

Oznaka poročila: ARRS-RPROG-LP-2009/117
Status poročila: Neoddano – delno digitalno podpisano
Datum oddaje poročila:

LETNO POROČILO O REZULTATIH RAZISKOVALNEGA PROGRAMA V LETU 2009

A. PODATKI O RAZISKOVALNEM PROGRAMU

1. Osnovni podatki o raziskovalnem programu

Šifra programa	P2-0103
Naslov programa	Tehnologije znanja
Vodja programa	8949 Nada Lavrač
Obseg raziskovalnih ur	skupaj: 62.220 od tega v letu 2009: 10.370
Cenovni razred	C
Trajanje programa	01.2009 - 12.2014
Izvajalci programa (raziskovalne organizacije in/ali koncesionarji)	106 Institut "Jožef Stefan" 1540 Univerza v Novi Gorici
Družbeno-ekonomski cilj	06. Industrijska proizvodnja in tehnologija

2. Povzetek¹

SLO

Področje dela programske skupine Tehnologije znanja so napredne informacijske tehnologije za predstavitev, zajemanje in upravljanje znanja, ki so praktično uporabne za razvoj informacijske in na znanju temelječe družbe. Uveljavljena področja tehnologij znanja vključujejo inteligentno analizo podatkov, besedil in spleta (strojno učenje, rudarjenje podatkov, odkrivanje zakonitosti v podatkih), semantični splet, analizo socialnih omrežij, jezikovne tehnologije in računalniško jezikoslovje, podporo odločanja in upravljanje znanja. Novejša področja raziskav programske skupine vključujejo še splet 2.0, upravljanje virtualnih organizacij ter nove medije in e-znanost. Poleg razvoja tehnologij znanja razvijamo tudi aplikacije teh tehnologij. V letu 2009 smo nadaljevali z razvojem aplikacij predvsem na področjih znanosti o okolju in upravljanja z okoljem, analize medijskih vsebin, medicine in bioinformatike.

ANG

Within the Knowledge Technologies programme we perform research in advanced information technologies, aimed at representing, acquiring and managing knowledge to be used in the development of knowledge-based applications. Established areas of knowledge technologies include intelligent data analysis (machine learning, data mining, knowledge discovery in databases), text and web mining, semantic web, social network analysis, language technologies and computational linguistics, decision support and knowledge management. Research areas of the programme also include Web 2.0, management of virtual organizations, new media and e-science. Apart from knowledge technologies research, we also develop applications in environmental

sciences and ecology, medicine and health care, biomedicine and genetics, economy and marketing.

3. Pomen raziskovalnih rezultatov programske skupine²

3.1. Pomen za razvoj znanosti³

SLO

Javna agencija za raziskovalno dejavnost Republike Slovenije nas je ocenila kot najboljši raziskovalni program na našem področju raziskav (v letih 2004-2008). Pomen programa je v doseganju znanstvene odličnosti, sooblikovanju novih svetovnih znanstvenih trendov ter vpetosti raziskav v mednarodne znanstvene projekte. Tudi v letu 2009 smo pomembno prispevali k svetovni zakladnici znanj, k nadaljnji uveljavitvi slovenske znanosti v evropskem in svetovnem merilu, k prenosu znanstvenih dosežkov v prakso ter k podiplomskemu in strokovnemu izobraževanju v slovenskem in evropskem prostoru.

Področja inteligentne analize strukturiranih podatkov (tabelarični, relacijski), analize večpredstavnih podatkov (teksta, slik, spleta) in semantičnega spleta doživljajo v svetu izjemen razmah predvsem zaradi dostopnosti podatkov po eni, in potreb industrije in vladnih inštitucij po obvladovanju kompleksnih situacij po drugi strani. Pomembnost tovrstnih znanj in tehnologij je še toliko večja, ker služijo kot infrastruktura drugim tehnologijam, kot so ekologija in varstvo okolja, bioinformatika, medicina, ekonomija in jezikoslovje. Programska skupina Tehnologije znanja je v samem evropskem in svetovnem vrhu na področju tovrstnih analiz, kar potrjujejo vključenost v številne mednarodne projekte (18 EU projektov v letu 2009), publikacije v zbornikih konferenc in prestižnih znanstvenih revijah (Journal of Machine Learning Research, IEEE internet computing, ...), vključenost v uredniške odbore znanstvenih revij (MLJ, DAMI, ACM TKDD, AIM, AAI, JMLR, AI Com, IJCL, JLRE, EcoMod, EcoInf), programske odbore konferenc (ICML, ECML/PKDD, ILP, KDD, ...), vabljen predavanja ter povabila na ugledne mednarodne znanstvene in komercialne institucije (Microsoft Research Cambridge, British Telecom). V letu 2009 smo organizirali prestižno konferenco ECML/PKDD 2009, ki predstavlja eno od svetovno najpomembnejših srečanj znanstvenikov s področja strojnega učenja in odkrivanja znanj v podatkih.

Z dosežki v letu 2009 smo ključno prispevali tudi k razvoju novih znanstvenih področij, predvsem rudarjenja relacijskih podatkov in semantičnega rudarjenja podatkov, odkrivanja enačb in rudarjenja tekstovnih podatkov. Prispevali smo k povezavi med področji odkrivanja znanja, semantičnega spleta in analize socialnih omrežij. Na področju jezikovnih tehnologij smo potrdili pomen uporabe metod strojnega učenja nad jezikovnimi podatki in razvoja standardiziranih metod označevanja jezikovnih podatkov. Na področju podpore odločanja tudi z novimi rezultati sodimo pri večparametrskem odločanju v svetovni vrh.

ANG

The Knowledge Technologies research programme has been evaluated as the best research programme in ICT (for the period 2004-2008) by the Slovenian Research Agency (ARRS). The role of the programme is in achieving scientific excellence, and at setting new research trends in connection with our EU and other international research projects. In 2009 we have significantly contributed to the global treasury of knowledge, advanced Slovene science on a European and global scale, advanced the transfer of scientific results into practice, and improved the development of postgraduate and professional education in Slovenia and Europe.

The area of intelligent analysis of structured data (tabular, relational) and multimedia mining (mining texts, images and Web) and the semantic Web have witnessed exceptional growth on a global scale, mainly because of the availability of data on the one hand, and the need of industry and government institutions to master information complexity on the other. The role of knowledge technologies is also increasingly important because they serve as the infrastructure for other disciplines and technologies, e.g. ecology and preservation of the environment, bioinformatics, medicine, health care, economy and linguistics. The Knowledge Technologies programme has been at the very top of such analysis efforts, in European and global terms, confirmed through our participation in numerous international projects (18 EU projects in 2009), papers in leading scientific publications (Journal of Machine Learning Research, Machine Learning Journal, etc.) and prestigious international conferences (ECML, IDA, ICDM, ILP, DS, etc.), organization of numerous workshops, seminars and summer schools, tutorials at international conferences, membership in editorial boards of scientific journals (MLJ, DAMI, ACM TKDD, AIM, AAI, JMLR, JAIR, AI Com, IJCL, JLRE, EcoMod, EcoInf), conference program

committees (ICML, ECML/PKDD, ICDM, ILP, KDD, IDA, ACL), invited lectures, invitations to distinguished international scientific (Stanford, MIT, Tokyo Institute of Technology, ...) and commercial institutions (Microsoft Research Cambridge, British Telecom). In 2009 we organized the ECML/PKDD 2009 conference which is one of the most important scientific meeting in the areas of machine learning and data mining.

In 2009 we have also made key contributions to the development of new areas of science, especially inductive logic programming, inductive databases, equation discovery and text mining. In the area of language technologies, our reputation is based on successful applications of machine learning methods to textual data and the development of standardized methods of annotating linguistic data. We also shifted from the morphological and lexical levels towards problems arising due to developing the semantic Web, notably learning and standardizing ontologies. In the area of decision support, we also play a leading role in supporting multi-parametric decision processes.

3.2. Pomen za razvoj Slovenije⁴

SLO

Po vključenosti v projekte EU smo najuspešnejša programska skupina v Sloveniji. V letih 2004-2009 smo sodelovali oz. sodelujemo v 30 projektih 6. in 7. okvirnega programa (OP). Med EU projekti naj izpostavimo sodelovanje v 6 integriranih projektih (IP) in dveh mrežah odličnosti (NoE).

Pomen raziskovalnega programa pri razvoju Slovenije je dvojen.

1. Neposreden pomen in vloga programa se udejanja v osnovnih raziskavah, ki prispevajo k splošnemu napredku znanja na področju informacijskih tehnologij. Raziskovalni program zagotavlja aktivno vključevanje raziskovalcev v domače in mednarodne znanstvene tokove (npr. sodelovanje v EU projektih), kar je nujen pogoj za razvoj informacijske in na znanju temelječe družbe. Pri tem je pomembno poudariti odličnost raziskovalcev programske skupine in njihove mednarodno odmevne dosežke v mednarodnih in domačih projektih.

2. Druga vloga programa je posredna in se izraža v pomenu, ki ga imajo visoko razvite informacijske tehnologije kot osnova za razvoj in podporo drugih področij znanja za trajnostni družbeno-ekonomski in kulturni razvoj. Naši rezultati so še posebej pomembni za Slovenijo na naslednjih področjih. (a) Metode za inteligentno analizo podatkov smo uporabili na vrsti praktičnih problemov s področja znanosti o okolju, za odkrivanje rizičnih skupin pacientov, za odkrivanje povezave med genskimi informacijami in kliničnimi simptomi ter na problemih analize mikromrež v bioinformatiki (v sodelovanju z Nacionalnim institutom za biologijo). (b) Naš sistem za podporo večparametrskega odločanja smo uporabili v aplikacijah na področju izobraževanja, zdravstva ter vrednotenja investicij in projektov, kar omogoča večjo transparentnost delovanja javnih zavodov in javne uprave. (c) Pomen programa za kulturni razvoj Slovenije se navezuje na dejstvo, da je kulturna identiteta Slovencev že od nekdaj tesno povezana z jezikom. Delo na področju jezikovnih tehnologij poteka v sodelovanju s komplementarnimi slovenskimi institucijami: FF, FDV, ZRC SAZU, NUK, DZS, Amebis in Alpineon. Naše delo je uporabno pri poučevanju slovenskega jezika, slovaropisju, in strojno podprtem prevajanju, v teku pa je tudi več projektov za razvoj strojnega prevajanja. (d) Pomen za tehnološki razvoj je povezan s prehodom slovenskega gospodarstva v inovacijsko družbo; strateška usmeritev uspešnih gospodarskih družb je izdelava proizvodov in storitev, pri katerih je dodana vrednost največja. S pomočjo tehnologij za obdelavo in analizo velikih količin podatkov in semantičnega spleta prispevamo k razvoju lastnih vrhunskih informacijskih tehnologij, ki jih uporabljamo za razvoj storitev z visoko dodano vrednostjo, ki so uporabne v aplikacijah za slovenske in tuje naročnike. Raziskovalni program ima močno povezovalno komponento, tako med raziskovalnimi skupinami, kot tudi z gospodarstvom.

Sodelavci programa smo aktivni v prenosu znanja preko aplikativnih projektov in seminarjev organiziranih v sodelovanju s Centrom za prenos znanja na področju informacijskih tehnologij IJS. Sodelujemo v izobraževalnih aktivnostih Univerze v Ljubljani in Mariboru, Univerze v Novi Gorici in vrste tujih univerz ter pri oblikovanju in izvajanju študijskega programa Novi mediji in e-znanost na Mednarodni podiplomski šoli Jožefa Stefana. V letu 2009 smo organizirali 7 Solomonovih seminarjev, ki so namenjeni izboljšanju splošne ravni izobraževanja na področju informacijskih tehnologij. Pomemben korak k realizaciji izobraževanja na daljavo je tudi naše spletišče Videolectures.net na katerem je na voljo že preko 9700 videoposnetkov predavanj s cele vrste znanstvenih dogodkov.

ANG

In terms of our collaboration in EU projects, our programme group is the most successful in Slovenia. In 2004-2009 we have collaborated in 30 FP6 and FP7 projects, including six large

integrated projects (IP) and two networks of excellence (NoE).

The importance of the programme for this development is twofold:

1. Direct significance and role of the programme is embodied in its basic research component, which contributes to the general advancement of knowledge in the area of information technologies. The research programme ensures active inclusion of researchers into national and international scientific trends (e.g. cooperation in EU projects), which is a necessary condition for the development of the emerging knowledge-based society. The specific results include authorships of papers, book chapters and scientific monographs, editing books and special issues of scientific journals and proceedings of international conferences.
2. The second role of the programme is indirect, and expressed in the significance of advanced information technologies as the basis for the development and support for other knowledge intensive areas, ensuring sustainable socio-economic and cultural development. Our results are especially important for Slovenia in the following areas. (a) Methods of intelligent data analysis are used to solve numerous practical problems of environmental sciences and management, biochemical engineering, medical informatics and the rationalization of the Slovene healthcare system, pharmacology, epidemiology and the prevention of occupational disease in industry, discovering risk groups, and discovering connections between genetic information and clinical symptoms. (b) Our system for multi-parametric decision support is used in a series of applications in education, healthcare and the evaluation of projects and investments, improving the transparency of activities in public services and administration. (c) The programme is also important for the cultural development of Slovenia. Our work has applications in Slovene language teaching, lexicography and machine or machine-aided translation, and also concerns digital libraries, esp. the preservation of cultural heritage. (d) The importance of our contributions to the general technological development is connected to the transition of the national economy towards an innovation-oriented and knowledge-based society with a large added value. With technologies for processing and analyzing large amounts of data and exploiting the possibilities of the semantic Web we contribute to the development of our own advanced information technologies that are used in developing added-value services for Slovene and international clients.

Members of the programme team take active part in the transfer of knowledge through applied projects and training/seminars organized in cooperation with the JSI Centre for knowledge transfer in information technologies. We participate in educational activities of the Universities of Ljubljana, Maribor, Nova Gorica and numerous universities abroad and have contributed to the constitution and curricula development of the study programme New media and e-science at the Jozef Stefan International Postgraduate School. In 2009 we organised 7 Solomon seminar, dedicated to improving the general level of education in the area of information technologies. An important step towards the goal of providing distance learning/education is our Web portal Videolectures.net which has already made publicly available more than 9,700 video recordings of lectures from a broad range of scientific events.

B. REZULTATI IN DOSEŽKI RAZISKOVALNEGA PROGRAMA

4. Poročilo o realizaciji programa raziskovalnega programa⁵

Na področju **inteligentne analize podatkov** smo (a) razvili algoritem in metodologijo iskanja kontrastnih množic z metodami odkrivanja podskupin in jo uspešno uporabili pri odkrivanju karakteristik dveh skupin pacientov z možgansko kapjo; delo smo objavili v Journal of Biomedical Informatics. (b) Dodelali smo konceptualni okvir za nadzorovano učenje opisnih pravil ter pokazali, da vključuje metode za odkrivanje podskupin, odkrivanje kontrastnih množic in metode odkrivanja pojavljajočih vzorcev; delo smo objavili v Journal of Machine Learning Research. (c) Nadaljevali smo z razvojem metod semantičnega rudarjenja podatkov, ki omogoča uporabo ontologij kot predznanja v učenju, ter ga uporabili na področju odkrivanja izraženosti skupin genov na osnovi ontologije njihovih funkcij, procesov in interakcij. (d) Nadgradili smo ta pristop z namenom bisociativnega odkrivanja zakonitosti, ki ga raziskujemo v okviru projekta BISON, tako da najprej s semantičnim odkrivanjem znanja odkrijemo skupine izraženih genov, zatem pa sledi kreativno preiskovanje bisociativnih povezav z orodjem Biomine (naših partnerjev iz

Univerze v Helsinkih) za preiskovanje grafov. Izboljšali smo tudi pristop za kreativno odkrivanje zakonitosti iz tekstovnih dokumentov ter ga objavili v Journal of biomedical informatics. (e) Nadaljevali smo z razvojem sistema za servisno-orientirano odkrivanje znanja ter razvili operativno platformo imenovano Orange4WS. (f) Nadaljevali smo z raziskavo novih pristopov za odkrivanje gensko spremenjenih organizmov v hrani in krmi ter predlagali novo metodologijo odkrivanja neavtoriziranih gensko spremenjenih organizmov, ki je bila objavljena v Anal. bioanal. chem. journal.

Uspešno smo zaključili projekt 6. OP EETP (European Embryonal Tumor Pipeline), ki se je ukvarjal z zbiranjem in analizo podatkov o embrionalnih tumorjih. V projektu so bile zbrane različne vrste podatkov, in sicer na področju genomike, transkriptomike in proteomike, kot tudi podatki o klinični sliki pacientov. Projekt je obravnaval različne vrste embrionalnih tumorjev (tako imenovane entitete), kot so nevroblastomi, meduloblastomi, Ewingov sarkom ter Wilmovi in rabdoidni tumorji. Poglavitne naloge analize podatkov znotraj projekta so bile identifikacija genov pomembnih za razvoj različnih vrst tumorjev kot tudi genov, ki so pomembni za več entitet hkrati. V ta namen smo razvili tehnike za evalvacijo oz. vrednotenje rangiranih seznamov genov kot tudi za njihovo združevanje oz. agregacijo. Omenjene tehnike smo uporabili na podatkih o izraženosti genov za različne vrste tumorjev ter dobili rangirane liste genov za vsako entiteto posebej, ki so potem bile združene v rangirane liste genov za več vrst tumorja hkrati oz. za vse vrste embrionalnih tumorjev hkrati. Za gene, ki so bili visoko rangirani, smo predlagali biološke razlage (z iskanjem »obogatenih« množic genov), ter rekonstruirali mreže interakcij med njimi.

Nadaljevali smo razvoj metod za analizo kompleksnih podatkov, kot so drevesa in pravila za napovedno razvrščanje kratkih časovnih vrst ter ansambli odločitvenih dreves za hierarhično več-značkovno označevanje, ter metod za odkrivanje enačb za modeliranje dinamičnih sistemov. Omenjene metode smo uporabili na vrsti praktičnih problemov iz področij okoljskih znanosti in znanosti o življenju, kot na primer napovedovanje strukture združb diatomej, modeliranje pretoka genov med gensko spremenjenimi in konvencionalnimi posevki ter označevanje medicinskih slik oz. anotacija. Razvoj metod poteka v okviru dveh domačih projektov in enega projekta 7. OP, ki so osredotočeni na razvoj metod za analizo podatkov v sistemski biologiji.

Dolgoročni cilj na področju **podpore odločanja** je razvijati metode in tehnike odločitvenega modeliranja, jih podpreti z računalniškimi orodji ter jih povezovati s sistemi za rudarjenje podatkov. V letu 2009 smo dvakrat razširili računalniški program za simbolično večparametrsko odločanje DEXi z možnostmi za uvažanje, izvažanje in iskanje komponent odločitvenih modelov ter z novimi operacijami združevanja atributov. Najnovejša verzija nosi oznako 3.02 in je brezplačno dostopna preko spletne strani <http://kt.ijs.si/MarkoBohanec/dexi.html>.

Za potrebe evropskega projekta Co-Extra smo razvili več odločitvenih modelov in aplikacij za podporo pri odločanju o uporabi gensko spremenjenih rastlin v prehrani. Posebej bi radi izpostavili zelo obsežna modela za vrednotenje postopkov laboratorijskega odkrivanja gensko spremenjenih organizmov, ki smo ju razvili v sodelovanju z Nacionalnim institutom za biologijo. Modela sta podprta s podatkovno bazo in spletno aplikacijo, ki omogoča njuno široko in preprosto rabo.

V okviru evropskega projekta HEALTHREATS, katerega cilj je razviti sistem za podporo odločanja ob pojavu epidemij, smo tudi v letu 2009 razvijali modele internega vrednotenja dela in rezultatov projekta.

Uspešno smo zaključili interdisciplinarni raziskovalni projekt Metodološki vidiki raziskovanja kognitivnih procesov-učenje in odločanje ter rezultate objavili v knjigi Konteksti odločanja.

Na področju **analize besedil in spleta** je naše delo v evropskem projektu SMART (Statistical Multilingual Analysis for Retrieval and Translation) obsegalo predvsem evalvacijo sistema s strani končnih uporabnikov (strojno prevajanje in med-jezično iskanje

informacij) in laboratorijsko evalvacijo različnih metod razvitih v projektu. Razvili smo tudi metode za čezjezično klasifikacijo dokumentov na osnovi CCA in SVM metod. Uspešno smo zaključili delo na evropskem projektu IMAGINATION (Image-based Navigation in Multimedia Archives), kjer smo se v zadnjem letu projekta predvsem ukvarjali z integracijo razvitih metod za razširitev obstoječih tekstovnih metapodatkov slik. Aktivnosti na mreži odličnosti PASCAL2 (Pattern Analysis, Statistical Modeling and Computational Learning 2) so potekale predvsem na področju analize tekstovnih podatkov in procesiranja naravnega jezika.

Na področju **semantičnega spleta** smo uspešno zaključili delo na dveh evropskih projektih. V projektu TAO (Transitioning Applications to Ontologies) smo zaključili delo na razvoju metodologije za analizo čezmodalnih podatkov, ki sloni na povezovanju podatkov predstavljenih v obliki besedil in v obliki grafov oz. omrežji, ki smo jo ovrednotili na treh realnih problemih. V okviru evropskega projekta SWING (Semantic Web Services Interoperability for Geospatial Decision Making) smo zaključili razvoj orodja za semantično anotacijo OntoBridge, ki smo ga evalvirali na realnih podatkih. V okviru evropskega projekta NeOn (Lifecycle Support for Networked Ontologies) smo razširili NeOn Toolkit platformo z nekaj prototipi. (1) Razširitev spletne aplikacije SearchPoint, ki omogoča spletno iskanje ontologij z uporabo drugih ontologij. (2) Razvoj spletne aplikacije, ki integrira Cyc ontologijo in AnswerArt tehnologijo ter omogoča odgovarjanje na vprašanja postavljena v naravnem jeziku na osnovi množice dokumentov relevantnih za uporabnika. Cyc ontologija je skupaj z domensko ontologijo ASFA in WordNet ontologijo uporabljena za pridobivanje semantične informacije v obliki sinonimov in posplošitev besedila iz uporabniških dokumentov (ASFA povzetki). Delo na evropskem projektu ACTIVE (Enabling the Knowledge Powered Enterprise, 2008-2011) je potekalo predvsem v smeri modeliranja neformalnih procesov velikih organizacij z uporabo modeliranja kontekstov.

Smo predstavniki instituta Jožef Stefan v World Wide Web konzorciju (W3C), ki razvija in uveljavlja bodoče standarde za svetovni splet. Aktivno sodelujemo v delovni skupini za standardizacijo formata za izmenjavo pravil Rule Interchange Format (RIF). Skupina je že pripravila osnutke in zadnja priporočila za več standardov: RIF Core, Basic Logic Dialect, Framework for Logic Dialects, RDF and OWL Compatibility, Datatypes and Built-Ins, Production Rules Dialect, Use Cases and Requirements.

Na področju **upravljanja znanja** smo letos nadaljevali delo in raziskave s področja mrežnih organizacij, modelov in orodij za podporo interoperabilnosti in sodelovanja med organizacijami v integriranem projektu 7. OP COIN (COllaboration and INteroperability for networked enterprises). Podrobno smo se usmerili in nadaljevali z razvoje aplikacij na bazi tehnologij znanj za modeliranja procesov in kompetenc. Nadaljevali smo uspešen razvoj inteligentnega sistema za nadzor in upravljanje transporta kontejnerjev po Evropi v projektu 7. OP EURIDICE (European Inter-Disciplinary Research on Intelligent Cargo for Efficient, Safe and Environment-friendly Logistics), kjer smo že pričeli z združevanjem orodij za strojno učenje z Cyc sistemom in tako pripravljati osnovo za končni integriran sistem. Razvite prototipe implementiramo v orodja za zajem in formalizacijo tihega znanja v znanjsko intenzivnih podjetjih (projekt 7. OP ACTIVE - Enabling the Knowledge Powered Enterprise).

Na področju **jezikovnih tehnologij** smo v letu 2009 končali z delom na temeljnem projektu JOS, Jezikoslovno označevanje slovenskega jezika: metode in viri, v okviru katerega smo razvili jezikoslovno označene korpuse slovenskega jezika, na ravneh oblikoslovja, skladnje in semantike. Razviti viri so prosto dostopni pod licenco Creative Commons na <http://nl.ijs.si/jos/>. Kot partner smo v temeljnem projektu Slovensko prevodoslovje - viri in raziskave, ki ga vodi Oddelek za prevajanje Filozofske fakultete Univerze v Ljubljani začeli z delom na jezikoslovnem označevanju vzporednih

dvojezičnih korpusov, ki vsebujejo slovenski in glavne evropske jezike. Kot partner v aplikativnem projektu Neznani rokopisi slovenskega slovstva 17. in 18. stoletja: informacijsko-tehnološko podprta evidenca, znanstvene objave in analize in ciljnem raziskovalnem projektu Elektronsko besedilno središče z multimedijско komunikacijo, ki ju vodi ZRC SAZU smo delali na digitalnih izdajah slovenskega leposlovja, predvsem (v sodelovanju z odsekom F9 IJS) na razvoju platforme za predstavitev in iskanje po dokumentih in na spletnem vmesniku za pretvorbo dokumentov v XML, ki so skladni s smernicami za zapis besedil TEI.

V okviru projekta EU MondiLex Conceptual Modeling of Networking of Centers for High-Quality Research in Slavic Lexicography and their Digital Resources smo v sodelovanju z odsekom F9 IJS raziskovali možnosti uporabe tehnologij Grid za obdelave in diseminacijo velikih korpusov, razširili večjezične vire MULTTEXT-East z dodatnimi slovanskimi jeziki in organizirali odprto delavnico z objavljenim zbornikom. V projektu EU FlareNet, Fostering Language Resources Network smo nadaljevali z izdelavo jezikovnih virov za slovenski jezik, v okviru slovensko-francoskega bilateralnega projekta Definicija sintaktično-semantične strukture slovenskega glagola, pa smo vzpostavili sodelovanje z inštitutom INALCO iz Pariza.

5. Ocena stopnje realizacije zastavljenih raziskovalnih ciljev⁶

Glavni cilji programa so bili razvoj novih metod, tehnik in orodij za inteligentno analizo podatkov, jezikovne tehnologije in računalniško jezikoslovje, podporo odločanja in upravljanje znanja. Glavno področje dela so bile raziskave in razvoj metod inteligentne analize podatkov, t.j., metode strojnega učenja, rudarjenja podatkov in odkrivanja zakonitosti v podatkih ter rudarjenje tekstovnih, spletnih in multimedijskih vsebin. Glavni cilji programa so bile vrhunske raziskave in razvoj aplikacij tehnologij znanja, njihov prenos v tehnološko, poslovno, upravno in družbeno okolje ter izobraževanje kadrov na področju tehnologij znanja.

Programska skupina Tehnologije znanja je dosegla zadane cilje znanstvene odličnosti na področju razvoja in aplikacij tehnologij znanja, vključno z objavami v vrhunskih znanstvenih revijah.

Programska skupina je bila tudi izjemno aktivna na pedagoškem področju ter izjemno uspešna na področju mednarodnega sodelovanja in črpanja/uporabe sredstev evropskih projektov, v sodelovanju z mednarodnimi znanstvenimi skupinami. Ugotavljamo, da smo presegli zastavljene cilje programske skupine, da imamo stabilno finančno stanje, proaktivne sodelavce, aktivno mednarodno sodelovanje in širok nabor zanimivih raziskovalnih tematik za raziskave.

6. Utemeljitev morebitnih sprememb programa raziskovalnega programa⁷

Ni sprememb programa.

7. Najpomembnejši znanstveni rezultati programske skupine⁸

Znanstveni rezultat			
1.	Naslov	SLO	Nadzorovano učenje opisnih pravil: poenoten pregled odkrivanja kontrastnih množic, odkrivanja porajajočih se vzorcev in odkrivanja podskupin
		ANG	Supervised descriptive rule discovery: a unifying survey of contrast set, emerging pattern and subgroup mining

	Opis	SLO	V članku definiramo termin nadzorovano učenje opisnih pravil, v okviru katerega združimo odkrivanje podskupin, odkrivanje kontrastnih množic, odkrivanje porajajočih se vzorcev in druga sorodna področja. V skupnem teoretskem okviru poenotimo terminologijo, kar omogoči tudi poenotenje definicij. Prednost uporabe poenotenega teoretskega okvira je posplošitev vizualizacijskih metod, ki so bile razvite za odkrivanje podskupin, na nadzorovano učenje opisnih pravil.
		ANG	This paper gives a survey of contrast set mining (CSM), emerging pattern mining (EPM), and subgroup discovery (SD) in a unifying framework named supervised descriptive rule discovery. The paper contributes a novel understanding of these sub-areas of data mining by presenting a unified terminology, by explaining the apparent differences between the learning tasks and by exploring the apparent differences between the approaches. The paper also provides a critical survey of existing supervised descriptive rule discovery visualization methods.
	Objavljeno v		KRALJ NOVAK, Petra, LAVRAČ, Nada, WEBB, Geoffrey I. Supervised descriptive rule discovery : a unifying survey of contrast set, emerging pattern and subgroup mining. J. mach. learn. res.. [Print ed.], 2009, vol. 10, str. 377-403. http://jmlr.csail.mit.edu/papers/v10/
	Tipologija		1.01 Izvirni znanstveni članek
	COBISS.SI-ID		22475303
2.	Naslov	SLO	CSM-SD: Metodologija za odkrivanje kontrastnih množic z odkrivanjem podskupin
		ANG	CSM-SD: Methodology for Contrast Set Mining through Subgroup Discovery
	Opis	SLO	V članku predstavimo metodološko novost, kako problem odkrivanja kontrastnih množic uspešno rešimo s preoblikovanjem v bolj znano in dobro razumljeno metodo odkrivanja podskupin. Pokažemo tudi, da se razumljivost najdenih razlik še poveča, če jim dodamo dodatne opise, ki se imenujejo podporni dejavniki. Predlagana metodologija je bila uporabljena za odkrivanje značilnosti dveh skupin bolnikov: eni s trombozo in drugi z embolijo.
		ANG	As a methodological novelty, we show that the contrast set mining problem can be effectively solved by transforming it to a more common and well-understood subgroup discovery task. We also show that the explanatory potential of discovered contrast sets can be improved by offering additional contrast set descriptors, called the supporting factors. The proposed methodology has been applied to data about two groups of patients, with ischemia caused by thrombosis and by embolism, respectively.
	Objavljeno v		KRALJ NOVAK, Petra, LAVRAČ, Nada, GAMBERGER, Dragan, KRSTAIČIĆ, Antonija. CSM-SD: methodology for contrast set mining through subgroup discovery. Journal of biomedical informatics, 2009, vol. 42, no. 1, str. 113-122.
	Tipologija		1.01 Izvirni znanstveni članek
3.	COBISS.SI-ID		22479655
	Naslov	SLO	Modeliranje križanja med genetsko modificirano in konvencionalno koruzo z odkrivanjem enačb
		ANG	Modelling the outcrossing between genetically modified and conventional maize with equation discovery
	Opis	SLO	Razvitih je bilo veliko modelov križanja med posevki, ki pa so večinoma mehanistični, zelo kompleksni in le redko evalvirani na realnih, na terenu pridobljenih podatkih. Naš pristop uporablja realne podatke in ekspertno znanje za razvoj natančnih modelov križanja med GM in konvencionalnimi posevki, pri čemer imajo modeli obliko enačb. Analizirali smo tudi relativni vpliv klimatskih in geografskih parametrov na križanje posevkov ter testirali prenosljivost naših modelov med različnimi podatkovnimi bazami.
		ANG	Many models of outcrossing between crops were developed, but most of them are mechanistic, very complex and rarely evaluated against real data. Our approach uses field measurements and background knowledge to develop accurate equation-based models of the outcrossing between GM and conventional maize fields. We have also analyzed the relative influence of climatic and geographic parameters on the outcrossing and tested the transferability of the equation-based models on different datasets.
	Objavljeno v		TRAJANOV, Aneta, TODOROVSKI, Ljupčo, DEBELJAK, Marko, DŽEROSKI, Sašo. Modelling the outcrossing between genetically modified and

			conventional maize with equation discovery. Ecol. model.. [Print ed.], 2009, vol. 220, no. 8, str. 1063-1072.
	Tipologija		1.01 Izvirni znanstveni članek
	COBISS.SI-ID		22574375
4.	Naslov	SLO	Semantična tehnologija za zajemanje notranjih komunikacij v organizaciji
		ANG	Semantic technology for capturing communication inside an organisation
	Opis	SLO	Komunikaciji med zaposlenimi v neki organizaciji je vir, ki odraža neformalno organizacijsko strukturo. Predlagan je pristop, ki z uporabo tehnologije semantičnega spleta na osnovi elektronske komunikacije med zaposlenimi omogoča izgradnjo neformalne organizacijske strukture. Pristop je pol-avtomatski in združuje različne tehnike odkrivanja zakonitosti v podatkih. Evalvacijo smo naredili na realnih podatkih srednje velike organizacije, kjer smo pokazali kako se pridobljena neformalna organizacijska struktura ujema s formalno organizacijsko strukturo.
		ANG	Informal organizational structure established via a day-to-day communication between the employees reflects informal processes in an organization and in general can substantially differ from the formal organizational structure. Our approach is based on Semantic Technologies and analyzes an internal social network to propose an informal organizational structure. The output is a lightweight ontology. The approach is evaluated on the real-world data set of a mid-size organization.
	Objavljeno v		GROBELNIK, Marko, MLADENIČ, Dunja, FORTUNA, Blaž. Semantic technology for capturing communication inside an organisation. IEEE internet computing, 2009, vol. 13, no. 4, str. 59-66.
	Tipologija		1.01 Izvirni znanstveni članek
	COBISS.SI-ID		22843175
5.	Naslov	SLO	Potencial več ciljnega modeliranja za kartiranje rizika upornosti tal: ponazoritev uporabe na nacionalnem nivoju
		ANG	Potential of multi-objective models for risk-based mapping of the resilience characteristics of soils: demonstration at a national level
	Opis	SLO	Z uporabo več ciljnih regresijskih dreves smo odkrili nove povezave med kazalniki upornosti in prožnosti tal ter dejavniki biološkega in fizičnega stresa tal. Modele za napovedovanje več neodvisnih spremenljivk hkrati smo uporabili za izdelavo karte upornosti tal za celotno Škotsko.
		ANG	With the application of multi-objective regression tree analysis we have determined underlying relationships between soil properties and its resistance and resilience capacity. Models with multiple dependent variables (multi-objective models) for the simultaneous prediction of interdependent resilience and resistance were used to produce maps of resilience characteristics of soils for Scotland.
	Objavljeno v		DEBELJAK, Marko, KOCEV, Dragi, TOWERS, W., JONES, M., GRIFFITHS, Bryan, HALLETT, P. Potential of multi-objective models for risk-based mapping of the resilience characteristics of soils : demonstration at a national level. Soil use manage., 2009, vol. 25, no. 1, str. 66-77.
	Tipologija		1.01 Izvirni znanstveni članek
	COBISS.SI-ID		22691623

8. Najpomembnejši družbeno-ekonomsko relevantni rezultati programske skupine⁹

	Družbeno-ekonomsko relevantni rezultat		
1.	Naslov	SLO	Razvoj sistema GMO-track
		ANG	Development of the GMO-track system
	Opis	SLO	Zaradi velikega števila gensko spremenjenih organizmov (GSO) na tržišču je pomembno vzpostaviti sistem, s pomočjo katerega z najmanjšimi stroški pridemo do zanesljivega odkrivanja in identifikacije GSO v vzorcih hrane, krme in semen. V sodelovanju z Nacionalnim inštitutom za biologijo smo formalno definirali optimizacijski problem odkrivanja prisotnosti in identifikacije GSO in razvili sistem GMOTrack za podporo odločanja pri sledenju GSO, ki kombinacije testov oceni glede na ceno in pokritost GSO.

		Program GMOTrack je javno dostopen na spletni strani http://kt.ijs.si/software/GMOTrack/ .
	ANG	Due to a large number of genetically modified organisms (GMOs) on the market, it is important to develop a system for cost-effective detection and identification of GMOs in food, feed and seed samples. In collaboration with the National institute of biology, we have formally defined the laboratory-level GMO traceability optimization problem and developed a decision support system which evaluates combinations of assays according to their cost and coverage of GMOs. The GMOTrack program is publicly available on the website http://kt.ijs.si/software/GMOTrack/ .
Šifra		F.15 Razvoj novega informacijskega sistema/podatkovnih baz
Objavljeno v		KRALJ NOVAK, Petra, GRUDEN, Kristina, MORISSET, Dany, LAVRAČ, Nada, ŠTEBIH, Dejan, ROTTER, Ana, ŽEL, Jana. GMOTrack : generator of cost-effective GMO testing strategies. J. AOAC Int., 2009, vol. 92, no. 6, str. 1739-1746. http://www.atypon-link.com/AOAC/doi/abs/10.5555/jaoi.92.6.1739 .
Tipologija		1.01 Izvirni znanstveni članek
COBISS.SI-ID		2182735
2. Naslov	SLO	Organizacija konference ECML/PKDD 2009
	ANG	Organization of the ECML/PKDD 2009 conference
Opis	SLO	Uspešno smo organizirali 20. evropsko konferenco o strojnem učenju in principih in praksi odkrivanja znanja v podatkovnih bazah (European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases – ECML/PKDD 2009) na Bledu od 7. do 11. septembra 2009. Gre za enega od svetovno najpomembnejših dogodkov s področja strojnega učenja in odkrivanja znanj v podatkih.
	ANG	We have successfully organized the 20th European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD 2009) that took place in Bled from September 7th to 11th, 2009. It is one of the most important events in the field of machine learning and knowledge discovery.
Šifra		B.01 Organizator znanstvenega srečanja
Objavljeno v		BUNTINE, Wray (ur.), GROBELNIK, Marko (ur.), MLADENIĆ, Dunja (ur.), SHAW-TAYLOR, John (ur.). Machine learning and knowledge discovery in databases : European Conference, ECML PKDD 2009, Bled, Slovenia, September 7-11, 2009 : proceedings. Part I, (Lecture notes in computer science, Lecture notes in artificial intelligence, LNAI 5781). Berlin; Heidelberg: Springer, cop. 2009. XXIX, 756 str., ilustr. ISBN 978-3-642-04179-2. ISBN 3-642-04179-5.
Tipologija		2.31 Zbornik recenziranih znanstvenih prispevkov na mednarodni ali tuji konferenci
COBISS.SI-ID		23373351
3. Naslov	SLO	Strojno prevajanje s poudarkom na južnoslovanskih in balkanskih jezikih
	ANG	Machine translation focused on South Slavic and Balkan languages
Opis	SLO	V tem SEE.ERA EU projektu smo izdelali večjezični vzporedni jezikoslovno označen korpus za slovenščino, srbsščino, romunščino, bolgarščino in grščino, ter angleščino. Razviti korpus smo uporabili za šolanje sodobnega statističnega sistema za strojno prevajanje. Evalvacija je pokazala, da uporaba oblikoskladenjskih oznak signifikantno poveča kvaliteto dobljenih prevodov.
	ANG	This SEE.ERA EU project developed multilingual parallel linguistically annotated corpora, for Slovene, Serbian, Romanian, Bulgarian and Greek (and English). Using these corpora, an advanced statistical machine translation system was trained and evaluated. The results showed that using morphosyntactic information significantly increases translation accuracy.
Šifra		F.04 Dvig tehnološke ravni
Objavljeno v		TUFIŠ, Dan, KOEVA, Svetla, ERJAVEC, Tomaž, GAVRILLIDOU, Maria, KRSTEV, Cvetana. Machine translation focused on South Slavic and Balkan languages. V: MACHAČOVÁ, Jana (ur.), ROHSMANN, Katarina (ur.). Scientific results of the SEE-ERA.NET : pilot joint call. 1st ed. Vienna: Centre for Social Innovation, 2009, str. 37-49.
Tipologija		1.16

		Samostojni znanstveni sestavek ali poglavje v monografski publikaciji	
COBISS.SI-ID		22990887	
4.	Naslov	SLO	Organizacija tretje mednarodne delavnice o strojnem učenju v sistemski biologiji (MLSB 2009)
		ANG	Organization of the Third International Workshop on Machine Learning in Systems Biology (MLSB 2009)
	Opis	SLO	V Ljubljani smo organizirali mednarodno delavnico MLSB 2009, katere so-predsednik je bil S. Džeroski, s šestimi tujimi vabljenimi predavatelji in okrog 100 udeleženci. Recenziran zbornik delavnice, katerega so-urednik je pravtako S. Džeroski, je v tisku.
		ANG	We have organized an international workshop MLSB 2009, co-chaired by S. Džeroski in Ljubljana, with six invited speakers and around 100 attendees. A refereed post-workshop proceedings also co-edited by S. Džeroski is in press.
	Šifra		B.01 Organizator znanstvenega srečanja
	Objavljeno v		DŽEROSKI, Sašo (ur.), GEURTS, Pierre (ur.), ROUSU, Juho (ur.). Machine learning in systems biology : proceedings of the Third International Workshop, September 5-6, 2009, Ljubljana, Slovenia, (Julkaisusarja - Helsingin yliopisto. Tietojenkäsittelytieteiden laitos, report B-2009-1). Helsinki: Department of Computer Science, University, 2009. 180 str., ilustr. ISBN 978-952-10-5699-4.
Tipologija		1.20 Predgovor, spremna beseda	
COBISS.SI-ID		22827559	
5.	Naslov	SLO	Spletni portal VideoLectures.Net in World Summit nagrada OZN
		ANG	Web portal VideoLectures.Net and the World Summit award of the UN
	Opis	SLO	Spletni portal VideoLectures.Net je prejel nagrado World Summit Award na področju e-znanosti in tehnologije (Monterrey, Mehika). Nagrado podeljuje oddelek informacijske družbe Organizacije združenih narodov. Gre za svetovno tekmovanje za priznanje najboljših e-vsebine in ustvarjalnosti.
		ANG	VideoLectures.Net web portal received the World Summit Award in the area of E-Science & Technology (Monterrey, Mexico). This was launched as part of the United Nations Summit on Information Society, a global competition recognizing the best e-Content and Creativity solutions.
	Šifra		E.02 Mednarodne nagrade
	Objavljeno v		http://www.wsis-award.org/winners/winners.wbp JERMOL, Mitja. World summit award and future plans : intervju. [S. l.]: Videlectures.net, 20. okt. 2009. http://videlectures.net/jun09_jermol_int/.
Tipologija		3.25 Druga izvedena dela	
COBISS.SI-ID		23551271	

9. Drugi pomembni rezultati programske skupine¹⁰

Partnerstvo v 18 evropskih projektih 6. in 7.OP in 6 bilateralnih projektih (točka 13.): Sodelovali smo v 28 EU projektih 6. in 7. OP in smo najuspešnejša programska skupina v Sloveniji. Bili smo pobudniki in posredniki vključitve 8 slovenskih podjetij v projekte 6. in 7. OP. V teh evropskih projektih, smo v obdobju 2004-2008 pridobili okrog 7.3 mio EUR, v letu 2009 pa preko 900.000 EUR sredstev. Razvite metodološke in tehnološke rešitve imajo direkten učinek na družbeni in gospodarski razvoj Slovenije in Evrope, naše rešitve že uporabljajo British Telecom, Microsoft Research.

Nagrada za inovativnost za gospodarstvo:

Boštjan Pajntar je prejel nagrado za najbolj inventivne in inovative ideje, ki so s svojo aplikativno vrednostjo zanimive in uporabne tudi v gospodarstvu, podeljeno na Konferenci za prenos tehnologij na Institutu Jožef Stefan. Nagrada je bila podeljena za spletno storitev Searchpoint, ki inovativno nadgrajuje spletni iskalnik Google. Delo je bilo opravljeno pod mentorstvom Marka Grobelnika in doc. dr. Dunje Mladenčić.

English

Partnership in 18 European FP6 and FP7 projects and 6 bilateral projects (see 13.):

In terms of integration into EU projects, we are the most successful research programme in Slovenia. We initiated and mediated the inclusion of 8 Slovene companies in EU projects. We have collaborated in 28 FP6 and FP7 projects. Work on these projects in 2004-2008 brought us about 7.3 MIL EUR, and in 2009 more than 900,000 EUR in addition. Solutions developed in these projects are, for instance, used by British Telecom, Microsoft Research.

Innovations for economy award:

Boštjan Pajntar was awarded an industry innovation prize at the Technology Transfer Conference at the Jožef Stefan Institute. The prize was awarded for the Searchpoint Web search engine which upgrades the Google search engine in a novel way. This work was performed under the supervision of Marko Grobelnik and Dr. Dunja Mladenić.

10. Zaključena mentorstva članov programske skupine pri vzgoji kadrov¹¹

Vrsta izobraževanja	Število mentorstev	Od tega mladih raziskovalcev
- magisteriji	2	
- doktorati	2	
- specializacije		
Skupaj:	4	0

11. Zaposlitev vzgojenih kadrov po usposabljanju

Organizacija zaposlitve	Število doktorjev	Število magistrrov	Število specializantov
- univerze in javni raziskovalni zavodi	3	1	
- gospodarstvo		2	
- javna uprava			
- drugo			
Skupaj:	3	3	0

12. Vključenost raziskovalcev iz podjetij in gostovanje raziskovalcev, podoktorandov ter študentov iz tujine, daljše od enega meseca

Sodelovanje v programski skupini	Število
- raziskovalci-razvijalci iz podjetij	1
- uveljavljeni raziskovalci iz tujine	1
- podoktorandi iz tujine	1
- študenti, doktorandi iz tujine	1
Skupaj:	4

13. Vključenost v raziskovalne programe Evropske unije in v druge mednarodne raziskovalne in razvojne programe ter drugo mednarodno sodelovanje v letu 2009¹²

EU projekti:

- IMAGINATION: Image-based Navigation in Multimedia Archives (6.OP FP6-034626) doc. Dr. Dunja Mladenić, 1.5.2006-30.4.2009
- SMART: Statistical Multilingual Analysis for Retrieval and Translation, 1.10.2006-30.9.2009
- NeOn: Lifecycle Support for Networked Ontologies (6.OP IP-027595) Marko Grobelnik, doc.

dr. Dunja Mladenec, 1.3.2006-28.2.2010

- SWING: Semantic Web Services Interoperability for Geospatial Decision Making (6.OP FP6-026514) Marko Grobelnik, doc. dr. Dunja Mladenec, 1.3.2006-28.2.2009
- TAO: Transitioning Applications to Ontologies (6.OP FP6-026460) Marko Grobelnik, doc. Dr. Dunja Mladenec, 1.3.2006-28.8.2009
- E.E.T Pipeline: (6.OP STREP 037260) "European Embryonal Tumor Pipeline" prof. dr. Sašo Džeroski 1.1.2007-31.12.2009
- Healththreats: Integrated Decision Support System for HEALT THREATS and crises management, (e-Health 2006203), prof. Dr. Nada Lavrač, 31.5.2007-30.5.2010
- Co-Extra: "GM and non-GM supply chains: their CO-Existence and Traceability (6.OP, IP-007158), prof. dr. Marko Bohanec 1.4.2007.-30.9.2009
- COIN: "Collaboration and Interoperability for networked enterprises" (7.OP-216256) doc. dr. Dunja Mladenec, Marko Grobelnik 1.1.2008-31.12.2011
- EURIDICE: European Inter-Disciplinary Research on Intelligent Cargo for Efficient, Safe and Environment-friendly Logistics (7.OP, IP-216271), doc. dr. Dunja Mladenec, Marko Grobelnik, 1.2.2008-31.1.2011
- ACTIVE: Enabling the Knowledge Powered Enterprise (7.OP, IP-215040), doc. dr. Dunja Mladenec, Marko Grobelnik, 1.3.2008-28.2.2012
- BISON: Bisociation Networks for Creative Information Discovery (7.OP 211892) prof.dr. Nada Lavrač, 1.6.2008 – 31.5.2011
- PASCAL2: Pattern Analysis, Statistical Modelling and Computational Learning 2 (7.OP NoE-216886) 1.3.2008-28.2.2013
- MONDILEX: Conceptual Modelling of Networking of Centres for High-Quality Research in Slavic Lexicography and their Digital Resources (7.OP, CA 211938), dr. Tomaz Erjavec, 1.4.2008 – 31.3.2010
- PHAGOSYS: Systems biology of phagosome formation and maturation - modulation by intracellular pathogens (7.OP 223451, Health) 1.11.2008 – 31.10.2011
- FlaReNet: Fostering Language Resources Network (7.OP eContent, ECP-2007-LANG-617001), dr. Tomaz Erjavec, 1.9.2008 – 31.8.2011
- VIDi: Visualising the impact of the legislation by analysing public discussions using statistical means (E-participation, EP-08-01-014), Marko Grobelnik, doc. dr. Dunja Mladenec, 1.1.2009-31.12.2010
- GENDERA: Gender Debate in the European Research Area (7.OP, CSA 244499), doc. dr. Dunja Mladenec, 1.11.2009 - 30.10.2011

Ostali projekti:

- LAND-EME: Sustainable Crop System in Scotland (naročnik SCRI, Living Technology, Škotska), prof. dr. Marko Debeljak, 31.3.-31.12.2009)

Mednarodno sodelovanje:

- Rudarjenje heterogenih virov podatkov v bioinformatiki in biomedicine (bioinfuse) (projekt znanstveno-tehnološkega sodelovanja med Republiko Slovenijo ter Republiko Češko) (prof. dr. Nada Lavrač) 2008-2009
- Japonsko-slovenski viri za učence japonščine (sofinanciranje kratkoročnih obiskov slovenskih raziskovalcev na Japonskem) doc. dr. Tomaž Erjavec 2008-2009
- Induktivno učenje pravil (projekt znanstveno-tehnološkega sodelovanja med Republiko Slovenijo ter Republiko Hrvaško) (prof. dr. Nada Lavrač) 2009-2010
- Definicija sin taktično-semantične strukture slovenskega glagola v modelu češkega leksikona Vallex (projekt koordiniran s projektom Barrande) (projekt znanstvenoraziskovalnega projekta s Francosko republiko Program PROTEUS) doc. dr. Tomaž Erjavec, 2009-2010
- Vpliv gensko modificirane korne na njeno konvencionalno pridelavo: system svetovanja za zagotavljanje soobstoja (projekt znanstvenoraziskovalnega projekta s Francosko republiko Program PROTEUS) prof. dr. Marko Debeljak, 2009-2010
- Sistem za podporo odločanja pri zagotavljanju varne pridelave povrtnin na vrtnih tleh urbanih okolji (Gardentox) projekt znanstvenoraziskovalnega projekta s Francosko republiko Program INRA, prof. dr. Marko Debeljak, 2009-2010

14. Vključenost v projekte za uporabnike, ki so v letu 2009 potekali izven financiranja ARRS¹³

- VIZIPIN (naročnik Iskra Zaščite), Marko Grobelnik, 1.10.2007-30.9.2009
- GMOtrack Prototype Decision Support System (Prototip sistema za podporo odločanja) (naročnik Nacionalni Inštitut za biologijo), prof. dr. Nada Lavrač, dr. Petra Kralj Novak, 26.2.-30.4.2009

- SM-RIS - Mladinska mreža razvoja raziskovalnih vrednot mladih (Ministrstvo za visoko šolstvo, znanost in tehnologijo), doc.dr. Dunja Mladenec, 1.5.2009 - 31.8.2012
- Sodelovanje pri izdelavi spletne strani SBL1 (naročnik ZRC SAZU), doc.dr. Tomaž Erjavec, 1.10.2009 - 30.11.2009
- Vpliv prenosa genov, genske raznovrstnosti in raznolikosti ter tehnologije pridelovanja oljne ogrščice na soobstoj in izpolnjevanje pogojev trajnostne pridelave ter razvoj metod za sledljivost (Ministrstvo za kmetijstvo), prof.dr. Marko Debeljak, 2008-2010
- Analiza in scenarij razvoja in rabe gozdov v Sloveniji, (Ministrstvo za kmetijstvo), prof.dr. Marko Debeljak 2008-2011

15. Opisno poročilo o porabi sredstev, pridobljenih na podlagi dodatnega letnega sofinanciranja zaradi mednarodnega sodelovanja (sodelovanje v projektih okvirnih programov EU) v letu 2009¹⁴

Sredstva smo porabili v naslednje namene:

- Nakup opreme (strežniki, osebni računalniki, prenosni računalniki)
- Del sredstev smo prenesli v 2010 za materialne stroške in amortizacijo (kratkoročno odloženi prihodki)

16. Opisno poročilo o porabi sredstev, pridobljenih na podlagi dodatnega letnega sofinanciranja zaradi povečane raziskovalne aktivnosti v Republiki Sloveniji in EU v letu 2009¹⁵

Sredstva smo porabili za financiranje dodatnega raziskovalnega dela.

17. Opisno poročilo o porabi sredstev, pridobljenih na podlagi dodatnega letnega sofinanciranja zaradi sodelovanja v pedagoškem procesu v letu 2009¹⁶

Sredstva smo porabili za financiranje mentorskega dela raziskovalcev, vključenih v pedagoške procese.

V sodelovanju z Mednarodno podiplomsko šolo Jožefa Stefana (S. Džeroski kot mentor doktorantke Anete Ivanovske, v tem procesu je sodeloval tudi M. Debeljak) smo razvijali metode strojnega učenja, natančnejše odkrivanja enačb, ter jih uporabili za modeliranje procesa križanja med gensko spremenjenimi in konvencionalnimi posevki. Pri tem smo uporabili podatke s terenskih poskusov v Sloveniji in Nemčiji. Analiziral smo tudi relevanten vpliv različnih dejavnikov (vetra in razdalje) na stopnjo križanja.

V sodelovanju z Univerzo v Novi Gorici (M. Bohanec kot mentor mlade raziskovalke Mojce Arsenijević) smo sodelovali pri razvoju odločitvenih modelov na področju ekologije ter na področju rudarjenja tekstovnih podatkov za iskanje skritih zakonitosti v medicinski literaturi (T. Urbančič in B. Cestnik kot somentorja doktorske disertacije Ingrid Petrič, v tem delu je sodelovala tudi N. Lavrač).

18. Ocena tehnološke zrelosti rezultatov programa in možnosti za njihovo implementacijo v praksi¹⁷

Nekaj programskih rešitev programske skupine, razvitih v letu 2009, je zrelih za trženje in uporabo v praksi. Mednje sodijo program GMOtrack, ki je javno dostopen na spletni strani <http://kt.ijs.si/software/GMOtrack/>, je v redni rabi na Nacionalnem inštitutu za biologijo za sledenje in identifikacijo gensko spremenjenih organizmov v hrani, krmi in semenih.

Spletna storitev Searchpoint, dostopna na <http://searchpoint.ijs.si>, ki inovativno nadgrajuje spletni iskalnik Google, je javno dostopna in uporabna za preiskovanje spletnih vsebin.

Program DEXi (verzija 3.02 je bila razvita oktobra 2009) je javno dostopen na spletni strani <http://kt.ijs.si/MarkoBohanec/dexi.html> in je namenjen za podporo pri reševanju večparametrskih odločitvenih problemov.

Na naslovu <http://nl.ijs.si/jos/analyse/> je dostopen oblikoskladenjski označevalnik in lematizator za slovenska besedila. Spletni servis se uporablja za izdelavo označenih korpusov slovenskega jezika na Filozofski fakulteti Univerze v Ljubljani, Filozofski fakulteti Univerze v Mariboru in Fakulteti za humanistiko Univerzi v Novi Gorici; načrte za uporabo so izkazala tudi slovenska internetna podjetja.

19. Ocenite, ali bi doseženi rezultati v okviru programa lahko vodili do ustanovitve spin-off podjetja, kolikšen finančni vložek bi zahteval ta korak ter kakšno infrastrukturo in opremo bi potrebovali.

Možnost ustanovitve spin-off podjetja	☒ DA ☐ NE
Potrebni finančni vložek	15.000 EUR
Ocena potrebne infrastrukture in opreme¹⁸	Za trženje programske rešitve Searchpoint, za katero je Poštjan Pajntar v letu 2009 prejel nagrado za inovacije, je v postopku ustanavljanje spin-off podjetja. Finančni vložek je sorazmerno majhen (okrog 15.000 EUR). Potrebna oprema je računalnik in programska oprema.

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 3., 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 letnega poročila seznanjeni in se strinjajo vsi izvajalci programa

Podpisi:

vodja raziskovalnega programa	in	zastopnik oz. pooblaščen oseba RO
Nada Lavrač		Institut "Jožef Stefan"

Kraj in datum:

Oznaka poročila: ARRS-RPROG-LP-2009/117

¹ Napišite kratek povzetek vsebine oziroma opisa raziskovalnega programa. Največ 3.000 znakov vključno s presledki (približno pol strani, velikosti pisave 11). [Nazaj](#)

² Pomen raziskovalnih rezultatov za razvoj znanosti in za razvoj Slovenije bo objavljen na spletni strani: <http://sicris.izum.si> [Nazaj](#)

³ Največ 4.000 znakov vključno s presledki [Nazaj](#)

⁴ Največ 4.000 znakov vključno s presledki. Opišite tudi morebiten vpliv na gospodarstvo in okolje. [Nazaj](#)

⁵ Napišite kratko vsebinsko poročilo, kjer boste predstavili raziskovalno hipotezo in opis raziskovanja. Navedite ključne ugotovitve, znanstvena spoznanja ter rezultate in učinke raziskovalnega programa. Največ 21.000 znakov vključno s presledki (približno tri in pol strani, velikosti pisave 11). [Nazaj](#)

⁶ Največ 3000 znakov vključno s presledki (približno pol strani, velikosti pisave 11). [Nazaj](#)

⁷Samo v primeru bistvenih odstopanj in sprememb od predvidenega programa raziskovalnega programa, kot je bil zapisan v predlogu raziskovalnega programa. Največ 3.000 znakov vključno s presledki (približno pol strani, velikosti pisave 11). [Nazaj](#)

⁸ Navedite največ pet najpomembnejših znanstvenih rezultatov programske skupine, ki so nastali v času trajanja programa v okviru raziskovalnega programa, 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/>.

PRIMERR (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., PREMZL, 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](#)

⁹ Navedite največ pet najpomembnejših družbeno-ekonomsko relevantnih rezultatov programske skupine, ki so nastali v času trajanja programa v okviru raziskovalnega programa, 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, 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](#)

¹⁰Navedite rezultate raziskovalnega programa v primeru, da katerega od rezultatov ni mogoče navesti v točkah 6 ali 7 (npr. ker se ga v sistemu COBISS ne vodi). Največ 2.000 znakov vključno s presledki. [Nazaj](#)

¹¹Za raziskovalce, ki niso habilitirani, so pa bili mentorji mladim raziskovalcem, se vpiše ustrezen podatek samo v stolpec MR [Nazaj](#)

¹²Navedite oziroma naštejite konkretne projekte. Največ 12.000 znakov vključno s presledki. [Nazaj](#)

¹³Navedite konkretne projekte, kot na primer: industrijski projekti, projekti za druge naročnike, državno upravo, občine ipd. in ne sodijo v okvir financiranja pogodb ARRS. Največ 9.000 znakov vključno s presledki. [Nazaj](#)

¹⁴Izpolnijo izvajalke raziskovalnega programa, prejemnice sredstev iz naslova dodatnega letnega sofinanciranja raziskovalnega programa zaradi mednarodnega sodelovanja (sodelovanja v projektih okvirnih programov Evropske unije, na podlagi 13. čl. Pravilnika o ocenjevanju in financiranju raziskovalnih in infrastrukturnih programov). Izvajalka, ki je na podlagi pogodbe prejela sredstva iz navedenega naslova, vsebinsko opiše porabo prejetih sredstev pri izvajanju tega raziskovalnega programa. V primeru, da so bile v okviru raziskovalnega programa prejemnice sredstev različne izvajalke, vsaka pripravi vsebinsko poročilo za svoj delež pogodbenih sredstev. Vodja raziskovalnega programa poskrbi, da je vsebinsko poročilo, ločeno za vsako izvajalko, vključeno v navedeno točko poročila. Največ 2.000 znakov vključno s presledki. [Nazaj](#)

¹⁵Izpolnijo izvajalke raziskovalnega programa, prejemnice sredstev iz naslova dodatnega letnega sofinanciranja raziskovalnega programa zaradi povečane raziskovalne aktivnosti v Republiki Sloveniji in EU (na podlagi 15a. čl. Pravilnika o ocenjevanju in financiranju raziskovalnih in infrastrukturnih programov). Izvajalka, ki je na podlagi pogodbe prejela sredstva iz navedenega naslova, vsebinsko opiše porabo prejetih sredstev pri izvajanju tega raziskovalnega programa. V primeru, da so bile v okviru raziskovalnega programa prejemnice sredstev različne izvajalke, vsaka pripravi vsebinsko poročilo za svoj delež pogodbenih sredstev. Vodja raziskovalnega programa poskrbi, da je vsebinsko poročilo, ločeno za vsako izvajalko, vključeno v navedeno točko poročila. Največ 2.000 znakov vključno s presledki. [Nazaj](#)

¹⁶Izpolnijo izvajalke raziskovalnega programa, prejemnice sredstev iz naslova dodatnega letnega sofinanciranja raziskovalnega programa zaradi sodelovanja v pedagoškem procesu (na podlagi 14. čl. Pravilnika o ocenjevanju in financiranju raziskovalnih in infrastrukturnih programov). Izvajalka, ki je na podlagi pogodbe prejela sredstva iz navedenega naslova, vsebinsko opiše porabo prejetih sredstev pri izvajanju tega raziskovalnega programa. V primeru, da so bile v okviru raziskovalnega programa prejemnice sredstev različne izvajalke, vsaka pripravi vsebinsko poročilo za svoj delež pogodbenih sredstev. Vodja raziskovalnega programa poskrbi, da je vsebinsko poročilo, ločeno za vsako izvajalko, vključeno v navedeno točko poročila. Največ 2.000 znakov vključno s presledki. [Nazaj](#)

¹⁷Opišite možnosti za uporabo rezultatov v praksi. Opišite izdelke oziroma tehnologijo in potencialne trge oziroma tržne niše, v katere sodijo. Ocenite dodano vrednost izdelkov, katerih osnova je znanje raz vito v okviru programa, oziroma dodano vrednost na zaposlenega, če jo je mogoče oceniti (na primer v primerih, ko je rezultat izboljšava obstoječih tehnologij oziroma izdelkov). Največ 3.000 znakov vključno s presledki. [Nazaj](#)

¹⁸Največ 1.000 znakov vključno s presledki. [Nazaj](#)

Obrazec: ARRS-RPROG-LP/2009 v1.00a