

Oznaka poročila: ARRS-RPROJ-ZP-2010-1/31

ZAKLJUČNO POROČILO O REZULTATIH RAZISKOVALNEGA PROJEKTA

A. PODATKI O RAZISKOVALNEM PROJEKTU

1. Osnovni podatki o raziskovalnem projektu

Šifra projekta	J2-9180
Naslov projekta	Jezikoslovno označevanje slovenskega jezika: metode in viri
Vodja projekta	5023 Tomaž Erjavec
Tip projekta	J Temeljni projekt
Obseg raziskovalnih ur	3.150
Cenovni razred	C
Trajanje projekta	01.2007 - 12.2009
Nosilna raziskovalna organizacija	106 Institut "Jožef Stefan"
Raziskovalne organizacije - soizvajalke	581 Univerza v Ljubljani, Filozofska fakulteta
Družbeno-ekonomski cilj	13. Splošni napredek znanja - RiR financiran iz drugih virov (ne iz splošnih univerzitetnih fondov - SUF)

2. Sofinancerji¹

1.	Naziv	
	Naslov	
2.	Naziv	
	Naslov	
3.	Naziv	
	Naslov	

B. REZULTATI IN DOSEŽKI RAZISKOVALNEGA PROJEKTA

3. Poročilo o realizaciji programa raziskovalnega projekta²

Slovenščina je sodoben evropski jezik, ki ima status državnega in uradnega jezika v Republiki Sloveniji in Evropski uniji. Slovenska jezikovna skupnost za sodobno komunikacijsko in socialno učinkovitost potrebuje prožno in učinkovito jezikovno zmožnost. Ta je čedalje bolj odvisna tudi od stopnje informacijske podpore slovenskemu jeziku na področjih, kot so iskanje informacij, strojno (podprto) prevajanje, avtomatska sinteza in prepoznavanje govora, računalniško podprto učenje jezika, avtomatska sumarizacija itd. Za vse te aplikacije pa je predpogoj zmožnost avtomatske jezikovne analize (slovenskega) jezika, saj lahko aplikacije ustrezno obravnavajo

besedilo šele s kontekstno pogojeno informacijo o oblikoslovni kategoriji in lemi (npr. »hotela« kot samostalnik/hotel ali kot glagol/hoteti), skladijski vlogi (npr. osebek ali predmet) in pomenu (npr. hrošč kot žival ali kot napaka v programu).

Tradicionalni računalniški pristopi k jezikoslovnim analizam so temeljili na ročno izdelanih pravilih in leksikonih, vendar pa je bila njihova izdelava zelo draga, večinoma pa tudi niso dosegli zadostne stopnje robustnosti in pokritja, da bi postali resnično uporabni. V zadnjih dveh desetletjih se bolj in bolj uveljavlja drugačna paradigma pri izdelavi takšnih programov, ki temelji na izdelavi (oblikoslovno, skladijsko, semantično) označenih zbirkah besedil, korpusih. Takšni korpusi nato služijo kot učna množica induktivnim programom, ki se iz njih naučijo modela jezika. Ti programi, ki temeljijo na principih strojnega učenja in statističnih analizah, so ne samo bistveno robustnejši, pač pa so tudi lažje prenosljivi med aplikacijskimi domenami, jezikovnimi registri in jeziki. Programi tako postajajo razmeroma jezikovno neodvisni, težišče dela za posamezen jezik pa se prenese na izdelavo kvalitetnih, raznovrstnih in čim večjih označenih korpusov ter, vsaj deloma, iz korpusov izvedenih jezikovnih virov, predvsem (oblikoslovnih, skladijskih in semantičnih) leksikonov. Označeni korpusi dodatno služijo tudi kot testne množice razvitim programom, s čimer postavljajo napredek pri razvoju jezikovnih tehnologij na zdrave, empirične in s tem preverljive temelje, obenem pa koristijo tudi jezikoslovcem, ki se ukvarjajo z empiričnim raziskovanjem (slovenskega) jezika.

Projekt je razvil avtomatske induktivne metode za označevanje oblikoslovja, skladnje in semantike in te metode uporabil pri izdelavi prosto dostopnih jezikovnih virov slovenskega jezika. Označeni korpusi so bili ročno pregledani in popravljeni, s čimer je sedaj zagotovljena prepotrebna infrastruktura za nadaljnji razvoj jezikovnih tehnologij za slovenski jezik. Ta je dostopna ne samo prijaviteljem, pač pa tudi vsem ostalim slovenskim in evropskim raziskovalnim skupinam in lahko tako služi kot katalizator raziskav in razvoja na tem področju, izredno pomembnem za nadaljnjo informatizacijo slovenščine.

Projekt je bil sestavljen iz štirih sklopov. Prvi, horizontalni sklop je obravnaval tehnična in pravna vprašanja dostopnosti, da lahko izdelane vire uporabljajo tako razvijalci, katerim služijo kot podatkovne množice za učenje in testiranje razvitih tehnologij, kot tudi jezikoslovci. Ostali sklopi so obravnavali tri ravni jezikoslovne analize. Oblikoslovje in s tem povezana lematizacija predstavljata osnovni nivo označevanja, potreben pri skoraj vsaki jezikovno usmerjeni aplikaciji; projekt je nadgradil do sedaj razvite metode in pripravil dva označena korpusa. Avtomatska obravnava skladnje je ključnega pomena za poglobljene analize besedil, saj razkriva medsebojne odvisnosti stavčnih členov; projekt je zagotovil prvi večji ročno skladijsko označeni korpus za slovenski jezik. Zadnji sklop je obravnaval leksikalno semantiko slovenskega jezika, ki je potrebna npr. pri strojnem prevajanju in iskanju informacij. Projekt je dogradil obstoječi slovenski semantični leksikon sloWNet in označil korpus s koncepti iz tega leksikona.

V projektu smo v prvem letu pripravili potrebno teoretično in praktično infrastrukturo, oblikoslovno označili prvi korpus ter razvili več programov za oblikoslovno označevanje. Definirali smo specifikacije in nabor oblikoslovnih oznak, ki so dostopne na spletu v izvornem zapisu TEI/XML kot tudi v obliki HTML ter izvedenih tabelaričnih datotekah, ki preslikajo oblikoslovne oznake v več različnih formatov. Razvili smo programsko podporo za ročno označevanje in pregledovanje korpusov, izdelali metodologijo vzorčenja izvornega korpusa FidaPLUS ter uredili pravna vprašanja distribucije korpusa. Iz FidePLUS smo vzorčili dva korpusa, jos100k (100.000 besed) in jos1M (1.000.000 besed). Kot standard zapisa oblikoslovnih specifikacij in korpusov smo izbrali XML shemo, ki je skladna z mednarodnimi priporočili za označevanje besedil za namene znanstvene obravnave Text Encoding Initiative Guidelines P5. Oblikoslovne oznake in leme korpusa jos100k so bile v več fazah ročno pregledane in popravljene (3-kratno navzkrižno pregledovanje), s čimer smo dobili prvi dostopen, kvaliteten, reprezentativen in uravnotežen oblikoslovno označeni korpus slovenskega jezika, ki lahko služi kot referenčni korpus za oblikoslovno označevanje.

V drugem letu smo končali delo na večjem korpusu, jos1M. Ta ni bil v celoti ročno pregledan, pač pa smo razvili metodologijo izločanja tistih avtomatsko pripisanih oznak, ki imajo največjo verjetnost napake; te oznake (270.000) so bile nato ročno pregledane in popravljene. Tako je slovenščina dobila prvi korpus, ki je dovolj velik, da je uporaben kot učna množica za kvalitetne induktivne oblikoslovne označevalnike in lematizatorje. Izvedli smo več eksperimentov s

področja avtomatske lematizacije, kjer smo v različnih režimih delovanja primerjali natančnost in pokritje dveh induktivnih lematizatorjev, CLOG in RDR. Razvili smo dva oblikoslovna označevalnika, prvi, klasični, temelji na HMM (Hidden Markov Models), drugi, s precej večjo fleksibilnostjo izbire značilnik pa na SVM (Support Vector Machines). Pripravili smo prosto dostopen spletni konkordančni nad izdelanima korpusoma. Na osnovi označenih korpusov smo izšolali oblikoslovni označevalnik in lematizator za slovenski jezik ter v sklopu razvitega programa »totale« postavili spletni servis, ki omogoča tokenizacijo, oblikoslovno označevanje ter lematizacijo poljubnih slovenskih besedil. S tem smo omogočili študentom in raziskovalcem slovenistike in jezikovnih tehnologij, da na enostaven način označijo slovenska besedila, ki jih lahko nato uporabijo za nadaljnje raziskave. Začeli smo tudi z delom na skladiščnem in pomenskem označevanju korpusa jos100k.

V tretjem letu smo naredili glavno delo na skladiščno označenem korpusu slovenskega jezika ter korpus tudi pomensko označili. Na področju skladiščnega označevanja smo napisali priročnik za označevalce in označili standardni nabor stavkov, kjer se je, skozi vrsto sestankov, uskladilo načela skladiščnega označevanja v razvitem formalizmu. Razvili smo urejevalnik, ki omogoča skladiščno označevanje korpusa, ter v več fazah ročno označili korpus jos100k. S tem smo za slovenski jezik dobili prvi večjo skladiščno označeni korpus, primeren za šolanje induktivnih skladiščnih označevalnikov. Izvedli smo tudi več poskusov šolanja in testiranja takšnih označevalnikov nad korpusom z zadovoljivimi rezultati. Korpus smo označili tudi pomensko. V korpusu jos100k smo izbrali 100 najbolj pogostih samostalnikov (5.500 pojavitev), in jih označili s pomeni iz slovenskega semantičnega leksikona sloWNet. Napisali smo priročnik za označevalce in razvili sistem, skozi katerega so študentje lahko pomensko označevali te samostalnike oz. fraze, v katerih se pojavljajo. Označevanje je potekalo v več korakih, vzporedno z označevanjem korpusa pa je potekalo tudi popraviljanje in razširjanje sloWNeta. Ker je bilo to prvič, da so bila slovenska besedila semantično označevana, smo morali rešiti tudi razna metodološka vprašanja vezana na semantične relacije v slovenskih samostalnikih. Končni rezultat je prvi pomensko označeni korpus za slovenski jezik, ki predstavlja dragocen vir kot testna oz. učna množica za programe pomenskega razdvoumljanja besed.

Glavni rezultati projekta so torej:

- oblikoskladiščne specifikacije za slovenski jezik
- oblikoskladiščno, skladiščno in pomensko označeni korpus jos100k
- oblikoskladiščno označeni korpus jos1M
- spletni servis za oblikoskladiščno označevanje slovenskih besedil

Oba korpusa sta zapisana v skladu z mednarodnimi priporočili (TEI P5, MULTTEXT-East) ter v več izvedenih formatih. Oba korpusa smo naredili dostopna tako za pregledovanje preko spletnega konkordančnika kot tudi za prenos, pod licenco Creative Commons. S tem smo za slovenski jezik dobili prva dostopna, kvalitetna, reprezentativna in uravnotežena jezikoslovno označena korpusa.

Rezultati projekta so bili uporabljeni pri visokošolskem pouku (do- in podiplomskem ter kot osnova za magistrske naloge) slovenistike in prevajalstva na Univerzi v Novi Gorici in v Ljubljani in pri raziskavah na Filozofski fakulteti Univerze v Ljubljani in Mariboru. Projekt je intenzivno sodeloval s projektom Evropskih strukturnih skladov »Sporazumevanje v slovenskem jeziku«, ki je kot osnovo za izdelavo svojega učnega korpusa vzel oblikoskladiščna priporočila in korpus JOS, skupna pa so tudi priporočila za skladiščno označevanje. Modeli oblikoskladiščnega označevanja in lematizacije JOS so bili uporabljeni v več projektih: za poboljšanje sinteze govora v projektih CRP "VoiceTRAN II - Večjezični prenosni govorni komunikator za bojevnika 21. stoletja" in projektu e-vsebin »BMT: bralnik besedil za mobilne telefone«, za izdelavo slovenskega dela vzporednega korpusa projekta EU SEE-ERA.NET "Building Language Resources and Translation Models for Machine Translation focused on South Slavic and Balkan Languages", kjer je korpus omogočil izšolanje prevajalnega modela med slovenskim in štirimi drugimi jeziki (angleščina, grščina, romunščina, bolgarščina) in v projektu EU 6FP SMART "Statistical Multilingual Analysis for Retrieval and Translation". Metodologija priporočil JOS, ki se povezujejo z večjezičnimi priporočili za oblikoskladiščno označevanje MULTTEXT-East, je bila sprejeta kot ena osrednjih tem projekta EU Mondilex »Conceptual Modelling of Networking of Centres for High-Quality Research in Slavic Lexicography and their Digital

Resource«, kjer je služila kot metodološka osnova za izdelavo priporočil za slovaški, bolgarski, poljski in ukrajinski jezik.

4. Ocena stopnje realizacije zastavljenih raziskovalnih ciljev³

Hipoteza projekta je bila, da je za jezikovnotehnološko podporo slovenskemu jeziku potrebno zagotoviti kvalitetne, ročno jezikoslovno označene korpusne slovenščine, ki morajo biti široko uporabni, torej standardizirani, dobro dokumentirani in dostopni. Projekt je zagotovil takšne korpusne, prve potrditve hipoteze pa so se pokazale že med trajanjem projekta. Na osnovi korpusov smo lahko izšolali oblikoskladenjski in skladdenjski označevalnik, ki imata bistveno boljše natančnost kot prejšnji poskusi za slovenščino. Omogočili smo enostavno uporabo oblikoskladenjskega označevalnika skozi spletni servis, ki se že intenzivno uporablja za izdelavo označenih korpusov v pedagoškem procesu in za raziskave slovenskega jezika, v zadnjem času pa tudi prihajajo prošnje za komercialno uporabo. Označevalnik smo tudi sami uspešno uporabili v več slovenskih in evropskih projektih.

Velik vpliv imajo tudi izdelane oblikoskladenjske specifikacije, saj so postale osnova za označene korpusne v projektu »Sporazumevanje v slovenskem jeziku«, kot tudi za naslednjo generacijo slovenskega referenčnega korpusa FidaPLUS; poleg tega so služile kot osnova za večjezične specifikacije MULTEXT-East v sklopu projekta MONDILEX. Odmevnost rezultatov je zaenkrat večinoma omejena na oblikoskladenjski nivo označevanja, saj so ti rezultati projekta že dve leti na voljo. Večjo odmevnost skladdenjskega in semantičnega označevanja bomo verjetno doživeli v naslednjem letu, skozi dostopnost in publikacije dodatno označenega korpusa jos100k, se pa razviti sistem skladdenjskega označevanja že uporablja v projektu »Sporazumevanje v slovenskem jeziku«.

5. Utemeljitev morebitnih sprememb programa raziskovalnega projekta⁴

--

6. Najpomembnejši znanstveni rezultati projektne skupine⁵

Znanstveni rezultat		
1.	Naslov	<i>SLO</i> Strojno učenje lematizatorjev
		<i>ANG</i> Machine learning of lemmatisers
	Opis	<i>SLO</i> Lemmatizacija je ena od osnovnih jezikovnotehnoloških komponent. V tem znanstvenem članku predstavimo metodo, s katero se lahko skozi strojno učenje nad oblikoslovnimi leksikoni nauči kvalitetnih modelov za lematizacijo poljubnega jezika in pokažemo njene prednosti glede na poprej razvite metode.
		<i>ANG</i> Lemmatisation is one of the basic language technology components. In this paper we present a supervised machine learning method that learns lemmatisation models from morphological lexica. We show its advantages over previously developed methods.
	Objavljeno v	PLISSON, Joël, LAVRAČ, Nada, MLADENIĆ, Dunja, ERJAVEC, Tomaž. Ripple down rule learning for automated word lemmatisation. AI commun., 2008, vol. 21, no. 1, str. 15-26.
	Tipologija	1.01 Izvirni znanstveni članek
COBISS.SI-ID	21593383	
2.	Naslov	<i>SLO</i> Oblikoslovno označevanje slovenskega jezika z metaoznačevalnikom
		<i>ANG</i> Morphosyntactic tagging of Slovene with a meta-tagger
	Opis	<i>SLO</i> Oblikoslovno označevanje je ena od osnovnih jezikovnotehnoloških komponent. V članku predstavimo metodo, s katero je moč poboljšati rezultate oblikoslovnega označevanja skozi kombinacijo oblikoslovnih označevalnikov.
		Morphosyntactic tagging is one of the basic language technology

		ANG	components. In this paper we introduce a method that enables increasing the accuracy of morphosyntactic tagging by combining the outputs of multiple taggers.
	Objavljeno v		RUPNIK, Jan, GRČAR, Miha, ERJAVEC, Tomaž. Improving morphosyntactic tagging of Slovene language through meta-tagging. <i>Informatica</i> (Ljublj.), 2008, vol. 32, no. 4, str. 437-444.
	Tipologija		1.01 Izvirni znanstveni članek
	COBISS.SI-ID		22416423
3.	Naslov	SLO	Oblikoslovno označeni korpus jos100k
		ANG	Morphosyntactically tagged corpus jos100k
	Opis	SLO	Članek predstavi prvo, samo oblikoslovno označeno različico korpusa jos100k.
		ANG	The paper introduces the first version of the jos100k corpus, linguistically annotated only on the morphosyntactic level.
	Objavljeno v		ERJAVEC, Tomaž, KREK, Simon. The JOS morphosyntactically tagged corpus of Slovene. V: 6th International Conference on Language Resources and Evaluation, Marrakech, Morocco, May 26 - June 1, 2008. LREC 2008 : proceedings. Marrakech: ELRA, 2008, 5 str.
	Tipologija		1.08 Objavljeni znanstveni prispevek na konferenci
	COBISS.SI-ID		21930023
4.	Naslov	SLO	Avtomatizacija izdelave slovenskega semantičnega leksikona
		ANG	Automating the creation of the Slovene semantic lexicon
	Opis	SLO	Izdelava semantičnih leksikonov je zamuden in drag proces. Članek predlaga metodo, pri kateri iz odprtih jezikovnih virov, kot npr. Wikipedia, avtomatsko dopolnjujemo semantične leksikone za poljubne jezike.
		ANG	Developing semantic lexica is a time consuming and expensive task. The paper develops a method where open source language resources, such as Wikipedia, can be used to automatically extend language wordnets.
	Objavljeno v		FIŠER, Darja, SAGOT, Benoît. Combining multiple resources to build reliable wordnets. <i>Lect. notes comput. sci.</i> , str. [61]-68.
	Tipologija		1.01 Izvirni znanstveni članek
	COBISS.SI-ID		40118626
5.	Naslov	SLO	Semantično označevanje slovenskega jezika
		ANG	Semantic annotation of Slovene
	Opis	SLO	Predstavitev procesa in rezultatov ročnega označevanja besednih pomenov v korpusu jos100k.
		ANG	The paper discusses the process and results of manual semantic annotation of the jos100k corpus.
	Objavljeno v		FIŠER, Darja, ERJAVEC, Tomaž. sloWnet: construction and corpus annotation. V: BHATTACHARYYA, Pushpak (ur.). Proceedings of The 5th Global WorldNet Conference, GWC 2010, 31st Jan.-4th Feb. 2010, Bombay, India. Principles, construction and application of multilingual wordnets. New Delhi [etc.]: Narosa Publishing House, cop. 2010, str. 177-183, ilustr.
	Tipologija		1.08 Objavljeni znanstveni prispevek na konferenci
	COBISS.SI-ID		42066018

7. Najpomembnejši družbeno-ekonomsko relevantni rezultati projektne skupine⁶

	Družbeno-ekonomsko relevantni rezultat		
1.	Naslov	SLO	Jezikovni viri JOS: http://nl.ijs.si/jos/
		ANG	The JOS language resources: http://nl.ijs.si/jos/
	Opis	SLO	Domača stran projekta JOS ponuja dostop do priporočil za oblikoskladenjsko označevanje in oblikoskladenjsko, skladenjsko in semantično označenih korpusov JOS, zapisanih po mednarodnih standardih in prosto dostopnih pod

		licenco CC.
	ANG	The home page of the JOS project offers access to the specifications for morphosyntactic annotation and the morphosyntactically, syntactically and semantically annotated JOS corpora, which are encoded according to international recommendations and freely available under the Creative Commons licence.
Šifra		F.15 Razvoj novega informacijskega sistema/podatkovnih baz
Objavljeno v		ERJAVEC, Tomaž, KREK, Simon. Oblikoskladenjske specifikacije in označeni korpusi JOS. V: ERJAVEC, Tomaž (ur.), ŽGANEC GROS, Jerneja (ur.). Zbornik Šeste konference Jezikovne tehnologije, 16. do 17. oktober 2008, : zbornik 11. mednarodne multikonference Informacijska družba - IS 2008, zvezek C : proceedings of the 11th International Multiconference Information Society - IS 2008, volume C, (Informacijska družba). Ljubljana: Institut Jožef Stefan, 2008, str. 49-53.
Tipologija		1.08 Objavljeni znanstveni prispevek na konferenci
COBISS.SI-ID		22123559
2.	Naslov	SLO Uporaba rezultatov projekta v pedagoške namene
	ANG	Use of project results in university curricula
	Opis	SLO Rezultati projekta, predvsem označena korpusa in spletni konkordančni in servis za označevanje slovenskih besedil se že uporabljajo pri visokošolskem izobraževanju na več slovenskih univerzah. Izšel je tudi učbenik, ki rezultate projekta predstavi in uporablja pri praktičnih nalogah iz korpusne analize.
	ANG	Project results, esp. the annotated corpora and web concordancer and annotation service for Slovene texts are already used in university education at several Slovene universities. The recently published textbook uses the results of the project for practical exercises in corpus analysis.
Šifra		D.10 Pedagoško delo
Objavljeno v		GORJANC, Vojko, FIŠER, Darja. Korpusna analiza. 1. natis. Ljubljana: Znanstvena založba Filozofske fakultete, 2010. 79 str., ilustr. ISBN 978-961-237-344-3.
Tipologija		2.03 Univerzitetni ali visokošolski učbenik z recenzijo
COBISS.SI-ID		249463296
3.	Naslov	SLO Organizacija in predsedovanje programskemu odboru Šeste konference Jezikovne tehnologije
	ANG	Organisational and programme chair of the Sixth Conference on Language Technologies
	Opis	SLO V sklopu meta-konference Informacijska družba - IS 2008, ki je potekala od 13. do 17. oktobra 2008 na Institutu "Jožef Stefan", je Slovensko društvo za jezikovne tehnologije organiziralo konferenco o jezikovnih tehnologijah. To srečanje je bilo šesto v vrsti konferenc o jezikovnih tehnologijah in organizirano kot mednarodna konferenca z mednarodnim programskim odborom. Uradna jezika konferenca sta bila slovenski in angleški. Sprejeti prispevki so bili objavljeni v tiskanem zborniku in na spletnih straneh konference, na http://nl.ijs.si/is-ltc08/index-sl.html
	ANG	The Slovenian Language Technologies Society held its biennial conference in the scope of the Information Society - IS 2008, October 15 - 17, 2008 at the Jožef Stefan Institute in Ljubljana, Slovenia. This event was the sixth in the series of Slovenian Language Technologies Conferences ('98, '00, '02, '04, '06), with the official languages of the conference being both English and Slovene. Accepted papers were published in printed proceedings, as well as on-line, c.f. http://nl.ijs.si/is-ltc08/index-en.html
Šifra		B.02 Predsedovanje programskemu odboru konference
Objavljeno v		ERJAVEC, Tomaž (ur.), ŽGANEC GROS, Jerneja (ur.). Zbornik Šeste konference Jezikovne tehnologije, 16. do 17. oktober 2008, Ljubljana, Slovenija : zbornik 11. mednarodne multikonference Informacijska družba - IS 2008, zvezek C. Ljubljana: Institut "Jožef Stefan", 2008. VI, 123 str., ilustr. ISBN 978-961-264-006-4.
Tipologija		2.32 Zbornik recenziranih znanstvenih prispevkov na domači konferenci
COBISS.SI-ID		241520896
		Predstavitev rezultatov skladdenjskega označevanja slovenskim raziskovalcem

4.	Naslov	SLO	s področja humanistike
		ANG	Presentation of project results in the area of syntactic annotation to Slovene humanities researchers
	Opis	SLO	Področje skladišnega označevanja korpusov je v Sloveniji povsem novo področje; na osrednji konferenci za slovenske jezikoslovce smo predstavili projektne rezultate na tem področju.
		ANG	The field of syntactic annotation of corpora is completely new in Slovenia; at the main conference for Slovenian linguists we presented the project results in this area.
	Šifra	D.04 Pobuda za uvedbo novega raziskovalnega področja v Sloveniji	
	Objavljeno v	LEDINEK, Nina, ERJAVEC, Tomaž. Odvisnostno površinskoskladišsko označevanje slovenščine: specifikacije in označeni korpusi. V: STABEJ, Marko (ur.). Infrastruktura slovenščine in slovenistike, (Obdobja, Simpozij, = Symposium, 28). 1. natis. Ljubljana: Znanstvena založba Filozofske fakultete, 2009, str. 219-224.	
	Tipologija	1.08 Objavljeni znanstveni prispevek na konferenci	
COBISS.SI-ID	30665261		
5.	Naslov	SLO	Predstavitev rezultatov semantičnega označevanja slovenskim raziskovalcem s področja humanistike
		ANG	Presentation of project results in the area of semantic annotation to Slovene humanities researchers
	Opis	SLO	Področje semantičnega označevanja korpusov je v Sloveniji povsem novo področje; na osrednji konferenci za slovenske jezikoslovce smo predstavili projektne rezultate na tem področju.
		ANG	The field of semantic annotation of corpora is completely new in Slovenia; at the main conference for Slovenian linguists we presented the project results in this area.
	Šifra	D.04 Pobuda za uvedbo novega raziskovalnega področja v Sloveniji	
	Objavljeno v	FIŠER, Darja. sloWNET - slovenski semantični leksikon. V: STABEJ, Marko (ur.). Infrastruktura slovenščine in slovenistike, (Obdobja, Simpozij, = Symposium, 28). 1. natis. Ljubljana: Znanstvena založba Filozofske fakultete, 2009, str. 145-149, graf. prikazi.	
	Tipologija	1.08 Objavljeni znanstveni prispevek na konferenci	
COBISS.SI-ID	41359714		

8. Drugi pomembni rezultati projektne skupine⁷

<p>Poletna šola ESSLLI 2007 Predsedovanje programskemu odboru: Tomaž Erjavec</p> <p>Avgusta 2007 je v Dublinu potekala dvotedenska poletna šola "19th European Summer School in Logic, Language and Information", kjer je sodelovalo več kot 300 študentov in 50 predavateljev. Postopek izbire predmetov poteka na podoben način kot pri konferencah, preko prijave predmetov-predavateljev, in, s pomočjo programskega odbora, večstopenjskim procesom selekcije kvalitetnega in uravnoveženega programa. https://www.cs.tcd.ie/esslli2007/</p> <p>Posvetovanje "Čas za vire se izteka.." Organizator znanstvenega srečanja: Simon Krek Oktobra 2007 je bil v Mariboru v sklopu Slovenskega društva za jezikovne tehnologije (eden od organizatorjev član projektne skupine Simon Krek) posvet o jezikovnih virih za slovenski jezik, ki se ga je udeležilo preko 30 razpravljavcev, poslušalcev in tudi predstavnikov medijev. Prispevki in sklepi srečanja so objavljeni na spletnih straneh društva SDJT. Sklepi potrjujejo usmeritev projekta, saj se zavzemajo za javno dostopnost v dokumentirani in razvidni oblik vseh javno financiranih virov jezika. http://www.sdjt.si/dogodki/MB2007/sdjtMB07.htm</p>

9. Pomen raziskovalnih rezultatov projektne skupine⁸

9.1. Pomen za razvoj znanosti⁹

SLO

Moduli in tehnologije, razviti v projektu, postavljajo slovenski jezik v razred jezikov z razvito osnovno računalniško-jezikovno infrastrukturo, kar bo omogočilo nadaljnje raziskave z besedili v slovenskem jeziku v Sloveniji in zunaj nje. Projekt spada v znanstveno področje računalniškega jezikoslovja, kjer je prispeval k naslednjim področjem:

Razvoj metod za strojno učenje jezikoslovnih modelov: za razvoj tehnologij, ki služil izdelavi modulov znotraj projekta, smo uporabili nekatere najsodobnejše metode za analizo nestrukturiranih in delno strukturiranih podatkov – te metode zajemamo predvsem s področja strojnega učenja, ki je v zadnjih letih naredilo pomembne korake v tej smeri. Zaradi specifik slovenskega jezika mnogo že razvitih metod ni uporabnih brez ustreznih adaptacij – v projektu smo te adaptacije izvajali, evalvirali in uporabljali za končni izdelek. Projekt je tako do sedaj zagotovil napredek pri razvoju stohastičnih metod razdvoumljanja oblikoslovja in induktivnega logičnega programiranja in drugih metod strojnega učenja za namene lematizacije. Dodaten napredek smo dosegli pri kombinaciji raznih metod za doseganje boljše točnosti in pokritja razvitih označevalnikov.

Razvoj empirično osnovanih analiz slovenski jezika: jezikoslovje v Sloveniji je v veliki meri še vedno vpeto v strukturalistično/generativno paradigmo, ki temelji na introspekciji in »umetnih« primerih in njihovi analizi. Projekt ponuja sodobnejše alternative, kjer so analizirani primeri vzeti iz dejanskega jezika, s čimer bo pomagal pri razvoju sodobnega, empirično podprtega jezikoslovja.

Razvoj na področju zapisa in standardizacije jezikoslovnih podatkov: glede na vse večjo kompleksnost dodanih analitičnih oznak v korpuse postaja področje besedišča, zapisa in kombiniranja oznak v svetu aktivno raziskovalno področje. Projekt v izdelanih virih kombinira več ravni oznak in mora omogočiti orodjem, da s temi oznakami operirajo in ponuditi korpuse, ki jih vsebujejo. Zato smo pri projektu upoštevali in nadgradili obstoječe standarde in priporočila (predvsem TEI P5 in MULTEXT-East), kar predstavlja znanstveno zanimive rezultate.

ANG

The modules and technology developed in this project position the Slovene language in the family of languages with at least a basic computerised language infrastructure. This enables further research on Slovene texts, in Slovenia as well as in a wider European context. The project belongs to the scientific discipline of computational linguistics, where it advances the state of the art in the following fields:

Development of methods for machine learning of language models: for the development of technology, which serves in the production of the project software modules we used some of the state-of-the-art methods for the analysis of unstructured and partially structured data – these methods have been taken primarily from the field of machine learning, which has made significant strides in this direction in the last few years. Due to the specificity of Slovene (esp. compared to English) many of the existing methods have not been useful without adaptations. In the course of the project we developed these adaptations, evaluated them and used them for the final software modules. The project has developed machine learning methods for disambiguation of word-level morphosyntactic tags and for the purpose of lemmatisation. Additional advances have been made in the combination of various learning methods in order to achieve better accuracy of taggers.

Research on empirically-grounded linguistic analyses of several levels of the Slovene language: linguistics in Slovenia is, to a large extent, still bound to the generative paradigm, which is based on introspection and »artificial« examples used in the analyses. The project offer alternatives, where the examples are taken from actual language, so it supports the development of contemporary, empirically based linguistics.

Development in the area of encoding and standardisation of linguistic data: given the growing complexity of analytical annotations added to corpora, the area of annotated vocabularies, coding and annotation combination has been attracting increasing interest. The resources developed in this project combine three levels of linguistic annotation; the project had to ensure that the tools can operate on these annotations and offered standardised corpora that contain them. Therefore it was imperative to take into account international standards and recommendations in these areas. The project also recommended new solutions (esp. TEI P5 and MULTEXT-East based), which represent a scientific advance in this area.

9.2. Pomen za razvoj Slovenije¹⁰

SLO

Tako kot je bilo nekoč pomembno imeti lastno knjigo, kasneje lastne časopise, še kasneje lastne elektronske medije, kot so radio, televizija in internet, je dandanes nujno imeti računalniško podprt lastni jezik. Zaradi specifik jezika in kulture tako nalogo lahko izvedejo le pripadniki naroda samega. Razvoj ustrezne računalniške jezikovne infrastrukture je predpogoj za pripadnost jezika družini jezikov, ki že razvijajo poglobljene ravni obravnave onstran leksikalnih in skladenjskih nivojev. Lahko bi rekli, da eden od možnih pogledov na pomen jezika v globalnem smislu postaja njegova dostopnost in povezljivost z drugimi jeziki. Brez široko dostopnih rezultatov, kot smo jih zagotovili v okviru projekta, bi slovenski jezik tako povezljivost težko dosegel.

Izdelani viri bodo tudi pripomogli k ohranjanju kulturne dediščine, saj s pomočjo tovrstne računalniške jezikovne tehnologije postajajo materiali, ki opredeljujejo jezikovno dediščino precej bližji in dostopnejši, kot bi bili sicer.

Ključni del projekta je bilo maksimizirati odmevnost rezultatov s tem, da so vsi izdelani viri slovenskega jezika prosto dostopni. Predvideni uporabniki teh virov so:

- neposredno slovenisti in drugi jezikoslovci, ki so dobili možnost označevanje svojih besedil, internetnega pregledovanja in analize izdelanih virov, pa tudi prenosa celotnih zbirk in obdelave z lastnimi analitičnimi orodji,
- neposredno razvijalci jezikovnih tehnologij, ki lahko obdelujejo besedila v slovenskem jeziku na podobnem tehnološkem nivoju, kot je to mogoče za besedila v angleškem in drugih »velikih« jeziki; uporaba razvitih tehnologij tako omogoča slovenskim akademskim in komercialnim uporabnikom vključevanje v projekte in sodelovanja na globalnem nivoju z lastnimi prispevki, ki podpirajo delo s slovenskim jezikom,
- posredno vsi »uporabniki« slovenskega jezika, saj bodo rezultati projekta spodbudili napredek jezikovnih tehnologij in s tem povezanih aplikativnih programov, kot so iskanje informacij, strojno prevajanje, sinteza in analiza govora itd.

ANG

Just as it used to be important to have books written in one's own language, then newspapers, and later electronic media such as radio, television and Internet, it is today imperative to have computer support for a language. Due to the specifics of languages and cultures, this task can only be accomplished by native speakers. The development of sufficient computational infrastructure is a prerequisite for a language to belong to the family of languages which are already developing methods of analysis that go beyond the lexical and syntactic levels. It could be said that one of the possible views on the importance of a language on a global scale is its accessibility and connectedness with other languages. Without widely available results, such as which we have ensured in the project, the Slovene language will have difficulties in attaining this connection.

The developed resources will also help in preserving cultural heritage, as the foreseen language technology development ensures that the materials which define the language heritage of Slovene become much closer and more accessible to the general public than they would be otherwise.

A key part of the project is to maximise the impact of its results, by making all the developed Slovene language resources freely available. The foreseen users of these resources are:

- The developers of language technologies, as they are able to process texts in Slovene at a technological level similar to that available for other, »larger« languages. The use of developed technologies will enable Slovene academic and commercial partners to participate in projects and global cooperation with their own contributions that support work with the Slovene language.
- Linguists, esp. those studying the Slovene language, who are now able to annotate their own texts, analyse the developed resources via Internet tools, as well as having the option of downloading the complete dataset for research with their own analytic tools.
- Indirectly, all »users« of the Slovene language, as the project results stimulate the development of language technologies for Slovene, and hence the development of directly usable applications, such as information retrieval, machine translation, speech synthesis and analysis, etc.

10. Samo za aplikativne projekte!

Označite, katerega od navedenih ciljev ste si zastavili pri aplikativnem projektu, katere konkretne rezultate ste dosegli in v kakšni meri so doseženi rezultati uporabljeni

Cilj	
F.01	Pridobitev novih praktičnih znanj, informacij in veščin

	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.02	Pridobitev novih znanstvenih spoznanj	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.03	Večja usposobljenost raziskovalno-razvojnega osebja	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.04	Dvig tehnološke ravni	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.05	Sposobnost za začetek novega tehnološkega razvoja	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.06	Razvoj novega izdelka	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.07	Izboljšanje obstoječega izdelka	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.08	Razvoj in izdelava prototipa	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.09	Razvoj novega tehnološkega procesa oz. tehnologije	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.10	Izboljšanje obstoječega tehnološkega procesa oz. tehnologije	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>

	Uporaba rezultatov	<input type="text"/>
F.11	Razvoj nove storitve	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.12	Izboljšanje obstoječe storitve	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.13	Razvoj novih proizvodnih metod in instrumentov oz. proizvodnih procesov	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.14	Izboljšanje obstoječih proizvodnih metod in instrumentov oz. proizvodnih procesov	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.15	Razvoj novega informacijskega sistema/podatkovnih baz	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.16	Izboljšanje obstoječega informacijskega sistema/podatkovnih baz	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.17	Prenos obstoječih tehnologij, znanj, metod in postopkov v prakso	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.18	Posredovanje novih znanj neposrednim uporabnikom (seminarji, forumi, konference)	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.19	Znanje, ki vodi k ustanovitvi novega podjetja ("spin off")	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>

F.20	Ustanovitev novega podjetja ("spin off")	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.21	Razvoj novih zdravstvenih/diagnostičnih metod/postopkov	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.22	Izboljšanje obstoječih zdravstvenih/diagnostičnih metod/postopkov	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.23	Razvoj novih sistemskih, normativnih, programskih in metodoloških rešitev	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.24	Izboljšanje obstoječih sistemskih, normativnih, programskih in metodoloških rešitev	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.25	Razvoj novih organizacijskih in upravljavskih rešitev	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.26	Izboljšanje obstoječih organizacijskih in upravljavskih rešitev	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.27	Prispevek k ohranjanju/varovanje naravne in kulturne dediščine	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.28	Priprava/organizacija razstave	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.29	Prispevek k razvoju nacionalne kulturne identitete	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE

	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.30	Strokovna ocena stanja	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.31	Razvoj standardov	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.32	Mednarodni patent	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.33	Patent v Sloveniji	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.34	Svetovalna dejavnost	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.35	Drugo	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>

Komentar

11. Samo za aplikativne projekte!

Označite potencialne vplive oziroma učinke vaših rezultatov na navedena področja

	Vpliv	Ni vpliva	Majhen vpliv	Srednji vpliv	Velik vpliv	
G.01	Razvoj visoko-šolskega izobraževanja					
G.01.01.	Razvoj dodiplomskega izobraževanja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.01.02.	Razvoj podiplomskega izobraževanja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.01.03.	Drugo:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02	Gospodarski razvoj					

G.02.01.	Razširitev ponudbe novih izdelkov/storitev na trgu	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.02.	Širitev obstoječih trgov	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.03.	Znižanje stroškov proizvodnje	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.04.	Zmanjšanje porabe materialov in energije	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.05.	Razširitev področja dejavnosti	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.06.	Večja konkurenčna sposobnost	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.07.	Večji delež izvoza	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.08.	Povečanje dobička	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.09.	Nova delovna mesta	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.10.	Dvig izobrazbene strukture zaposlenih	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.11.	Nov investicijski zagon	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.12.	Drugo:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.03	Tehnološki razvoj					
G.03.01.	Tehnološka razširitev/posodobitev dejavnosti	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.03.02.	Tehnološko prestrukturiranje dejavnosti	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.03.03.	Uvajanje novih tehnologij	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.03.04.	Drugo:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.04	Družbeni razvoj					
G.04.01	Dvig kvalitete življenja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.04.02.	Izboljšanje vodenja in upravljanja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.04.03.	Izboljšanje delovanja administracije in javne uprave	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.04.04.	Razvoj socialnih dejavnosti	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.04.05.	Razvoj civilne družbe	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.04.06.	Drugo:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.05.	Ohranjanje in razvoj nacionalne naravne in kulturne dediščine in identitete	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.06.	Varovanje okolja in trajnostni razvoj	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.07	Razvoj družbene infrastrukture					
G.07.01.	Informacijsko-komunikacijska infrastruktura	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.07.02.	Prometna infrastruktura	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.07.03.	Energetska infrastruktura	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.07.04.	Drugo:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.08.	Varovanje zdravja in razvoj zdravstvenega varstva	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.09.	Drugo:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

Komentar

--

12. Pomen raziskovanja za sofinancerje, navedene v 2. točki¹¹

1.	Sofinancer		
	Vrednost sofinanciranja za celotno obdobje trajanja projekta je znašala:		EUR
	Odstotek od utemeljenih stroškov projekta:		%
	Najpomembnejši rezultati raziskovanja za sofinancerja		Šifra
	1.		
	2.		
	3.		
	4.		
	5.		
	Komentar		
Ocena			
2.	Sofinancer		
	Vrednost sofinanciranja za celotno obdobje trajanja projekta je znašala:		EUR
	Odstotek od utemeljenih stroškov projekta:		%
	Najpomembnejši rezultati raziskovanja za sofinancerja		Šifra
	1.		
	2.		
	3.		
	4.		
	5.		
	Komentar		
Ocena			
3.	Sofinancer		
	Vrednost sofinanciranja za celotno obdobje trajanja projekta je znašala:		EUR
	Odstotek od utemeljenih stroškov projekta:		%
	Najpomembnejši rezultati raziskovanja za sofinancerja		Šifra
	1.		

	2.		
	3.		
	4.		
	5.		
	Komentar		
	Ocena		

C. IZJAVE

Podpisani izjavljam/o, da:

- so vsi podatki, ki jih navajamo v poročilu, resnični in točni
- se strinjamo z obdelavo podatkov v skladu z zakonodajo o varstvu osebnih podatkov za potrebe ocenjevanja, za objavo 6., 7. in 8. točke na spletni strani <http://sicris.izum.si/> ter obdelavo teh podatkov za evidence ARRS
- so vsi podatki v obrazcu v elektronski obliki identični podatkom v obrazcu v pisni obliki
- so z vsebino zaključnega poročila seznanjeni in se strinjajo vsi soizvajalci projekta

Podpisi:

Tomaž Erjavec	in	
podpis vodje raziskovalnega projekta		zastopnik oz. pooblaščen oseba RO

Kraj in datum:

Ljubljana

15.4.2010

Oznaka poročila: ARRS-RPROJ-ZP-2010-1/31

¹ Samo za aplikativne projekte. [Nazaj](#)

² Napišite kratko vsebinsko poročilo, kjer boste predstavili raziskovalno hipotezo in opis raziskovanja. Navedite ključne ugotovitve, znanstvena spoznanja ter rezultate in učinke raziskovalnega projekta. Največ 18.000 znakov vključno s presledki (približno tri strani, velikosti pisave 11). [Nazaj](#)

³ Realizacija raziskovalne hipoteze. Največ 3.000 znakov vključno s presledki (približno pol strani, velikosti pisave 11). [Nazaj](#)

⁴ Samo v primeru bistvenih odstopanj in sprememb od predvidenega programa raziskovalnega projekta, kot je bil zapisan v predlogu raziskovalnega projekta. Največ 3.000 znakov vključno s presledki (približno pol strani, velikosti pisave 11). [Nazaj](#)

⁵ Navedite največ pet najpomembnejših znanstvenih rezultatov projektne skupine, ki so nastali v času trajanja projekta v okviru raziskovalnega projekta, ki je predmet poročanja. Za vsak rezultat navedite naslov v slovenskem in angleškem jeziku (največ 150 znakov vključno s presledki), rezultat opišite (največ 600 znakov vključno s presledki) v slovenskem in angleškem jeziku, navedite, kje je objavljen (največ 500 znakov vključno s presledki), izberite ustrezno šifro tipa objave po Tipologiji dokumentov/del za vodenje bibliografij v sistemu COBISS ter napišite ustrezno COBISS.SI-ID številko bibliografske enote. Navedeni rezultati bodo objavljeni na spletni strani <http://sicris.izum.si/>.

PRIMER (v slovenskem jeziku):

Naslov: Regulacija delovanja beta-2 integrinskih receptorjev s katepsinom X;

Opis: Cisteinske proteaze imajo pomembno vlogo pri nastanku in napredovanju raka. Zadnje študije kažejo njihovo povezanost s procesi celičnega signaliziranja in imunskega odziva. V tem znanstvenem članku smo prvi dokazali... (največ 600 znakov vključno s presledki)

Objavljeno v: OBERMAJER, N., PREMŽL, A., ZAVAŠNIK-BERGANT, T., TURK, B., KOS, J.. Carboxypeptidase cathepsin X mediates $\beta 2$ - integrin dependent adhesion of differentiated U-937 cells. Exp. Cell Res., 2006, 312, 2515-2527, JCR IF (2005): 4.148

Tipologija: 1.01 - Izvirni znanstveni članek

COBISS.SI-ID: 1920113 [Nazaj](#)

⁶ Navedite največ pet najpomembnejših družbeno-ekonomsko relevantnih rezultatov projektne skupine, ki so nastali v času trajanja projekta v okviru raziskovalnega projekta, ki je predmet poročanja. Za vsak rezultat navedite naslov (največ 150 znakov vključno s presledki), rezultat opišite (največ 600 znakov vključno s presledki), izberite ustrezen rezultat, ki je v Šifrantu raziskovalnih rezultatov in učinkov (Glej: <http://www.arrs.gov.si/sl/gradivo/sifranti/sif-razisk-rezult.asp>), navedite, kje je rezultat objavljen (največ 500 znakov vključno s presledki), izberite ustrezno šifro tipa objave po Tipologiji dokumentov/del za vodenje bibliografij v sistemu COBISS ter napišite ustrezno COBISS.SI-ID številko bibliografske enote.

Navedeni rezultati bodo objavljeni na spletni strani <http://sicris.izum.si/>. [Nazaj](#)

⁷ Navedite rezultate raziskovalnega projekta v primeru, da katerega od rezultatov ni mogoče navesti v točkah 6 in 7 (npr. ker se ga v sistemu COBISS ne vodi). Največ 2.000 znakov vključno s presledki. [Nazaj](#)

⁸ Pomen raziskovalnih rezultatov za razvoj znanosti in za razvoj Slovenije bo objavljen na spletni strani: <http://sicris.izum.si/> za posamezen projekt, ki je predmet poročanja. [Nazaj](#)

⁹ Največ 4.000 znakov vključno s presledki [Nazaj](#)

¹⁰ Največ 4.000 znakov vključno s presledki [Nazaj](#)

¹¹ Rubrike izpolnite/prepišite skladno z obrazcem "Izjava sofinancerja" (<http://www.arrs.gov.si/sl/progproj/rproj/gradivo/>), ki ga mora izpolniti sofinancer. Podpisan obrazec "Izjava sofinancerja" pridobi in hrani nosilna raziskovalna organizacija – izvajalka projekta. [Nazaj](#)

Obrazec: ARRS-RPROJ-ZP/2010 v1.00a

55-94-3C-E2-90-49-C1-0D-8A-94-A3-B1-2A-90-65-E2-71-E7-5A-F2