта. Једна од најпопуларнијих платформи на којој се може спровести пројекат групне расподеле рада преко микрозадатака јесте *Mechanical Turk* (MTurk), компаније Amazon. Микрозадаци на MTurk платформи носе назив НИТ (енг. Human Intelligence Tasks). Приликом постављања микрозадатака, вођа пројекта може да одреди више параметара – колико одговора је потребно за сваки микрозadataк; може се одредити коме је дозвољено да решава неке задатке, рецимо, у зависности од нивоа познавања неког страног језика ( у случају пројекта који се баве побољшавањем резултата машинског превођења), итд. Amazon Mechanical Turk уствари пружа API (енг. Application Program Interface) то јест апликациону програмску сумеђу за веб сервисе (ова сумеђа дефинише начине на које веб сервис може да се користи, те одређује речник и правила позивања која програмер треба да примени да би користио те сервисе). Она се може користити да би се преко Amazon Mechanical Turk веб сајта задаци поставили, да би се они који су испуњени одобрили и да би се одговори уврстили у друге софтверске апликације. (Kittur 2008; Vukovic et. al. 2010; Doan et.al. 2011)

### **3.3 Како мотивисати учеснике да допринесу пројекту групне расподеле рада?**

Ово је једно од најважнијих питања која се постављају приликом организовања једног пројекта групне расподеле рада.

Аутори (Estellés-Arolas and González-Ladrón-de-Guevara 2012) повезују надокнаду учесницима са Масловљевом пирамидом потреба (Maslow 1943). Тако мотивација, поред новчане надокнаде, може бити и прилика да се развију креативне способности и друге вештине, да се забавимо, да поделимо знање с другима, љубав према заједници, приврженост задатку који треба обавити, жеља да у друштву коме припадамо будемо препознати захваљујући својим доприносима неком пројекту. У (Chamberlain et.al. 2012) су наведена три главна начина за подстицање учесника пројекта групне расподеле рада: лични, друштвени и финансијски начин.

1) Лични мотиви огледају се у жељи учесника да једноставно буду део неког пројекта. Уопштено речено, лични мотив је присутан увек када је кориснику сáмо учествовање у пројекту довољна награда и када неко жели да допринесе пројекту за који сматра да је вредан. Дobar пример мотивације учесника види се у пројекту Duolingo, у коме учесници преводе делове веба, а оно што добијају заузврат јесте прилика да бесплатно уче стране језике, и то на добро структуриран и стручан начин (Vesselinov 2012).

2) Друштвени мотив и начин подстицања односи се на награђивање побољшавањем друштвеног статуса корисника у контексту друштвених интернет мрежа и у односу на друге учеснике у пројекту групне расподеле рада. Показало се да је коришћење система бодовања и нивоа ефикасно, јер учесници често испуњавају више задатака прописаних пројектом управо да би достигли већи број бодова и прешли на следећи ниво (Von Ahn and Dabbish 2008).

3) Када говоримо о новчаној надокнади, то јест подстицању учесника пројекта групне расподеле рада на финансијски начин, Amazon Mechanical Turk се намеће као платформа која се у те сврхе најчешће

користи. За сваки задатак који учесник обави преко ове платформе, одређена је сума новца коју ће добити. Amazon.com, као власник ове платформе, за сваку појединачну исплату коју учесник у пројекту добије узима додатних десет посто од особе или организације која је осмислила дати пројекат. Најнижа цена која се може понудити за појединачни задатак на овој платформи јесте \$0.005 по задатку.

### **3.4. Како проверити квалитет доприноса учесника у пројекту групне расподеле рада?**

Расподелом задатака на платформи каква је Amazon Mechanical Turk и друге сличне платформе, увек постоји опасност од злоупотреба, то јест дешава се да учесници намерно дају погрешне одговоре или решавају неке задатке само да би испунили минимум прописаних услова и добили одређену компензацију. Зато се користе разне методе евалуације, почев од провере редунданцијом (енг. Redundancy check) на основу модела провере у којима више учесника обавља исти задатак (Oleson et.al. 2011) чиме се, поред утврђивања тачног одговора, може измерити квалитет доприноса учесника у једном пројекту групне расподеле рада. Даље, често коришћен метод евалуације описан је у (Dawid and Skene 1979), где се провера поузданости доприноса учесника заснива на статистичком алгоритму под називом Очекивање-максимизација (енг. Expectation Maximization – EM) чији је резултат скуп процењених тачних одговора за сваки задатак, као и матрица која, за сваког учесника понаособ, даје вероватноћу прављења грешке. Из те матрице се директно може измерити квалитет рада сваког радника. Сличан приступ представљен је у (Ireigotis et.al. 2010) где се EM алгоритам користи за израчунавање квалитета доприноса за сваког учесника у пројекту, а прилагођен је

случајевима када може доћи до грешака, те се за сваког учесника добија и стопа грешке, чиме се може израчунати право стање квалитета рада појединачних учесника у пројекту групне расподеле рада.

Решење које се успешно примењује у многим пројектима групне расподеле рада јесте успостављање система „златних задатака“ (у литератури се јављају и термини „златни стандард“ и „златна јединица“) којима се унапређује основна функционалност платформи на којима се извршавају задаци групне расподеле рада. Златни задаци су они задаци за које је одговор већ познат, или је тривијалан, те се прецизност одговора учесника у пројекту одређује на основу одговора на златне задатке. Помоћу њих се може установити да ли се неким учесницима може веровати, јер ако су дали тачне одговоре на питања за која систем већ зна одговор, постоји одређени степен сигурности да ће и одговори који систему нису унапред познати бити тачни. На сличном принципу функционише и технологија ReCAPTCHA која се користи у многим пројектима дигитализације књига, као и у другим пројектима групне расподеле рада из области културе (Митровић 2013). ReCAPTCHA је бесплатна CAPTCHA технологија која помаже у дигитализацији књига, новина, и старих радио емисија. CAPTCHA (акроним од речи Completely Automated Public Turing Test to Tell Computers and Humans Apart) је програм који може да утврди да ли је његов корисник људско биће или рачунарски програм. ReCAPTCHA технологија заснива се на укуцавању две речи у понуђени оквир на веб страници у циљу аутентикације. Једна од те две речи је систему позната, док је друга реч остала препозната техником оптичког препознавања карактера (енг. Optical Character Recognition – OCR). Ако корисници, то јест особе које користе

интернет, правилно укуцају реч за коју је одговор познат, систем претпоставља да је њихов одговор за другу реч такође тачан. Затим се та иста слагалица с две речи нуди другим корисницима, статистичким моделима се одређује проценат тачности одговора, те је крајњи резултат адекватно препозната реч. Можемо закључити да је ReCAPTCHA светски пројекат групне расподеле рада заснован на микрозадацима јер корисници интернета широм света свакодневно помажу дигитализацију књига, реч по реч.

#### 4. Евалуација у пројектима обраде природних језика

Ако се постојеће платформе групне расподеле рада користе уз додатне технике управљања и уз проверу квалитета обављених задатака, то јест уз евалуацију доприноса корисника, многи комплексни задаци могу се уз овај начин расподеле посла обавити веома успешно. Тако се микрозадаци могу користити и у многим пројектима из области обраде природних језика, обично за разне врсте анотације и валидације. Такође, у неким пројектима у којима се користи више платформи и више начина сакупљања доприноса корисника, потребно је изнаћи решење за другачији начин евалуације.

У раду (Munro et.al. 2010) је за евалуацију у неколико пројеката коришћен *Kana коефицијент*. Капа коефицијент (Carletta 1996) се користи у анализи садржаја у многим пројектима из области рачунарске лингвистике, јер дозвољава упоређивање различитих резултата корисника и користи се у случајевима када одређивање поузданости резултата уобичајеним методама (на пример, помоћу златних задатака) није довољно.

У пројекту CROWDMAP (Sarasua et.al. 2012) поред MTurk платформе коришћена је и платформа CrowdFlower, која је основа имплементације пројекта. CrowdFlower

делује као посредник јер истовремено поставља микрозадатке на различитим платформама (укључујући MTurk, Crowd Guru, getpaid, Snapvertise). У овом пројекту се групна расподела рада користи да би се преко микрозадатака истражила веома значајна област у обради природних језика, поклапање онтологија (енг. ontology alignment), то јест да би се већ постојећи, аутоматизовани процеси, побољшали доприносом људи. Евалуација је извршена на основу златних задатака, као и помоћу *Крипендорфовог алфа коефицијента*.

*Крипендорфов алфа коефицијент* (енг. Krippendorff's alpha coefficient) развијен је за мерење поклапања доприноса посматрача, оцењивача, или неких инструмената за мерење којима се утврђују разлике између типично неструктурираних феномена, као и у случајевима када су подаци непотпуни. Крипендорфов алфа коефицијент првобитно је коришћен за анализу садржаја, али примену проналази у свим пројектима у којима се на исти скуп објеката или јединица анализе примењују две или више метода генерисања података, а потребно је закључити колико се резултирајућим подацима може веровати. (Krippendorff 2011).

У пројекту генерисања и анализе специјализованог корпуса који се може користити за идентификовање сарказма и ироније у текстовима (Filatova 2012), такође су коришћене две контролне процедуре. Прва врста провере је просто *већинско гласање* (енг. majority voting) – учесници су гласали за најбоља решења која су понудили други корисници, док је друга врста провере заснована на претходно поменутом Крипендорфовом алфа коефицијенту.

У пројекту Phrase Detectives (Chamberlain et.al. 2012), снага друштвене мреже Фејсбук искоришћена је за пројекат групне расподеле рада у коме учесници играјући игрице анотирају, то јест обележавају делове

текстова. Играчи на почетку „тренирају“ на текстовима који су одређени као златни стандард, или златни задаци (то су текстови које су аотирани стручни лингвисти). Када играч заврши аотирање текстова одређених за златни стандард, додељује му се оцена на основу процента тачно аотираних „златних“ текстова. Овај пројекат спада у групу пројеката групне расподеле рада под називом *Игре са сврхом* (енг. Games with a Purpose). Значајан пројекат заснован на Играма са сврхом је и пројекат дигитализације Народне библиотеке Финске, Digitalkoot, у коме се видео игрице користе за исправљање грешака насталих у процесу оптичког препознавања карактера. (Митровић, и др. 2010/2011). Неке од првих игара са сврхом биле су представљене пројектима ESP и Verbosity (Von Ahn and Dabbish 2008). Transcribe Bentham је такође веома важан пројекат групне расподеле рада. Настао је на Лондонском универзитетском колеџу, а циљ му је да преко интернета ангажује велики број људи у циљу транскрипције оригиналних и непроучених рукописа које је написао велики филозоф и реформатор Џереми Бентам (енг. Jeremy Bentham) (1748-1832). До сада је у оквиру овог пројекта, захваљујући непроцењивој помоћи учесника, транскрибовано 95 посто сачуваних рукописа (Causser et.al 2012).

### **5. Дограђња Српског wordneta путем групне расподеле рада**

Wordnet (Fellbaum 1998) је хијерархијска лексичко-семантичка мрежа која има непроцењиву вредност у свим областима истраживања које комбинују рачунарске технологије и лингвистику, било да те области називамо обрадом природних језика, рачунарском лингвистиком, или једноставно неком граном дигиталне хуманистике (Biemann and Nygaard 2010; Rumshisky 2011; Biemann 2012). У овој јединственој

мрежи, лексички концепти организовани су у такозване синсетове, то јест концепте груписане по сличности. Српски wordnet (енг. Serbian WordNet – SWN) почео је да се развија у оквиру пројекта BalkaNet (енг. Balkan WordNet project – BWN), као што је то био случај и са бугарским, чешким, румунским, грчким и турским wordnetom (Stamou et.al. 2002). SWN се заснива на структури принстонског wordneta (енг. Princeton WordNet – PWN), те је изграђен на принципу такозваног модела проширивања (енг. expand model), у складу с правилима која је налагао пројекат BalkaNet, то јест копирањем синсетова из принстонског wordneta у српски и преведећи их, уз очување хијерархијске структуре принстонског wordneta.

По завршетку пројекта BalkaNet, 2004. године, нови синсетови додавани су захваљујући удруженом раду професора, сарадника и студената Катедре за библиотекарство и информатику, Филолошког факултета у Београду и Групе за језичке технологије Математичког факултета у Београду, као и захваљујући бројним волонтерима. Ипак, данас српски wordnet броји око 20,000 синсетова, док принстонски броји око 117,000 синсетова. Зато би процес доградње овог важног ресурса требало на неки начин убрзати. У том смислу бисмо могли да искористимо модел групне расподеле рада.

Један од пројеката групне расподеле рада који би могао да се спроведе у смислу доградње српског wordneta могао би да буде одређивање најфреквентнијих придева у српском језику, а у односу на енглески језик, то јест, у односу на придеве који су увршћени у принстонски wordnet, јер нам је он основ доградње. Тако се у принстонском wordnetу налази 20479 придева, у поређењу са 1488 у српском wordnetу. Уз значајно повећање броја придева, српски wordnet би био кориснији за значајан процес у обради природних језика –

анализу осећања (енг. Sentiment analysis), за шта је веома погодан због своје веома добро развијене семантичке структуре (Torii et.al. 2011).

Пројекат одређивања релевантних придева би се заснивао на микрозадацима. Веома је важно да задаци буду што јаснији и што лакши, како би што већи број учесника желео да својим радом допринесе пројекту. Системом једноставног гласања понудили бисмо корисницима да одреде да ли је неки придев из енглеског језика (извучен из принстонског wordneta), то јест његов пандан на српском језику, релевантан за тог корисника. Број задатака које би индивидуални учесник морао да испуни би био ограничен. Исти скупови задатака би били понуђени великом броју учесника. По истеку одређеног временског периода, постављани би нови скупови задатака, док не бисмо дошли до већег броја података, то јест до већег броја релевантних придева које бисмо додали у српски wordnet. Из скупа података које бисмо добили гласањем били би елиминисани одговори свих оних учесника који нису задовољили одговоре на златна питања (која су неопходна у једном оваквом пројекту) док ће се над преосталим подацима вршити статистичка обрада израчунавањем средње вредности ДА и НЕ одговора, тако што ће ДА имати вредност 1, док ће одговор НЕ имати вредност 0. Тако ћемо добити процентуалну вредност релевантности придева, то јест видећемо да ли је дати придев, по мишљењу учесника у пројекту, релевантан и често у употреби.

Овај пројекат би захтевао коришћење групне мудрости релативно хомогене групе људи, с обзиром да би учесници морали да поседују одређени ниво знања енглеског језика и српског језика, како би могли да донесу добру одлуку, јер би понуђени придеви били на енглеском језику, а задатак учесника би био да одреде да ли би тај придев преведен на

српски језик био релевантан. Овакав пројекат бисмо, за почетак, могли да спроведемо преко алата компаније Гугл под називом Google form, који се користи у оквиру сервиса Google Drive. Google form олакшава употребу различитих врста упутника који лако могу да се поделе на друштвеним мрежама, али и да се пошаљу на адресе електронске поште потенцијалних учесника у пројекту. Резултате упитника организатори пројекта добијају у облику табеларног приказа, у коме се види и време уноса.

Мотивација учесника у оваквом пројекту била би личне и друштвене природе, то јест, то би била спремност и воља да се помогне у процесу побољшавања српског wordneta, као важног ресурса за обраду српског језика. Учесници би били ангажовани путем друштвених мрежа, као и путем веб презентација удружења љубитеља српског језика у целом свету, студената енглеског језика на факултетима у Србији, те студената српског језика на енглеским говорним подручјима – свим потенцијалним учесницима који могу допринети једном оваквом пројекту.

### **Закључак**

Модел групне расподеле рада се показао веома успешним у пројектима из области културе и науке, а посебно је значајан у обради природних језика, јер је још увек неопходно да људи проверавају аутоматске методе које се у неким истраживањима примењују с мање или више успеха. Када говоримо о платформама групне расподеле рада као што је MTurk, учесници добијају новчану надокнаду, те није неопходно мотивисати их на неки други начин. У случајевима када учесници једног пројекта групне расподеле рада том пројекту доприносе на добровољној основи, неопходно их је на неки начин заинтересовати, што представља основу успеха сваког пројекта



групне расподеле рада. Докле год можемо да пронађемо адекватан начин да употребимо људску креативност, под условом да обратимо пажњу на све аспекте пројекта групне расподеле рада, овај модел сарадње може да донесе велику корист и у многим домаћим пројектима. Српски wordnet (Krstev et.al. 2004) се тако техником групне расподеле рада и коришћењем неких од метода евалуације наведених у овом раду може унапредити и надоградити, што је правац истраживања који је неопходан за развијање алата и ресурса који нам стоје на располагању за обраду српског језика путем рачунарских технологија.

### Литература

- Arazy, Ofer, Morgan, Wayne, Patterson, Raymond. 2006. Wisdom of the Crowds: Decentralized Knowledge Construction in Wikipedia. *16th Annual Workshop on Information Technologies & Systems*. Milwaukee, USA.
- Biemann, Chris and Nygaard, Valerie. 2010. Crowdsourcing Wordnet. *Proceedings of the 5th Global WordNet conference*. Mumbai, India.
- Biemann, Chris. 2012. Turk Bootstrap Word Sense Inventory 2.0: A Large-Scale Resource for Lexical Substitution. *Language Resources and Evaluation*.
- Brabham, Daren, C.a. 2008. Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. *Convergence: The International Journal of Research into New Media Technologies*:75-90.
- Brabham, Daren, C.б. 2008. Moving the crowd at iStockphoto: The composition of the crowd and motivations for participation in a crowdsourcing application. *First Monday* 13(6).
- Buecheler, Thierry, Sieg, Jan, Henrik, Fuchslin Rudolf M, Pfeifer, Rolf. 2010. Crowdsourcing. Open Innovation and Collective Intelligence in the Scientific Method. *Proceedings of the Twelfth International Conference on the Synthesis and Simulation of Living Systems*: 679-686, Odense, Denmark.
- Carletta, Jean. 1996. Assessing Agreement on Classification Tasks: The Kappa Statistic. *Computational Linguistics* 22(2).
- Causer, Tim, Tonra, Justin and Wallace, Valerie. 2012. Transcription maximized; expense minimized: Crowdsourcing and editing The Collected Works of Jeremy Bentham. *Literary and Linguistic Computing* 27(2).
- Chamberlain, Jon, Kruschwitz, Udo, Poesio, Massimo. 2012. Motivations for Participation in Socially Networked Collective Intelligence Systems. *Collective Intelligence*.
- Dawid, Alexander, Skene, A.M. 1979. Maximum likelihood estimation of observer error-rates using the EM algorithm. *Applied Statistics* 28(1) : 20-28.
- Dinucci, Darcy. Fragmented Future. 1999. [http://darcy.d.com/fragmented\\_future.pdf](http://darcy.d.com/fragmented_future.pdf) (приступљено 3.03.2013).
- Doan, Anhai, Raghu Ramakrishnan, Alon Y. Halevy. 2011. Crowdsourcing Systems on the World-Wide Web. *Communications of the ACM* 54:86-96.
- Estellés-Arolas, Enrique and González-Ladrón-de-Guevara, Fernando. 2012. Towards an Integrated Crowdsourcing Definition. *Journal of Information Science* 38(2) : 189-200.
- Fellbaum, Christiane, ed. 1998. WordNet: *An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Filatova, Elena. 2012. Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing. *Language Resources and Evaluation (LREC)*, Istanbul, Turkey.
- Grier, David Alan. 2011. Not for All Markets. *Computer* 44(5): 6-8.
- Howe, Jeff. 2008. *Crowdsourcing: How the Power of the Crowd is Driving the Future of Business*, Business Books, Great Britain.
- Howe, Jeff. *The Rise of Crowdsourcing*. 2006. <http://www.wired.com/wired/archive/14.06/crowds.html>

(приступљено 20.12.2012).

Ipeirotis, Panagiotis, Provost, Foster, Wang, Jing. 2010. Quality management on Amazon Mechanical Turk. *Proceedings of the ACM SIGKDD Workshop on Human Computation*: 64-67.

Kazai, Gabriela. 2011. In Search of Quality in crowdsourcing for Search Engine Evaluation. *In Proceedings of the 33rd Proceedings of the European conference on Advances in Information retrieval*: 165-176.

Kittur, Aniket. Crowdsourcing User Studies with Mechanical Turk. 2008. *Computer Human Interaction*, Florence, Italy.

Krippendorff, Klaus. 2011. *Computing Krippendorff's Alpha-Reliability*. Annenberg School for Communication Departmental Papers (ASC), University of Pensilvania.

Krstev, Cvetana, Pavlović-Lažetić, Gordana, Vitas, Duško, Obradović, Ivan. 2004. Using Textual and Lexical Resources in Developing Serbian Wordnet. *Romanian Journal of Information Science and Technology* 7 : 147-161.

Maji, Subhransu. 2011. Large Scale Image Annotations on Amazon Mechanical Turk. Technical report, Electrical Engineering and Computer Sciences, University of California at Berkley.

Митровић, Јелена, Андоновски, Јелена, Софронијевић, Адам. 2010/2011. Иновативне технологије у функцији развоја и промоције дигиталних збирки завичајне грађе: Microtasks Platform и Yahoo Pipes. *Гласник Народне библиотеке Србије*.

Митровић, Јелена. 2013. Пројекти групне расподеле посла (краудсорсинг) у области културе настали захваљујући САРТСНА технологији. *Култура: у потрази за новом парадигмом*. Београд: Филолошки факултет.

Maslow, Abraham, Harold. 1943. A Theory of Human Motivation. *Psychological review* 50.

Munro, Robert, Bethard, Steven, Kuperman,

Victor, Tzuyin, Lai, Melnick, Robin, Potts, Christopher, Schnoebelen, Tyler, Tily, Harry. 2010. Crowdsourcing and Language Studies: The New Generation of Linguistic Data. *Human Language Technologies: Conference of the North American Chapter of the Association for Computational Linguistics, Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*.

Murugesan, San. Understanding web 2.0. 7 2007. [http://hcotuk.etu.edu.tr/bil554/Understanding\\_web\\_2.pdf](http://hcotuk.etu.edu.tr/bil554/Understanding_web_2.pdf) (приступљено 1.09. 2012).

Oleson, David, Sorokin, Alexander, Laughlin, Greg, Hester, Vaughn, Le, John, Biewald Lukas. 2011. Programmatic Gold: Targeted and Scalable Quality Assurance in Crowdsourcing. *Proceedings of the Third Association for the Advancement of Artificial Intelligence Human Computation Workshop*.

Rumshisky, Anna. 2011. Crowdsourcing Word Sense Definition. *Linguistic Annotation Workshop*: 74-81.

Sarasua, Christina, Simperl, Elena, Noy, Natalya F. 2012. Crowdmap: Crowdsourcing Ontology Alignment with Microtasks. *11th International Semantic Web Conference*, Boston, USA.

Schreibman, Susan, Siemens, Ray, Unsworth, John, eds. 2004. A Companion to Digital Humanities. <http://www.digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-2-1&toc.depth=1&toc.id=ss1-2-1&brand=default> (приступљено 15.06.2013)

Stamou Sofia, Oflazer Kemal, Pala Karel, Christodoulakis Dimitris, Cristea Dan, Tufis Dan, Koeva Svetla, Totkov George, Dutoit Dominique, Grigoriadou Maria. 2002. Balkanet: A Multilingual Semantic Network for Balkan Languages. *1st International Global WordNet Conference*. Mysore, India.

Torii, Yoshimitsu, Das, Dipankar, Bandyopadhyay, Sivaji, Okumura, Manabu. 2011. Developing Japanese Wordnet Affect for Analyzing Emotions. *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, ACL-Human Language Technologies*. Portland,

USA.

Vesselinov, Roumen, Grego, John. 2012. Duolingo Effectiveness Report. <http://tcrn.ch/Vshb1C> (приступљено 20.03. 2013).

Von Ahn, L. Dabbish, L. 2008. Designing Games with a Purpose. *Communications of the ACM* 51(8): 57.

Vuković, Maja and Bartolini, Claudio. 2010. Towards a Research Agenda for Enterprise Crowdsourcing. *Leveraging Applications of Formal Methods, Verification, and Validation - 4th International Symposium on Leveraging Applications*. Heraklion, Crete.

Vukovic, Maja, Sundar, Kumara, Greenshpan, Ohad. 2010. Ubiquitous Crowdsourcing. *ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Copenhagen, Denmark.

Примљено: 31. III 2013.

Прихваћено: 21. VI 2013.