

Oznaka poročila: ARRS_ZV_RPROJ_ZP_2008/255

ZAKLJUČNO POROČILO O REZULTATIH RAZISKOVALNEGA PROJEKTA

A. PODATKI O RAZISKOVALNEM PROJEKTU

1. Osnovni podatki o raziskovalnem projektu

Šifra projekta	J2-9742
Naslov projekta	Modeliranje spontanega govora in zelo pregibnih jezikov za podporo vodenim telekomunikacijskim storitvam
Vodja projekta	20032 Andrej Žgank
Tip projekta	J Temeljni projekt
Obseg raziskovalnih ur	2.400
Cenovni razred	B
Trajanje projekta	01.2007 - 12.2008
Nosilna raziskovalna organizacija	796 Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko
Raziskovalne organizacije - soizvajalke	
Družbeno-ekonomski cilj	11 Neusmerjene raziskave (temeljne)

2. Sofinancerji¹

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

B. REZULTATI IN DOSEŽKI RAZISKOVALNEGA PROJEKTA

3. Poročilo o realizaciji programa raziskovalnega projekta²

Prva faza projekta je bila posvečena raziskavam s področja modeliranja spontanega govora za avtomatske razpoznavalnike govora. Osnovno vodilo je bilo izboljšati delovanje sistema avtomatskega razpoznavanja govora za najkompleksnejši tip vhodnega signala, ki ga predstavlja spontan govor. Šele bistvene izboljšave na področju temeljnih raziskav o razpoznavalnikih spontanega govora bodo v prihodnje omogočile njihovo vključitev v sodobne uporabniške vmesnike telekomunikacijskih storitev.

V uvodnem delu smo izvedli analizo pojavljanja posameznih kategorij efektov spontanega govora

v različnih govornih virih. Tukaj smo poseben poudarek dali primerjavi slovenskega spontanega govora v govorni bazi Turdis in bazi BNSI Broadcast News, ki je predstavljala naš temeljni govorni vir. Analizirali smo pogostost pojavljanja posameznih diskurznihih označevalcev, vpliv razlike med žanri, vpliv konteksta besedila v katerem se pojavijo, ter možnosti za njihovo avtomatsko označevanje pri prihodnji izgradnji govornih virov. Izvedli smo tudi primerjavo vpliva validatorjev na ročno označevanje diskurznihih označevalcev, kar je še posebej pomembno s stališča vrednotenja kvalitete avtomatskega označevanja, ki ga potrebujemo za obdelavo večjih govornih baz, kot je na primer baza SloParl s posnetki razprav iz Državnega zbora RS v skupni dolžini 100 ur govornega materiala. Na področju izločanja značilnk smo izvedli primerjalno določitev optimalnega nabora značilnk za predprocesiranje govornega signala. Pomembnost te primerjave leži v dejstvu, da želimo pri detekciji mašil ohraniti razlikovanje med njihovimi akustičnimi, fonetičnimi in prozodičnimi značilnostmi. Uvodna analiza je za primer mašil in onomatopej pokazala, da so razlike med značilnostmi pogosto bistveno manjše od razlik pri ostalih kategorijah govora.

Na področju predprocesiranja pri modeliranju spontanega govora smo se osredotočili na eksplicitno in implicitno detekcijo mašil z Gaussovimi modeli in prikritimi modeli Markova. Vrednotenje rezultatov je pokazalo, da dosežemo boljše rezultate v primeru, kadar mašila in onomatopeje vključujemo v akustično modeliranje, kot pa v primeru, kadar smo takšne dele spontanega govora na osnovi odločitve modula za detekcijo neposredno izločali. Kljub uporabi najobsežnejših govornih virov za slovenski jezik smo imeli težave s pogostostjo pojavljanja mašil v učnem korpusu. Kot rešitev smo vpeljali prenos akustičnih modelov med različnimi govornimi viri, kjer smo ocenjevanje podobnosti izvedli s podatkovno vodeno metriko. Prilagoditev na ciljni govorni vir je bila uspešnejša, ko smo namesto ponovnega učenja modelov izvedli akustično adaptacijo in tako generirali kombinirane modele. Na osnovi uvodoma izpeljane analize lastnosti diskurznihih označevalcev smo predlagali novo metodo za implicitno modeliranje mašil in onomatopej pri razpoznavanju spontanega govora. Posamezne kategorije mašil, ki jih modeliramo so definirane na osnovi fonetičnih razredov. Najprej smo uporabili ekspertno definirane razrede, kasneje pa še razrede definirane na osnovi podatkovno vodene metrike, s čemer smo se izognili potrebi po ekspertnem znanju in jezikovni odvisnosti metode.

V naslednjem koraku smo modelirane efekte spontanega govora vključili v jezikovne modele. Tvorili smo različne osnovne jezikovne modele, ki so implicitno modelirali posamezne kategorije. Končni, interpolirani, jezikovni model smo zgradili na osnovi razvojnega korpusa, ki je vseboval transkripcije televizijskih intervjujev. Pri tem smo zaradi doseženih dobrih rezultatov pri akustičnem modeliranju mašil in onomatopej posvetili posebno pozornost njihovi vlogi v jezikovnem modelu. Najboljše skupne rezultate smo dosegli v primeru, ko smo definirali brezkontekstno modeliranje mašil in onomatopej v jezikovnem modelu. Na tem področju raziskav se je kot posebna slabost pokazalo pomanjkanje večje količine besedilnih virov spontanega govora za slovenski jezik.

V zadnjem koraku modeliranja spontanega govora smo se posvetili fonetičnemu slovarju razpoznavalnika govora. Glede na fonetične lastnosti mašil in onomatopej v slovenskem jeziku, ki smo jih raziskali v okviru analize diskurznihih označevalcev, smo definirali njihove različne transkripcije. Hkrati smo v slovarju razpoznavalnika spontanega govora na osnovi pojavljanja v učnem korpusu definirali še fonetične transkripcije za delno izgovorjene besede. Zaradi pogosto specifičnega načina izgovorjave smo na osnovi ekspertnega znanja definirali še dodatne variante izgovorjav z redukcijami končnih samoglasnikov. Za določene vrste delno izgovorjenih besed se je pokazalo, da lahko vodi večanje števila variant fonetičnih transkripcij v povečanje akustične zamenljivosti. Ker smo iz tega dela vrednotenja želeli izločiti vse ostale možne vplive (variacije med govorci, vpliv domene...) smo to vrednotenje izpeljali z razpoznavalnikom govora za odvisnega govorca. Analiza rezultatov vrednotenja je pokazala, da prevladujoči delež izboljšanja rezultatov avtomatskega razpoznavanja govora še vedno pripada modeliranju mašil na akustičnem nivoju. Da smo lahko vključili vse zasnovane spremembe in predlagane izboljšave, je bilo potrebno prilagoditi tudi dekodirnik razpoznavalnika govora, predvsem s podporo za eksplicitno in implicitno modeliranje mašil in onomatopej. Hkrati s prilagoditvami dekodirnika, ki so bile potrebne za vključitev akustičnih in jezikovnih modelov spontanega govora, smo pripravili tudi različico dekodirnika za razpoznavanje podbesednih enot osnova – končnica, ki je služil kot okvir za nadaljevanje projekta, ki je bilo posvečeno izboljšavam pri modeliranju močno pregibnih

jezikov, med katere tako kot večina drugih slovanskih jezikov sodi tudi slovenščina.

Modeliranje močno pregibnih jezikov smo začeli z definiranjem možnosti kombiniranja različnih govornih virov za slovenski jezik, saj to pogosto predstavlja eno izmed ključnih ovir na področju govornih tehnologij. Vloga kombiniranja govornih virov postane še posebej pomembna v primeru, ko želimo uporabiti kakšen nov govorni vir z nepopolnimi transkripcijami. Za izpeljavo rahlo nadzorovanega učenja akustičnih modelov neobhodno potrebujemo možnost kombiniranja govornih virov, s katerimi smo izpeljali iskanje napak v nepopolnih transkripcijah govorne baze SloParl. Tako pripravljene učne transkripcije so bile osnova za postopek rahlo nadzorovanega učenja akustičnih modelov. S ponavljanji prisilne poravnave smo dodatno izločali učni korpus pri katerem je prihajalo do odstopanj med govornim signalom in transkripcijami. Na osnovi statistike izločenega materiala smo nadzorovali kvaliteto posameznih govorcev in v primeru preseganja definiranega praga začasno v celoti izločili takšne govorce. Naknadno kombinirano vključevanje tako izločenih govorcev v naučene akustične modele ni prineslo dodatnega izboljšanja rezultata, kljub povečanemu obsegu učnega materiala.

Hkrati s postopkom rahlo nadzorovanega in nenadzorovanega učenja akustičnih modelov je potekalo tudi dodatno zbiranje govornega materiala. Tako smo dodatno zbrali 15 ur takšnih posnetkov z nepopolnimi transkripcijami. Količino tako zbranega materiala je možno naknadno tudi relativno preprosto povečati. Trenutno namreč dodatno večanje govornega učnega korpusa z nepopolnimi transkripcijami ni smiselno, saj je kompleksnost učnega postopka zelo velika in bi jo tako samo še dodatno povečali.

Razcepljanje besed v močno pregibnih jezikih na manjši enoti osnova končnica je eden izmed načinov, kako je možno doseči boljše rezultate razpoznavanja govora v takšnih jezikih. Predlagali smo novi algoritem za razcepljanje na osnovo in končnico, ki je podatkovno voden, in kot takšen jezikovno neodvisen. Tako ga je možno uporabiti tudi za druge močno pregibne jezike. Metrika razcepljanja temelji na zmanjševanju entropije učnega korpusa. Rezultat razcepljanja je statistično pomembno zmanjšanje deleža besed izven slovarja, kar posledično vodi tudi k izboljšanju rezultatov razpoznavanja govora. Pri akustičnem modeliranju tako razcepljenih enot govornega korpusa smo na osnovi analize fonetičnih značilnosti spremenili mejno topologijo akustičnih modelov, ter na časovnih mejah dodatno iskali tiste segmente, ki so zaradi svojih značilnosti težavni za akustično modeliranje. V zadnjem koraku akustičnega modeliranja smo uporabili postopek diskriminativnega učenja, kjer smo minimizirali delež napak na kratkih razcepljenih enotah na osnovi postopka prisilne poravnave. Analiza vrednotenja na viru spontanega govora je pokazala, da so postopki razcepljenega učenja za močno pregibne jezike uspešni pri srednji in večji dolžini razcepljenih enot. Pri tistih enotah, ki so kot rezultat razcepljanja zelo kratke, pa je akustično modeliranje še vedno nezanesljivo, saj še vedno prihaja do akustične zamenljivosti v iskalnem prostoru, ki je tudi podpora razcepljenega jezikovnega modela ne odpravi. Verjetna smer raziskovanja, ki bi lahko odpravila te težave je v iskanju kombiniranega pristopa.

Končni rezultat integracije algoritmov predstavljajo demonstracijski posnetki z indeksiranimi avdio-vizualnimi vsebinami na multimedijem aplikacijskem strežniku. Mobilni uporabnik lahko do tako indeksiranih vsebin dostopa s sodobno multimodalno terminalno opremo, ki omogoča širokopasovni prenos podatkov.

4. Ocena stopnje realizacije zastavljenih raziskovalnih ciljev³

Analiza rezultatov razpoznavanja govora je pokazala, da je možno z uporabo modeliranja učinkov spontanega govora izboljšati rezultate razpoznavanja, kar potrjuje raziskovalno hipotezo, ki je bila postavljena pred začetkom projekta. Izpeljana analiza jezikovnih virov spontanega govora je pokazala na bistveni vpliv vrste in žanra ter konteksta na nivo pojavljanja učinkov spontanega govora, kar se je odražalo tudi v uspešnosti modeliranja. Med postopkom modeliranja spontanega govora se je pokazalo, da je za doseganje dobre uspešnosti potrebno predvsem kvalitetno modeliranje mašil in onomatopej na akustičnem nivoju. Tukaj sta bili izpeljana dva pristopa, prvi na osnovi ekspertnega znanja, drugi pa na osnovi podatkovno vodene metrike, kar je v skladu z zastavljenimi cilji o jezikovni neodvisnosti predlaganih metod. Ker smo pri modeliranju spontanega govora posegli tudi po vpeljavi prenosa akustičnih modelov mašil med različnimi

govornimi viri, smo tako delno odpravili težavo s pomanjkanjem primerne učnega materiala. Glede na zahtevnost zbiranja novega učnega materiala se je pokazalo za smiselno dodatno zbiranje učnega materiala z nepopolnimi transkripcijami.

Prav tako so rezultati razpoznavanja govora za modeliranje močno pregibnih jezikov pokazali na statistično pomembno izboljšanje rezultatov, kar potrjuje hipotezo, ki smo jo postavili pred začetkom projekta, da je možno z vpeljavo dodatnih metod, ki niso v uporabi za zahodnoevropske jezike, izboljšati delovanje razpoznavalnika spontanega govora za slovenski jezik. Tukaj je še posebej pomembno vlogo odigralo kombiniranje govornih virov, saj brez njega praktično ne bi mogli realizirati različnih tipov učenja akustičnih modelov na nepopolnih transkripcijah. Pri podbesednem modeliranju visoko pregibnih jezikov smo s predlagano metodo potrdili hipotezo o boljši funkcionalnosti podatkovno vodenega razcepljanja na podenote, kar se je odražalo tudi v statistično pomembnem izboljšanju delovanja razpoznavalnika govora, kar je bil eden izmed zastavljenih ciljev projekta.

Na teh osnovah je možno trditi, da so bili zastavljeni cilji projekta v celoti izpolnjeni.

5. Utemeljitev morebitnih sprememb programa raziskovalnega projekta⁴

Ni sprememb.

6. Najpomembnejši znanstveni rezultati projektne skupine⁵

Znanstveni rezultat			
1.	Naslov	SLO	Analiza vpliva konteksta na diskurzne označevalce v dveh pogovornih žanrih.
		ANG	The impact of context on discourse marker use in two conversational genres.
	Opis	SLO	Analizirane so bili kontekstne lastnosti v formalnem in neformalnem govoru, ki pomembno vplivajo na razlike v uporabi diskurzni označevalcev. Rezultati izvedene analize o pogostosti pojavljanja diskurzni označevalcev ter analize konteksta v katerem se pojavljajo so služili kot izhodišča pri postopkih modeliranja spontanega govora. Razlike med žanri so pokazale na upravičenost kombiniranja baz izgovorjav v postopku generiranja akustičnih modelov. Članek je bil objavljen v interdisciplinarni reviji, ki v skladu z ARRS metodologijo sodi v kategorijo A1".
		ANG	The impact of context in formal and informal conversations on the use of discourse markers was analysed. Several contextual factors, which contribute to the differences in the use of discourse markers, were identified. These results were used as a baseline for the modeling of spontaneous speech. The differences between genres indicated that the combining of various spoken language resources for acoustic modeling generation could be applied as possible solution. The achievement was published in an interdisciplinary journal, which belongs to the A1" category regarding the ARRS methodology.
	Objavljeno v		VERDONIK, Darinka, ŽGANK, Andrej, PISANSKI PETERLIN, Agnes. The impact of context on discourse marker use in two conversational genres. Discourse studies, 2008, vol. 10, no. 6, str. 759-775, JCR IF (2007): 0.929, SSE (20/45), communication, x: 0.847
	Tipologija		1.01 Izvirni znanstveni članek
	COBISS.SI-ID		12612886
2.	Naslov	SLO	Podatkovno vodeno modeliranje podbesednih enot za močno pregibni slovenski jezik.
		ANG	Data-driven modeling of subword units for highly inflective Slovenian language.
	Opis	SLO	V članku smo predstavili rezultate modeliranja slovenskega jezika za uporabo v razpoznavalniku spontanega govora. Slovenski jezik sodi v skupino visoko pregibnih jezikov s kompleksno morfologijo. V članku smo predlagali podatkovno vodeno razcepljanje besed v dve podenoti, tj. osnovo in končnico z uporabo nove metrike. Prednost predlaganega algoritma je

			jezikovna neodvisnost, saj razcep temelji na zmanjševanju entropije učnega korpusa. Eksperimentalno smo pokazali, da predlagani jezikovni modeli zmanjšajo delež besed izven slovarja in tako vodijo k boljšim rezultatom razpoznavanja govora.
		ANG	The paper presents a study on modeling the highly inflective Slovenian language for spontaneous speech recognition. The research focus is on data sparsity, which results from the complex morphology of the language. A new data-driven subword unit based method is proposed for the induction of inflectional morphology modeling. No prior knowledge of the language is used. We are searching for a decomposition which yields the minimum entropy of the training corpus. The experiments demonstrate that proposed models considerably reduce out-of-vocabulary rate and improve speech recognition performance.
	Objavljeno v		SEPEŠY MAUČEC Mirjam, ROTOVNIK Tomaž, KAČIČ Zdravko, BREST Janez. Using data-driven subword units in language model of highly inflective Slovenian language. International Journal of Pattern Recognition and Artificial Intelligence, iss. 2., vol. 23, marec 2009, pp. 287-312. JCR IF (2007): 0.374, SE (83/93), computer science, artificial intelligence.
	Tipologija		1.01 Izvirni znanstveni članek
	COBISS.SI-ID		13118230
3.	Naslov	SLO	Razpoznavanje slovenskega spontanega govora in modeliranje mašil ter onomatopej
		ANG	Slovenian spontaneous speech recognition and acoustic modeling of filled pauses and onomatopoeas
	Opis	SLO	Predlagali smo novo metodo modeliranja mašil in onomatopej za razpoznavanje spontanega govora z velikim slovarjem besed. Akustični modeli so bili zasnovani na osnovi rezultatov analize diskurznihih označevalcev. Pri tem smo upoštevali predvsem njihove akustično-fonetične značilnosti, kjer smo ločeno modelirali onomatopeje. Predlagana metoda je tako zasnovana na osnovi implicitnega modeliranja s fonetičnimi razredi brez upoštevanja konteksta v jezikovnem modelu. Primerjava s tremi drugimi metodami modeliranja je pokazala na statistično pomembno izboljšanje rezultatov razpoznavanja govora.
		ANG	A new method for modeling filled pauses and onomatopoeas for large vocabulary spontaneous speech recognition was proposed. The acoustic models were defined using the results of discourse markers analysis. The major influence was given to their acoustic-phonetic properties, where separated modeling of onomatopoeas was necessary. The proposed method is based on implicit modeling using the phonetic broad classes. The context of these models is ignored in the language model. The comparison with three other modeling methods showed statistically significant improvement of speech recognition results.
	Objavljeno v		ŽGANK, Andrej, ROTOVNIK, Tomaž, SEPEŠY MAUČEC, Mirjam. Slovenian spontaneous speech recognition and acoustic modeling of filled pauses and onomatopoeas. WSEAS trans. signal process., Jul. 2008, vol. 4, iss. 7, str. 388-397.
	Tipologija		1.01 Izvirni znanstveni članek
	COBISS.SI-ID		12706070
4.	Naslov	SLO	Analiza pojavljanja diskurznihih označevalcev v spontanem slovenskem govoru.
		ANG	Analysis of discourse markers frequencies in Slovenian spontaneous speech.
	Opis	SLO	V članku smo opisali rezultate analize pojavljanja diskurznihih označevalcev v slovenskem spontanem govoru. Analiza je bila izvedena na osnovi primerjave med naborom spontanih stavkov iz govornih baze BNSI Broadcast News in Turdis (domena: novice oz. turizem). S stališča modeliranja učinkov spontanega govora v razpoznavalniku govora so posebej pomembni rezultati analize za kategorijo mašil, ter za kategorijo opornih signalov. Primerjava med obema žanroma je pokazala na statistično signifikantne razlike pri pogostosti pojavljanja različnih diskurznihih označevalcev.
			The paper presents the results of analysis of discourse markers in Slovenian spontaneous speech. The analysis was performed comparing sets of spontaneous utterances from the BNSI Broadcast News speech database and Turdis speech database. The first one comprises news shows and the second

		ANG	one conversation from tourist office. For modeling spontaneous speech for automatic speech recognition, the most important categories of results are those for filled pauses and background signals. The results showed a statistically significant difference in discourse markers frequency for both genres.
	Objavljeno v		VERDONIK Darinka, ŽGANK Andrej, PISANSKI PETERLIN Agnes, Diskurzni označevalci v dveh različnih pogovornih žanrih. Jezik in slovstvo, št. 6, let. 52, str. 19-33, nov-dec 2007.
	Tipologija		1.01 Izvirni znanstveni članek
	COBISS.SI-ID		36334434
5.	Naslov	SLO	
		ANG	
	Opis	SLO	
		ANG	
	Objavljeno v		
	Tipologija		
	COBISS.SI-ID		

7. Najpomembnejši družbeno-ekonomsko relevantni rezultati projektne skupine⁶

	Družbeno-ekonomsko relevantni rezultat		
1.	Naslov	SLO	Doprinos k ohranjanju slovenskega jezika v dobi digitalizacije.
		ANG	Preservation of Slovenian language in the era of digitalization.
	Opis	SLO	Raziskave o diskurzni označevalci v dveh različnih pogovornih žanrih slovenskega jezika so pomembne s stališča ohranjanja slovenskega jezika v dobi digitalizacije. Z večanjem vpliva informacijske družbe prihaja do večanja vpliva angleškega jezika, ki je še posebej opazen v povezavi z uporabo različnih telekomunikacijskih storitev. Analize izvedene na osnovi različnih slovenskih govornih virov omogočajo dodatne raziskave na področju avtomatskega razpoznavanja spontanega govora, kar bo posledično imelo vpliv tudi na razvoj sodobnih telekomunikacijskih storitev s podporo za slovenski jezik.
		ANG	The discourse markers analysis for two different genres of spontaneous speech is important for preserving the Slovenian language in the era of digitalization. With the growing influence of information society also the influence of English language grows. This is specifically noticeable for state-of-the-art telecommunication services. The analysis applied to Slovenian speech databases enables new fundamental research in the area of spontaneous speech recognition. This will result in the additional development of state-of-the-art telecommunication services with support for Slovenian language.
	Šifra		F.29 Prispevek k razvoju nacionalne kulturne identitete
	Objavljeno v		VERDONIK Darinka, ŽGANK Andrej, PISANSKI PETERLIN Agnes, Diskurzni označevalci v dveh različnih pogovornih žanrih. Jezik in slovstvo, št. 6, let. 52, str. 19-33, nov-dec 2007.
	Tipologija		1.01 Izvirni znanstveni članek
	COBISS.SI-ID		36334434
2.	Naslov	SLO	Vrednotenje označevanja diskurzni označevalcev v korpusu spontanega govora
		ANG	Validation of discourse markers annotation in a spontaneous speech corpus
	Opis	SLO	V članku predstavljamo postopek označevanja diskurzni označevalcev v korpusih spontanega govora. Vrednotenje je pokazalo, da pri določenih kategorijah diskurzni označevalcev sicer prihaja do odstopanja pri ročnem označevanju, vendar razlike ne vplivajo bistveno na kasnejšo uporabo korpusa za razvoj modelov razpoznavalnika govora. Zelo pomemben rezultat vrednotenja je določitev skupine tistih označevalcev, ki vedno nastopajo v vlogi diskurzni označevalcev. Tako bo v prihodnje možno delno avtomatizirati postopek označevanja korpusov spontanega govora.

		ANG	The paper presents the validation results for annotating the discourse markers in spontaneous speech databases. The analysis showed that during manual annotation variances occur for specific categories of discourse markers, but are insignificant for later usage of speech databases for building acoustic models for automatic speech recognition. An important result is the definition of those tokens that always occur in the role of discourse markers. This will ease the procedure of automatic annotation of spontaneous spoken language resources for spontaneous speech recognition in the future.
	Šifra		B.03 Referat na mednarodni znanstveni konferenci
	Objavljeno v		VERDONIK, Darinka, ŽGANK, Andrej, PISANSKI PETERLIN, Agnes. Validacija označevanja diskurzivnih označevalcev v korpusih Turdis-2 in BNSInt. Zbornik Šeste konference Jezikovne tehnologije, 16. do 17. oktober 2008, Ljubljana, Slovenia : 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. 29-32.
	Tipologija		1.08 Objavljeni znanstveni prispevek na konferenci
	COBISS.SI-ID		12719894
3.	Naslov	SLO	Modeliranje mašil v spontanem govoru
		ANG	Modeling of filled pauses in spontaneous speech
	Opis	SLO	Zasnovali smo sistem za modeliranje mašil pri razpoznavanju slovenskega spontanega govora. Osnovo je predstavljala analiza diskurzivnih označevalcev v formalnem in neformalnem govoru. Na nivoju akustičnih modelov smo modeliranje izvedli s štirimi različnimi implicitnimi topologijami, ki so predstavljale različno kompleksnost zgradbe mašil. Pri tem smo uporabljali pravila na osnovi ekspertnega jezikovnega znanja. Eksperimente smo izvedli na slovenskih govornih virih. Analiza rezultatov je pokazala pomemben vpliv modeliranja mašil na performance razpoznavanja spontanega govora.
		ANG	A framework for modeling filled pauses for Slovenian spontaneous speech recognition was designed. The baseline was set with the analysis of discourse markers in formal and informal genres. The acoustic modeling for speech recognition was performed using four proposed implicit topologies with various complexities. The expert defined language dependent rules were used at the start. The experiment was carried out using Slovenian spoken language resources. The results analysis showed a significant influence of modeling filled pauses on the performance of spontaneous speech recognition
	Šifra		B.03 Referat na mednarodni znanstveni konferenci
	Objavljeno v		ŽGANK, Andrej, ROTOVNIK, Tomaž, SEPESY MAUČEC, Mirjam. Modeling filled pauses for spontaneous speech recognition applications. V: Advanced applications of electrical engineering : proceedings of the 7th WSEAS International Conference on Application of Electrical Engineering (AEE`08), Trondheim, Norway, July 2-4, 2008, (Recent Advances in Electrical Engineering). [S. l.]: WSEAS Press, 2008, str. 42-47.
	Tipologija		1.08 Objavljeni znanstveni prispevek na konferenci
	COBISS.SI-ID		12454422
4.	Naslov	SLO	Kombiniranje govornih virov za razpoznavanje spontanega govora
		ANG	Combining Spoken Language Resources for Spontaneous Speech Recognition
	Opis	SLO	Pri jezikih z majhnim številom govorcev predstavlja dostopnost govornih in jezikovnih virov eno izmed glavnih ovir za razvoj govornih tehnologij. To je še posebej izrazito pri razpoznavanju spontanega govora. V članku smo zasnovali postopke kombiniranja obstoječih govornih virov za razpoznavanje novih domen in nenadzorovanega oziroma rahlo nadzorovanega učenja akustičnih modelov razpoznavalnika spontanega govora. Na osnovi govorne baze BNSI Broadcast News smo tako uporabili govorne baze SloParl, ki vsebuje parlamentarne razprave ter govorni bazi PoliDat in Plattos TTS.
			Languages with small number of speakers often suffer from limited availability of spoken language resources, due to immense costs to develop them. One of possible solutions in such case is to combine various existing spoken language resources to recognize speech for a new task or domain.

		ANG	This combined approach was also applied to unsupervised and lightly supervised training of acoustic models for spontaneous speech recognition. The BNSI Broadcast News database was used to recognize SloParl Parliamentary debates, utterances from the PoliDat database and utterances from the Plattos database.
	Šifra		B.03 Referat na mednarodni znanstveni konferenci
	Objavljeno v		ŽGANK, Andrej. Combining Spoken Language Resources for Spontaneous Speech Recognition. Proc. Advances in Speech Technology Workshop 2008. Maribor, Slovenia. (v tisku).
	Tipologija		1.08 Objavljeni znanstveni prispevek na konferenci
	COBISS.SI-ID		000000000
5.	Naslov	SLO	
		ANG	
	Opis	SLO	
		ANG	
	Šifra		
	Objavljeno v		
	Tipologija		
	COBISS.SI-ID		

8. Pomen raziskovalnih rezultatov projektne skupine⁷

8.1. Pomen za razvoj znanosti⁸

SLO

Rezultati projekta so pomembno doprinesli k dvema področjema avtomatskega razpoznavanja govora, ki ga med drugim uporabljamo tudi na področju razvoja telekomunikacijskih storitev. Dosegli smo pomembno izboljšanje rezultatov razpoznavanja spontanega govora, predvsem na področju modeliranja mašil in onomatopej. Predlagano metodo smo sicer eksperimentalno ovrednotili na govornih virih za slovenski jezik, vendar je zaradi uporabe podatkovno vodenih metod modeliranja jezikovno neodvisna. Predlagano metodo bo tako možno prenesti tudi na modeliranje spontanega govora v drugih svetovnih jezikih. Avtomatsko razpoznavanje govora je vključeno tudi v druga področja informacijsko-komunikacijskih tehnologij (npr.: strojno prevajanje iz govora v govor, uporabniški vmesniki), tako da lahko posledično pričakujemo vpliv tudi na njih. Analiza rezultatov modeliranja spontanega govora je pokazala, da predstavlja izredno kompleksen problem modeliranje delno izgovorjenih besed, tako da bo smiselno temu področju posvetiti dodatno pozornost v prihodnjih raziskavah, kar potrjujejo tudi tematike aktualnih objav na mednarodnih konferencah.

Rezultati raziskav o kombiniranju govornih virov so pokazali, da je na takšen način možno izboljšati tako akustično modeliranje spontanega govora kot tudi modeliranje močno pregibnih jezikov. To je bilo še posebej opazno pri učenju na osnovi nepopolnih transkripcij. Smernice o možnostih kombiniranja različnih govornih virov so še posebej pomembno za jezike s slabše razvitimi govornimi viri. V to kategorijo sodi velika večina današnjih živih jezikov, saj je razvoj novih govornih virov povezan z zelo velikimi časovnimi in finančnimi vložki.

Opravljenе raziskave na področju modeliranja pregibnih jezikov so zmanjšale zaostanek teh jezikov za zahodnoevropskimi jeziki na področju razpoznavanja govora. Rezultati raziskav vzpodbujajo nadaljnje delo na raziskovalnem področju dekodiranja s podbesednimi enotami, ki še vedno predstavlja eno izmed glavnih metod modeliranja močno pregibnih jezikov.

Skupni vpliv raziskav v okviru izvedenega projekta je pomembno doprinesel k povečanju sposobnosti performanc razpoznavalnika spontanega govora. Eksperimentalni sistem za slovenski jezik je tako že uporaben za indeksiranje različnih avdio-vizualnih vsebin ter posledično na takšen način izboljšano možnost dostopa do informacij za nadaljnje procesiranje. Hkrati je pomemben rezultat projekta tudi interdisciplinarni doprinos k področju raziskav o diskurzni označevalci, izsledke česar smo objavili v reviji kategorije A1" po metodologiji ARRS.

ANG

The results of this project significantly contributed to two research areas of automatic speech

recognition, which is also being used for development of telecommunication services. A statistically significant improvement of spontaneous speech recognition performance was achieved, especially for modeling filled pauses and onomatopoeas. The proposed method was evaluated using Slovenian spoken language resources, but the method is due to the usage of data-driven metrics for modeling, language independent. Therefore the proposed method could be also used for modeling spontaneous speech in other world's languages. Automatic speech recognition also takes its role in other areas of information-communication technologies (e.g.: speech-to-speech translation, user interfaces,...), consequently influence on these areas could be also anticipated. The analysis of modeling spontaneous speech also showed that modeling of truncated words presents a highly complex problem. Therefore additional focus should be given to this topic in the future, which is also supported with recent publications on international conferences.

The results on combining spoken language resources showed that it is possible to significantly improve the acoustic modeling for spontaneous speech and modeling of highly inflectional languages using such approach. This was especially noticeable in the case, when imperfect transcriptions were used for training. The guidelines on combining spoken language resources are very important for languages with less well developed spoken language resources. The majority of present-day languages belong into this group, as is the development of spoken language resources connected with immense expenses and it is also very time consuming.

The performed research in the area of modeling highly inflectional languages reduced the gap of these languages behind other western European languages in the area of speech recognition. The research results contribute to increased interest in the area of subword unit decoding, which is still one of the most important methods for modeling highly inflectional languages.

The achieved project's results significantly contributed to the increased performance of spontaneous speech recognition. The Slovenian experimental system could be for example used for indexing various audio-visual contents and in such a way increase the availability of information for further processing. An important interdisciplinary result of this project is also the achievement in the area of discourse studies. These results were presented in an A1" publication according to the ARRS methodology.

8.2. Pomen za razvoj Slovenije⁹

SLO

V okviru projekta predlagane metode so jezikovno neodvisne, vendar še posebej primerne za modeliranje močno pregibnih jezikov, med katere sodi tudi slovenski jezik. Tako so rezultati projekta izrednega pomena za ohranjanje slovenskega jezika v obdobju razmaha digitalizacije. Podpora za razpoznavanje spontanega govora v materinem jeziku namreč posredno vzpodbuja razmah različnih informacijsko-komunikacijskih tehnologij z govorno podporo v slovenskem prostoru. Na ta način se zmanjšuje digitalna ločnica pri dostopu do različnih virov informacij, saj je tako možno uporabljati uporabnikom prijaznejše vmesnike za dostop do virov.

Izsledke o možnostih kombiniranja govornih virov za razpoznavanje spontanega govora bo možno v prihodnje uporabljati pri snovanju novih govornih virov za slovenski jezik, kjer lahko pričakujemo večji poudarek na razvoju cenejših govornih virov z nepopolnimi transkripcijami. To je še posebej pomembno, če upoštevamo, da se je v zadnjem obdobju s porastom prenosnih hitrosti bistveno povečala dostopnost različnih avdio-vizualnih vsebin na spletu (npr: seje DZ RS). Za avdio-vizualne vsebine, kjer transkripcije niso na voljo na spletu, je možno nepopolne transkripcije tvoriti tako ročno kot tudi avtomatsko z uporabo avtomatskega razpoznavalnika govora v načinu nenadzorovanega učenja.

Rezultati eksperimentalnega sistema nakazujejo tudi možnosti prihodnjega prenosa izsledkov raziskave v okviru aplikacijskih projektov v slovensko telekomunikacijsko industrijo, predvsem na področju indeksiranja avdio-vizualnih vsebin. Možnosti uporabe takšnega sistema so predvsem v arhivih različnih medijskih hiš.

S stališča skrbi za slovenski jezik so še posebej pomembni rezultati interdisciplinarnih raziskav o diskurzni označevalci, saj so bile te raziskave prve takšne vrste za slovenski jezik. Opravljena je bila analiza vpliva validatorjev na ročno označevanje diskurzni označevalcev, pogostost pojavljanja diskurzni označevalcev v različnih zvrsteh govora ter vpliv kontekstnih lastnosti na diskurzne označevalce. Ti rezultati bodo pomembno vplivali na prihodnje znanstveno delo na področju diskurza.

ANG

The methods proposed in the project are language independent, but are very suitable for modeling highly inflectional languages as is Slovenian language. The project's results are very significant for the preservation of Slovenian language in the era of digitalization. The support for spontaneous speech recognition in native language is necessary to increase the interest for the development of various voice driven information-communication services in Slovenia. This results in the reduction of digital divide, as the users can easily access various information sources, using user friendly interfaces.

The results on combining spoken language resources for spontaneous speech recognition could be applied to future development of new Slovenian speech databases. The anticipated trends are in increased importance of developing less expensive speech databases with imperfect transcriptions. This is of immense importance, if we take into account the increased penetration rate of broad-band internet access and the resulting availability of various audio-visual content on the internet (e.g. Parliamentary debates). For those audio-visual content, which is available without transcriptions, the imperfect transcription can be generated manually or automatically using automatic speech recognition system in an unsupervised training mode.

The results of experimental system pointed out the possibility of transfer of knowledge in a scope of an applied project in Slovenian telecommunication industry, specifically in the area of audio-video content indexing. Such system could be very valuable for different content providers.

From the point of view of care for Slovenian language are very important interdisciplinary results on discourse markers. These results were the first such results for Slovenian language. The following research was carried out: analysis of influence of human annotators, the analysis of discourse markers frequency in different genres and the analysis of influence of context on discourse markers in spontaneous speech. These results are important for future research work on the topic of discourse studies.

9. 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	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.02	Pridobitev novih znanstvenih spoznanj	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.03	Večja usposobljenost raziskovalno-razvojnega osebja	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.04	Dvig tehnološke ravni	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.05	Sposobnost za začetek novega tehnološkega razvoja	

	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.06	Razvoj novega izdelka	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.07	Izboljšanje obstoječega izdelka	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.08	Razvoj in izdelava prototipa	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.09	Razvoj novega tehnološkega procesa oz. tehnologije	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.10	Izboljšanje obstoječega tehnološkega procesa oz. tehnologije	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.11	Razvoj nove storitve	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.12	Izboljšanje obstoječe storitve	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.13	Razvoj novih proizvodnih metod in instrumentov oz. proizvodnih procesov	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.14	Izboljšanje obstoječih proizvodnih metod in instrumentov oz. proizvodnih procesov	

	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.15	Razvoj novega informacijskega sistema/podatkovnih baz	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.16	Izboljšanje obstoječega informacijskega sistema/podatkovnih baz	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.17	Prenos obstoječih tehnologij, znanj, metod in postopkov v prakso	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.18	Posredovanje novih znanj neposrednim uporabnikom (seminarji, forumi, konference)	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.19	Znanje, ki vodi k ustanovitvi novega podjetja ("spin off")	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.20	Ustanovitev novega podjetja ("spin off")	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.21	Razvoj novih zdravstvenih/diagnostičnih metod/postopkov	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.22	Izboljšanje obstoječih zdravstvenih/diagnostičnih metod/postopkov	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.23	Razvoj novih sistemskih, normativnih, programskih in metodoloških rešitev	

	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.24	Izboljšanje obstoječih sistemskih, normativnih, programskih in metodoloških rešitev	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.25	Razvoj novih organizacijskih in upravljavskih rešitev	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.26	Izboljšanje obstoječih organizacijskih in upravljavskih rešitev	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.27	Prispevek k ohranjanju/varovanje naravne in kulturne dediščine	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.28	Priprava/organizacija razstave	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.29	Prispevek k razvoju nacionalne kulturne identitete	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.30	Strokovna ocena stanja	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.31	Razvoj standardov	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.32	Mednarodni patent	

	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.33	Patent v Sloveniji	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.34	Svetovalna dejavnost	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.35	Drugo	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	

Komentar

--

10. 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	☐	☐	☐	☐	
G.01.02.	Razvoj podiplomskega izobraževanja	☐	☐	☐	☐	
G.01.03.	Drugo:	☐	☐	☐	☐	
G.02	Gospodarski razvoj					
G.02.01	Razširitev ponudbe novih izdelkov/storitev na trgu	☐	☐	☐	☐	
G.02.02.	Širitev obstoječih trgov	☐	☐	☐	☐	
G.02.03.	Znižanje stroškov proizvodnje	☐	☐	☐	☐	
G.02.04.	Zmanjšanje porabe materialov in energije	☐	☐	☐	☐	
G.02.05.	Razširitev področja dejavnosti	☐	☐	☐	☐	
G.02.06.	Večja konkurenčna sposobnost	☐	☐	☐	☐	
G.02.07.	Večji delež izvoza	☐	☐	☐	☐	
G.02.08.	Povečanje dobička	☐	☐	☐	☐	
G.02.09.	Nova delovna mesta	☐	☐	☐	☐	

Zaključno poročilo o rezultatih raziskovalnega projekta

G.02.10.	Dvig izobrazbene strukture zaposlenih	☐	☐	☐	☐	
G.02.11.	Nov investicijski zagon	☐	☐	☐	☐	
G.02.12.	Drugo:	☐	☐	☐	☐	
G.03	Tehnološki razvoj					
G.03.01.	Tehnološka razširitev/posodobitev dejavnosti	☐	☐	☐	☐	
G.03.02.	Tehnološko prestrukturiranje dejavnosti	☐	☐	☐	☐	
G.03.03.	Uvajanje novih tehnologij	☐	☐	☐	☐	
G.03.04.	Drugo:	☐	☐	☐	☐	
G.04	Družbeni razvoj					
G.04.01	Dvig kvalitete življenja	☐	☐	☐	☐	
G.04.02.	Izboljšanje vodenja in upravljanja	☐	☐	☐	☐	
G.04.03.	Izboljšanje delovanja administracije in javne uprave	☐	☐	☐	☐	
G.04.04.	Razvoj socialnih dejavnosti	☐	☐	☐	☐	
G.04.05.	Razvoj civilne družbe	☐	☐	☐	☐	
G.04.06.	Drugo:	☐	☐	☐	☐	
G.05.	Ohranjanje in razvoj nacionalne naravne in kulturne dediščine in identitete	☐	☐	☐	☐	
G.06.	Varovanje okolja in trajnostni razvoj	☐	☐	☐	☐	
G.07	Razvoj družbene infrastrukture					
G.07.01.	Informacijsko-komunikacijska infrastruktura	☐	☐	☐	☐	
G.07.02.	Prometna infrastruktura	☐	☐	☐	☐	
G.07.03.	Energetska infrastruktura	☐	☐	☐	☐	
G.07.04.	Drugo:	☐	☐	☐	☐	
G.08.	Varovanje zdravja in razvoj zdravstvenega varstva	☐	☐	☐	☐	
G.09.	Drugo: 	☐	☐	☐	☐	

Komentar

11. Pomen raziskovanja za sofinancerje, navedene v 2. točki¹⁰

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

Podpisi:

Andrej Žgank	in/ali	
podpis vodje raziskovalnega projekta		zastopnik oz. pooblaščen oseba RO

Kraj in datum:

Maribor

16.4.2009

Oznaka poročila: ARRS_ZV_RPROJ_ZP_2008/255

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 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](#)

⁸ Največ 4.000 znakov vključno s presledki [Nazaj](#)

Zaključno poročilo o rezultatih raziskovalnega projekta

⁹ Največ 4.000 znakov vključno s presledki Nazaj

¹⁰ 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-ZV-RPROJ-ZP/2008 v1.00