

Oznaka poročila: ARRS-RPROG-ZP-2015/1



ZAKLJUČNO POROČILO O REZULTATIH RAZISKOVALNEGA PROGRAMA

(za obdobje 1. 1. 2009 - 31. 12. 2014)

A. PODATKI O RAZISKOVALNEM PROGRAMU

1. Osnovni podatki o raziskovalnem programu

Šifra programa	P2-0103
Naslov programa	Tehnologije znanja Knowledge Technologies
Vodja programa	8949 Nada Lavrač
Obseg raziskovalnih ur (vključno s povečanjem financiranja v letu 2014)	62537
Cenovni razred	
Trajanje programa	01.2009 - 12.2014
Izvajalci raziskovalnega programa (javne raziskovalne organizacije - JRO in/ali RO s koncesijo)	106 Institut "Jožef Stefan" 1540 Univerza v Novi Gorici
Raziskovalno področje po šifrantu ARRS	2 TEHNIKA 2.07 Računalništvo in informatika
Družbeno-ekonomski cilj	06. Industrijska proizvodnja in tehnologija
Raziskovalno področje po šifrantu FOS	1 Naravoslovne vede 1.02 Računalništvo in informatika

B. REZULTATI IN DOSEŽKI RAZISKOVALNEGA PROGRAMA

2. Povzetek raziskovalnega programa¹

SLO

Program je nadaljeval uspešne raziskave prejšnjega programa Tehnologije znanja, ki ga je ARRS preteklem obdobju financiranja ocenil kot najboljši program na področju računalništva ter identificiral kot najuspešnejšo programsko skupino pri zagotavljanju evropskih raziskovalnih

sredstev.

Področje dela raziskovalnega programa Tehnologije znanja so napredne informacijske tehnologije za zajemanje, shranjevanje in upravljanje znanja, ki so praktično uporabne za razvoj informacijske in na znanju temelječe družbe. Raziskave so bile usmerjene v 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. V teku programa smo preučevali tudi novejša področja raziskav, kot so Splet 2.0, novi mediji in e-znanost. Poleg razvoja tehnologij znanja smo razvijali tudi aplikacije teh tehnologij na področju znanosti o okolju in upravljanja z okoljem, medicine in zdravstvenega varstva, biomedicine in genetike, ekonomije in tržništva.

V zadnjih šestih letih smo sodelovali oz. sodelujemo v 52 projektih 7. OP, vključno z našo koordinacijo projektov MAESTRA, NRG4Cast, TOPOSYS in XLike ter nedavno vključitvijo v prestižni FET Flagship "Human Brain Project". Objavili smo številne članke v priznanih revijah in napisali monografijo, ki je izšla pri založbi Springer in za katero je vodja programske skupine v 2013 prejela Zoisovo priznanje. ARRS je v letih 2011-2013 izpostavila štiri naše prispevke kot izjemne znanstvene dosežke slovenske znanosti. Prejeli smo "World Summit Award" za portal VideoLectures.net, bili smo dvakratni prejemniki "Google Digital Humanities Research Award" in prejeli pet nagrad za najboljše prispevke na konferencah. Raziskovalci programa smo uspešni tudi kot mentorji podiplomcev, saj je pod našim mentorstvom v tem programskem obdobju doktoriralo 24 študentov.

ANG

The program continued successful research of the previous program Knowledge Technologies, which was evaluated by the ARRS funding agency as the best Slovene Computer Science program and recognized as the most successful Slovenian research program in terms of established European research collaborations and funding.

The program Knowledge Technologies performs research in advanced information technologies, aimed at acquiring, storing and managing knowledge to be used in the development of knowledge-based applications. Research was focused on 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. In the course of the program we also addressed new research areas, such as Web 2.0, new media and e-science. Apart from research on knowledge technologies we also developed applications in environmental sciences and ecology, medicine and health care, biomedicine and genetics, economy and marketing.

In the last six years we have collaborated in 52 FP7 projects, including our coordination of FP7 projects MAESTRA, NRG4Cast, TOPOSYS and XLike, and recent inclusion in the prestigious FP7 FET Flagship "Human Brain Project". We published numerous research papers in top-ranked journals in the area, and wrote a Springer monograph for which the program group leader received the national Zois recognition award in 2013. In 2011-2013, four of our contributions were recognized by the ARRS agency as outstanding Slovenian research achievements. We received the "World Summit Award" for the VideoLectures.net portal, twice the "Google Digital Humanities Research Award", and were awarded 5 best conference paper awards. The program group researchers are successful also as supervisors of postgraduate students: 24 students obtained PhD degrees under our supervision in the last six years.

3. Poročilo o realizaciji predloženega programa dela na raziskovalnem programu, (vključno s predloženim dopolnjenim programom dela v primeru povečanja financiranja raziskovalnega programa v letu 2014)²

SLO

Na področju **inteligentne analize podatkov** smo razvili več metod rudarjenja podatkov in jih uporabili v raznovrstnih aplikacijah. Razvili smo metode za učenje napovednih in opisnih pravil ter pri založbi Springer izdali obširno monografijo Foundations of Rule Learning. Razvili smo metode za odkrivanje podskupin z uporabo predznanja v obliki ontologij SEGS, g-SEGS,

Hedwig in orodje SDM-X. Za potrebe biomedicinskih aplikacij smo razvili metodologijo SegMine, ki kombinira znanje iz ontologij genov in obsežnega grafa biomedicinskih relacij BioMine. Razvili smo platformo za rudarjenje podatkov Orange4WS in platformo za ansambelsko odkrivanje šuma in osamelcev NoiseRank. Razvili smo platformo za avtomatsko modeliranje dinamičnih sistemov iz podatkov ter domenskega predznanja, ki podpira uporabo različnih metod za optimizacijo parametrov in različnih kriterijskih funkcij. Platforma je bila uporabljena v več evropskih projektih za modeliranje vodnih ekosistemov, hidroloških procesov ter modeliranje izpiranja hranil v povodjih. Za napovedovanje strukturiranih vrednosti smo razvili metode napovednega razvrščanja za večciljno klasifikacijo in regresijo, večoznačno klasifikacijo ter analizo kratkih časovnih vrst. Največjo napovedno moč imajo ansambl dreves in pravil, ki jih razvijamo tudi v okviru evropskega projekta MAESTRA, ki ga koordiniramo na IJS. Metode razvijamo in uporabljamo na področjih znanosti o okolju ter znanosti o življenju, npr. za povezovanje okoljskih spremenljivk s strukturo biote, ocenjevanje stanja okolja preko podatkov daljinskega zaznavanja, označevanje slik v medicini ter napovedovanje funkcij genov v bakterijskih genomih. Za potrebe gradnje strukturnih modelov smo razvili metodo Bio3Graph, ki podpira gradnjo bioloških omrežij z luščenjem bioloških relacij iz znanstvenih člankov.

Z dodatnim financiranjem ARRS smo na področju strojnega učenja v sistemskih znanostih odsegli dva pomembna prispevka. Razvili smo metode za neparametrično modeliranje dinamičnih sistemov v diskretnem času, ki temeljijo na ansamblih mehkih modelnih dreves, ter jih uporabili na več problemih s področja vodenja sistemov. Razvili smo tudi metode za učenje ansamblov modelov dinamičnih sistemov v obliki navadnih diferencialnih enačb in jih uporabili na več problemih s področja modeliranje vodnih ekosistemov.

Na področju **analize besedil, spleta in heterogenih informacijskih omrežij** smo v okviru več evropskih projektov razvili orodja za analizo velikih količin dinamičnih in heterogenih virov finančnih informacij ter za rudarjenje tokov tekstovnih podatkov (finančnih novic, blogov, tvitov) v realnem času. S portalom NewStream spremljamo več kot 170 finančnih spletnih virov, kjer dnevno zajamemo okoli 40.000 novic. V projektu FIRST smo razvili pristope za analizo sentimenta o finančnih produktih, analizo ugleda finančnih institucij in ugotavljanje spletnih zavajanj. V projektu FOC smo iz tokov tekstovnih podatkov izluščili indikatorje za napovedovanje finančnih kriz. V projektu BISON smo razvili sistem CrossBee za odkrivanje novih meddomenskih povezav. Razvite tehnologije za analizo sentimenta v realnem času na omrežju Twitter smo v sodelovanju s podjetjem Gama System uporabili za spremljanje slovenskih predsedniških in bolgarskih parlamentarnih volitev.

Na področju **jezikovnih tehnologij in digitalne humanistike** smo izdelali številne temeljne digitalne jezikovne vire in metode za avtomatsko analizo jezika. Uredili smo četrto različico večjezičnih virov MULTEXT-East, izdelali ročno označene korpus slovenskega jezika JOS ter prve večje vire starejše slovenščine IMP (1584-1919) z digitalno knjižnico, ročno označenim korpusom in slovarjem. Sodelovali smo pri izdelavi nove generacije slovenskih referenčnih korpusov Gigafida (milijarda besed), korpusa govorjenega jezika Gos, kot tudi pri izdelavi številnih digitalnih izdaj slovenskega slovstva, npr. registra rokopisov, znanstvenokritičnih izdaj, slovenskega biografskega leksikona itd. Razvili smo metode, modele in orodja za avtomatsko analizo jezika, npr. večjezični lematizator LemmaGen in orodje za posodabljanje starejših slovenskih besedil. Na splet smo postavili visokozmogljiv konkordančnik, ki ponuja preko 20 različnih korpusov. Vsi izdelani viri in metode so prosto dostopni na spletu, pri čemer je večina virov tudi odprto dostopnih pod licenco CC, kar omogoča njihovo široko uporabo in vgradnjo v nove aplikacije s področja jezikovnih tehnologij.

Na področju **podpore pri odločanju** smo razvili nove metode za rangiranje alternativ v kvalitativnih večparametrskih modelih, ki temeljijo na uporabi kopul in izboljšujejo ločljivost modelov. Nadaljevali smo z razvojem, dopolnjevanjem in vzdrževanjem programa DEXi za simbolično večparametrsko modeliranje. Metode smo implementirali v odprtokodni knjižnici JDEXi in kot del mednarodnega sistema DecisionDeck. Najpomembnejše aplikacije so bile na področjih vrednotenja projektov, delotokov podatkovnega rudarjenja, vodnih virov, uporabe gensko spremenjenih rastlin, ekoturizma, kmetijstva (ocenjevanje kakovosti prsti in

primernosti pridelovalnih sistemov, izbire križancev hmelja, zmanjševanje onesnaženja površinskih voda s fitofarmacevtskimi sredstvi) in bančništva (odkrivanje zlorab ter ocenjevanje ugleda bank).

Na področju **statističnega modeliranja podatkov in semantičnega spleta** smo sodelovali v evropskih mrežah odličnosti PASCAL in PASCAL2. V projektu VIDI smo razvili pristope za vizualizacijo dinamičnih tekstovnih podatkov in svetovanje uporabnikom. V evropskih projektih METANET, MultilingualWeb in LT-Web smo razvili metode in servise za avtomatsko gradnjo večjezičnih korpusov, sodelovali pri razvoju standardov in praktičnih primerov orodji, ki podpirajo gradnjo, lokalizacijo in uporabo večjezičnih spletnih informacij. V projektu SMART smo razvili metode za medjezično klasifikacijo dokumentov, v projektu XLike, ki ga koordiniramo na IJS, pa tehnologije za medjezično povezovanje novic v realnem času, jezikovno analizo ter medjezikovno semantično anotacijo in povezovanje dokumentov. V okviru evropskih projektov TAO, SWING, NeOn, ACTIVE in ENVISION smo sodelovali pri razvoju metodologije za analizo čezmodalnih podatkov, orodja za semantično anotacijo, razširitvi aplikacije SearchPoint za spletno iskanje z združevanjem zadetkov v smiselne sklope, modeliranju neformalnih procesov velikih organizacij z uporabo modeliranja kontekstov, analizi tokov podatkov in razvoju metod za semantično bogatenje senzorskih podatkov.

4. Ocena stopnje realizacije programa dela na raziskovalnem programu in zastavljenih raziskovalnih ciljev³

SLO

Programska skupina je bistveno preseгла zastavljene cilje znanstvene odličnosti, saj imamo proaktivne sodelavce, aktivno mednarodno sodelovanje in širok nabor zanimivih tematik za naslednje šestletno raziskovalno obdobje. Navajamo le nekaj kazalcev izjemne uspešnosti.

Objave: Poleg monografije "Foundations of Rule Learning" (Springer, 2012) smo dela objavili v revijah z najvišjim faktorjem vpliva na področjih strojnega učenja, podatkovnega rudarjenja in upravljanja znanja (JMLR, MLJ, DMKD, Knowledge-Based Systems, ...) in drugih vrhunskih znanstvenih revijah (Nature Methods, PLOS Computational Biology, PLOS One, ...).

Nagrade in priznanja: Leta 2013 je vodja programa prejela Zoisovo priznanje za pomembne dosežke v znanosti. V letih 2011 - 2014 je ARRS štiri naša dela uvrstila v izbor izjemnih znanstvenih dosežkov. Na področju e-znanosti in tehnologije smo prejeli "World Summit Award" za portal Videolectures.net. Za delo na jezikovnih modelih starejše slovenščine smo prejeli "Google Digital Humanities Research Award". Na Konferenci za prenos tehnologij smo prejeli nagrado za inovativnost v gospodarstvu za orodje OntoGen.

Nagrajeni prispevki in referati: Prejeli smo priznanja za prispevke in referate na doktorskem simpoziju konference EDBT/ICDT 2011 v Uppsali, na 7th International Conference on Machine Learning and Data Mining (MLDM 2011) v Newarku, na konferenci 15th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2011) v Shenzhenu, na 26. e-konferenci na Bledu ter za demo na konferenci Extended Semantic Web Conference (ESWC 2011) na Kreti.

Predsedstvo programskih in organizacijskih odborov konferenc: European Conference on Machine Learning and Data Mining (ECML/PKDD 2009) na Bledu, 13th European Conference on Artificial Intelligence in Medicine (AIME 2011) na Bledu, 22nd International Conference on Inductive Logic Programming (ILP 2012) v Dubrovniku, v letošnjem letu pa organiziramo 5th International Conference on Computational Creativity (ICCC 2014) v Ljubljani ter 17th International Conference on Discovery Science in 25th International Conference on Algorithmic Learning Theory (DS/ALT 2014) na Bledu.

Poučevanje in mentorstvo: Poučujemo na Univerzi v Ljubljani, Univerzi v Novi Gorici, Univerzi na Primorskem, Fakulteti za informacijske študije v Novem mestu, Univerzi v Zagrebu, Univerzi v Gradcu, Katoliški univerzi v Leuvnu. Na Mednarodni podiplomski šoli Jožefa Stefana na programu IKT izvajamo modul Tehnologije znanja. Pod mentorstvom raziskovalcev programske skupine je v zadnjih šestih letih doktoriralo 24 študentov.

Mednarodno sodelovanje: Po naših podatkih smo najuspešnejša slovenska programska skupina pri pridobivanju sredstev evropskih projektov. Bili smo partnerji 52 projektov 7. OP, 3 drugih EU projektov in 11 bilateralnih projektov. Smo koordinatorji štirih 7. OP projektov MAESTRA, NRG4Cast, TOPOSYS in XLike, priključili pa smo se tudi prestižnemu 7. OP FET Flagship projektu "Human Brain".

5.Utemeljitev morebitnih sprememb programa raziskovalnega programa oziroma sprememb, povečanja ali zmanjšanja sestave programske skupine v letu 2014⁴

SLO

V programu raziskovalnega programa v letu 2014 ni bilo bistvenih sprememb.

Iz skupine se je izključil dr. Ivica Slavkov (šifra raziskovalca: 32908), ker je z dnem 28.2.2014 odšel v tujino.

V skupino sta se vključila dr. Darko Cherepnalkoski šifra raziskovalca 29890 in dr. Dragana Miljković (šifra raziskovalke 36912).

Dr. Dragana Miljković je začela z načrtovanjem in implementacijo semantičnih grafičnih modelov, ki temeljijo na asociacijah. Razvijala je mehanizme in mere za učenje osnovnih asociacij med idejami ali koncepti in razvijala metode, ki delujejo na teh predstavitev konceptov.

Dr. Darko Cherepnalkoski je začel s sodelovanjem pri razvoju platforme ClowdFlows. Prilagodil jo je za delo z velikimi količinami podatkov (Big Data) v oblaku (tehnologija Hadoop) in analizi tokov tekstovnih podatkov v realnem času (tehnologija Twitter Storm).

6.Najpomembnejši znanstveni rezultati programske skupine⁵

Znanstveni dosežek			
1.	COBISS ID	26327591	Vir: COBISS.SI
	Naslov	SLO	Osnove učenja pravil
		ANG	Foundations of rule learning
	Opis	SLO	Znanstvena monografija "Foundations of Rule Learning" (Springer 2012, 334 strani), katere vodilna avtorica je prof. dr. Nada Lavrač, je plod večletnega raziskovalnega dela na področju strojnega učenja. Monografija poda celovit pregled raziskovalnega dela s področja učenja pravil, predstavi pa tudi osnove, tehnike in izbrane aplikacije učenja pravil s področja klasičnega strojnega učenja in sodobnega podatkovnega rudarjenja. Knjiga se lahko uporablja tudi kot učbenik strojnega učenja in podatkovnega rudarjenja. Deli knjige so prosto dostopni na spletni strani založbe Springer http://link.springer.com/book/10.1007/9783540751977/ .
		ANG	Monograph "Foundations of Rule Learning" (Springer 2012, 334 pages), co-authored by Prof. Nada Lavrač as the leading author, is the result of several years of research work in the area of machine learning. The monograph presents the fundamentals of rule learning as investigated in classical machine learning and modern data mining. The book can be used as a comprehensive reference to research in the field of inductive rule learning and can also serve as a textbook for teaching machine learning. Parts of the book are available on the Springer website http://link.springer.com/book/10.1007/9783540751977/ .
	Objavljeno v	Springer; 2012; XVII, 334 str.; A": 1;A': 1; Avtorji / Authors: Fürnkranz Johannes, Gamberger Dragan, Lavrač Nada	
Tipologija		2.01	Znanstvena monografija

2.	COBISS ID		26510887	Vir: COBISS.SI
	Naslov	SLO	Računsko napovedovanje funkcij proteinov	
		ANG	Computational protein function prediction	
	Opis	SLO	Razvili smo novo metodo za računsko napovedovanje funkcij genov oz. proteinov, ki jih geni kodirajo. (i) Metoda, ki temelji na principih homologije in filogenetskih profilov in uporablja ansamble dreves za hierarhično klasifikacijo, je bila objavljena v reviji PLOS Computational Biology. Bila je eksperimentalno ovrednotena v laboratoriju, pri čemer se je pokazalo, da omogoča realistične ocene napovedne moči in se lahko uporabi za usmerjanje bioloških poskusov za določanje funkcije genov. (ii) Metoda je bila vključena v prvo obsežno primerjavo metod za računsko napovedovanje funkcij proteinov, ki je bila objavljena v reviji Nature methods, v kateri so primerjali 54 sodobnih metod za napovedovanje funkcij proteinov na ciljni množici 866 proteinov iz 11 organizmov.	
		ANG	We developed a new method for computational gene (or protein) function prediction. (i) The method, which was published in PLOS Computational Biology, is based on the principles of homology and phyletic profiles and uses ensembles of trees for hierarchical multi-label classification. In addition, the method was experimentally evaluated with wet lab experiments. The results show that the confidence estimates, obtained by applying our method, can be used to make informed decisions on experimental validation of computational predictions. (ii) The method was used in the first large-scale community-based critical assessment of protein function annotation (CAFA) experiment, which was published in Nature Methods, where 54 state of the art for protein function prediction methods were evaluated on a target set of 866 proteins from 11 organisms.	
	Objavljeno v		(i) Public Library of Science; PLoS computational biology; 2013; Vol. 9, no. 1; str. e10028521e100285214; Impact Factor: 4.867; Srednja vrednost revije / Medium Category Impact Factor: 1.957; A': 1; Avtorji / Authors: Škunca Nives, Bošnjak Matko, Kriško Anita, Panov Panče, Džeroski Sašo, Šmuc Tomislav (ii) Nature Publishing Group; Nature methods; 2013; Vol. 10, no. 3; str. 221-227; Impact Factor: 23.565; Srednja vrednost revije / Medium Category Impact Factor: 3.089; A'': 1; A': 1; WoS: CO; Avtorji / Authors: Radivojac Predrag, Panov Panče, Džeroski Sašo	
	Tipologija		1.01 Izvirni znanstveni članek	
3.	COBISS ID		27022119	Vir: COBISS.SI
	Naslov	SLO	Neuravnoteženi razredi in vplivne točke manjšinskega razreda	
		ANG	Class imbalance and the curse of minority hubs	
	Opis	SLO	Razvili smo nov način obravnave visokodimenzionalnih podatkov, ki uspešno uporabi zvezdičnost - tendenco visokodimenzionalnih podatkov, da vsebujejo posamezne točke v prostoru podatkov, ki so vplivne in se pogosto pojavljajo v soseščini drugih točk. Pokazali smo, da so pri klasifikaciji z metodo najbližjih sosedov vplivne točke iz manjšinskega razreda pogosto odgovorne za napačno klasifikacijo.	
		ANG	The paper deals with evaluating the impact of hubness on learning under class imbalance with nearest neighbor methods. Our results show that, contrary to the common belief, minority class hubs might be responsible for most misclassification in many high-dimensional datasets.	
	Objavljeno v		Butterworth Scientific Ltd.; Knowledge-based systems; 2013; Vol. 53; str. 157-172; Impact Factor: 3.058; Srednja vrednost revije / Medium Category Impact Factor: 1.609; A'': 1; A': 1; WoS: EP; Avtorji / Authors: Tomašev Nenad, Mladenović Dunja	

	Tipologija	1.01 Izvirni znanstveni članek	
4.	COBISS ID	26134055	Vir: COBISS.SI
	Naslov	SLO	Večciljna regresija z ansambli pravil
		ANG	Multi-target regression with rule ensembles
	Opis	SLO	Razvili smo algoritem FIRE za večciljno regresijo, ki uporablja pristop ansamblov pravil. Točnost algoritma smo izboljšali z dodajanjem preprostih linearnih funkcij v ansambel. Rezultati obsežne evalvacijske študije kažejo, da so ansambli pravil npr. bolj točni kot večciljna regresijska drevesa, a manj točni kot večciljni naključni gozdovi. Prednost ansamblov pravil je, da so bistveno bolj razumljivi kot naključni gozdovi. Možno pa je zgraditi tudi kompaktne množice pravil, ki so manjše od posameznega regresijskega drevesa, toda še vedno primerljive točnosti.
		ANG	We developed the FIRE algorithm for multi-target regression, which employs the rule ensemble approach. The accuracy of the algorithm was improved by adding simple linear functions to the ensemble. We also extensively evaluated the algorithm and the results show that multi-target regression rule ensembles are more accurate than, for instance, multi-target regression trees, but not quite as accurate as multi-target random forests. Rule ensembles have the advantage that they are significantly more concise than random forests, and it is also possible to create compact rule sets that are smaller than single regression trees but still comparably accurate.
	Objavljeno v	MIT Press; Journal of machine learning research; 2012; Vol. 13; str. 2367-2407; Impact Factor: 3.420; Srednja vrednost revije / Medium Category Impact Factor: 1.364; A'': 1; A': 1; WoS: AC, EP; Avtorji / Authors: Aho Timo, Ženko Bernard, Džeroski Sašo, Elomaa Tapio	
	Tipologija	1.01 Izvirni znanstveni članek	
5.	COBISS ID	22475303	Vir: COBISS.SI
	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	MIT Press; Journal of machine learning research; 2009; Vol. 10; str. 377-403; Impact Factor: 2.789; Srednja vrednost revije / Medium Category Impact Factor: 1.262; A'': 1; WoS: AC, EP; Avtorji / Authors: Kralj Novak Petra, Lavrač Nada, Webb Geoffrey I.	
	Tipologija	1.01 Izvirni znanstveni članek	

7. Najpomembnejši družbeno-ekonomski rezultati programske skupine⁶

	Družbeno-ekonomski dosežek		
1.	COBISS ID	23551271	Vir: COBISS.SI
	Naslov	<i>SLO</i>	Spletni portal VideoLectures.Net in World Summit nagrada OZN
		<i>ANG</i>	Web portal VideoLectures.Net and the World Summit award of the UN
	Opis	<i>SLO</i>	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 najboljšim e-vsebinam in ustvarjalnosti.
		<i>ANG</i>	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/winner/videolecturesnet-44820100621 Videolectures.net; 2009; Avtorji / Authors: Jermol Mitja, Dunja Mladenici, Marko Grobelnik	
	Tipologija	3.25	Druga izvedena dela
2.	COBISS ID	25004071	Vir: COBISS.SI
	Naslov	<i>SLO</i>	Okolje Orange4WS za servisno-orientirano rudarjenje podatkov
		<i>ANG</i>	Orange4WS environment for service-oriented data mining
	Opis	<i>SLO</i>	Razvili smo novo servisno orientirano okolje za rudarjenje podatkov ter referenčno implementacijo Orange4WS (Orange for Web Services). Orange4WS temelji na obstoječem programskem paketu za rudarjenje podatkov Orange ter njegovi komponenti za vizualno programiranje, ki omogoča ročno gradnjo delotokov, kateremu dodaja naslednje razširitve: (1) Uporaba spletnih servisov kot gradnikov delotokov za podatkovno rudarjenje. (2) Uporaba algoritmov za relacijsko rudarjenje podatkov v delotokih. (3) Ontologija področja odkrivanja znanja, ki opisuje komponente delotoka (podatki, znanje ter spletni servisi za podatkovno rudarjenje) na abstrakten način, ki omogoča strojno interpretacijo. Ontologija je uporabljena tudi za avtomatsko gradnjo delotokov. Po sklepu Znanstvenega sveta ARRS za Tehniko je bilo okolje Orange4WS razpoznano kot izjemni znanstveni dosežek za leto 2012.
		<i>ANG</i>	We developed a novel service-oriented knowledge discovery framework and its reference implementation in a service-oriented data mining environment Orange4WS (Orange for Web Services). Orange4WS is based on the existing Orange data mining toolbox and its visual programming environment, which enables manual composition of data mining workflows, including the following new features. (1) Simple use of web services as remote components that can be included into a data mining workflow. (2) Simple incorporation of relational data mining algorithms. (3) A knowledge discovery ontology to describe workflow components (data, knowledge and data mining services) in an abstract and machine-interpretable way, and its use for automated composition of data mining workflows. In 2012, the Scientific Board of the Slovenian research agency ARRS has recognized the Orange4WS platform as an outstanding research achievement.
	Šifra	F.15	Razvoj novega informacijskega sistema/podatkovnih baz
		Cambridge university press; The journals department; The Computer	

3.	Objavljeno v		journal; 2012; Vol. 55, no. 1; str. 82-98; Impact Factor: 0.755; Srednja vrednost revije / Medium Category Impact Factor: 1.024; WoS: ES, ET, EW, EX; Avtorji / Authors: Podpečan Vid, Žáková Monika, Lavrač Nada
	Tipologija		1.01 Izvirni znanstveni članek
	COBISS ID		26828583
	Vir: COBISS.SI		
	Naslov	SLO	Odločitveni model za odkrivanje zlorab v finančnem poslovanju
		ANG	A decision model for fraud detection in financial operation
	Opis	SLO	V okviru projekta FIRST (Large scale information extraction and integration infrastructure for supporting financial decision making) smo skupaj s strokovnjaki iz Nemčije razvili večparametrski odločitveni model za odkrivanje zlorab v finančnem poslovanju. Gre za zvrst zlorab, imenovanih "Pump and Dump", ki se nanašajo na nedovoljeno manipulacijo vrednosti finančnih instrumentov z razširjanjem lažnih informacij. Bistvena novost našega pristopa je, da pri ugotavljanju teh zlorab povezuje notranje finančne informacije z analizo sentimenta v dokumentih na medmrežju. Razvito rešitev smo predstavili v referatu, ki je dobil priznanje na mednarodni konferenci. Rešitev je tudi že vključena v znani in razširjeni informacijski sistem, ki ga proizvaja nemški projektni partner. Oboje nakazuje, da je rešitev uporabna v praksi, kjer lahko pomembno pomaga finančnim organizacijam pri odkrivanju zlorab in posledično zmanjševanju njihove škode.
		ANG	In the scope of the EU project FIRST (Large scale information extraction and integration infrastructure for supporting financial decision making), we have - in collaboration with experts from Germany - developed a multi-attribute model for fraud detection in financial operation. Specifically, we addressed the type of fraud known as "Pump and Dump", which refers to an illegal manipulation of financial instrument values by spreading false information. The novelty of our approach is in combining internal financial information with the analysis of sentiment in internet documents. The proposed solution was presented in an awarded conference paper. Also, it has been included in a well-known information system produced by the German project partner. This indicates that the solution is practically applicable and can substantially support financial institutions in detecting fraud and diminishing its consequences.
	Šifra		F.15 Razvoj novega informacijskega sistema/podatkovnih baz
	Objavljeno v		Moderna organizacija; eInnovation; 2013; Str. 64-77; Avtorji / Authors: Alić Irina, Siering Michael, Bohanec Marko
	Tipologija		1.08 Objavljeni znanstveni prispevek na konferenci
4.	COBISS ID		262711040
	Vir: COBISS.SI		
	Naslov	SLO	Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES
		ANG	Corpora of Slovene language: Gigafida, KRES, ccGigafida in ccKRES
	Opis	SLO	Sodelovali smo pri gradnji referenčnih korpusov slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES. Kulturna identiteta Slovencev je že od nekdaj tesno povezana z jezikom in zgrajeni korpusi predstavljajo pomemben mejnik pri opremljenosti slovenščine z referenčnimi in široko dostopnimi jezikovnimi viri. Delo na področju jezikovnih virov in tehnologij je potekalo v sodelovanju s komplementarnimi slovenskimi institucijami: UL-FF, UL-FDV, ZRC SAZU, NUK, Amebis in Alpineon. Naše delo na tem področju je bilo uporabljeno pri iskalniku dLib.si in pri poučevanju slovenskega jezika na gimnazijah, na Univerzah v Ljubljani in Mariboru ter pri raziskavah na univerzah in ZRC SAZU.
			We collaborated in the development of reference corpora of the Slovene language Gigafida, KRES, ccGigafida and ccKRES. The cultural identity of

5.			ANG	Slovenes has always been closely tied to our language, and the developed corpora represent an important milestone in equipping Slovene with widely available reference language resources. Our work in the field of language resources and technologies was conducted in collaboration with complementary Slovene institutions: UL-FF, UL-FDV, ZRC SAZU, NUK, Amebis and Alpineon. Our work in this field has been used in the search engine for the national digital library dLib.si, for teaching Slovene language in secondary schools, at the Universities of Ljubljana and Maribor and for research at universities and the Scientific Research Center of the Slovenian Academy of Sciences and Arts.
	Šifra		F.15	Razvoj novega informacijskega sistema/podatkovnih baz
	Objavljeno v		Trojina, zavod za uporabno slovenistiko; Fakulteta za družbene vede; 2012; 208 str.; A': 1; Avtorji / Authors: Logar Nataša, Grčar Miha, Brakus Marko, Erjavec Tomaž, Arhar Holdt Špela, Krek Simon	
	Tipologija		2.01	Znanstvena monografija
	COBISS ID		26363431	Vir: COBISS.SI
	Naslov	SLO	Modeliranje obrambnega odziva rastline s signalno mrežo	
		ANG	Signalling network construction for modelling plant defence response	
	Opis	SLO	Razvili smo metodologijo Bio3graph, ki omogoča avtomatsko luščenje relacij iz biološke literature: avtomatsko izluščene relacije v obliki trojk (komponenta1, reakcija, komponenta2) tvorijo omrežje, ki ga lahko vizualiziramo ter omogočimo njegovo biološko evalvacijo v primerjavi z ročno generiranim strukturnim modelom biološkega sistema. Orodje Bio3graph, ki vključuje slovar relevantnih bioloških izrazov, smo uporabili za ekstrakcijo relacij iz 9.586 relevantnih bioloških člankov, iz katerih smo izluščili 137 novih relacij, ki smo jih dodali v sprva ročno sestavljeno omrežje 175 komponent in 387 reakcij, s katerimi smo modelirali obrambni odziv rastlin na napad patogenov (virusov). Končno omrežje, ki predstavlja signalno omrežje prototipne rastline Arabidopsis thaliana sestavljeno iz 175 komponent in 524 reakcij, lahko služi kot koristen vir za nadaljnje modeliranje in interpretacijo omičnih podatkov.	
		ANG	Newly developed methodology Bio3graph allows for automated extraction of relations from biological literature: automatically extracted relations in the form of triplets (component1, reaction, component2) form a network structure, which can be visualized and examined by domain experts in comparison with manually constructed topology of a modeled biological system. The Bio3graph tool, which incorporates a vocabulary of relevant biological components and reaction types, was applied to a set of 9,586 relevant full text biological articles, which resulted in 137 newly detected reactions between the components. Finally, the initial manually constructed network topology, consisting of 175 components and 387 reactions, and the new reactions were merged to form a network structure consisting of 175 components and 524 reactions, modeling the plant defense response to virus attack of a model plant Arabidopsis thaliana. The resulting pathway diagram of plant defense signalling represents a valuable resource for further computational modeling and interpretation of omics data.	
	Šifra		F.15	Razvoj novega informacijskega sistema/podatkovnih baz
	Objavljeno v		Public Library of Science; PloS one; 2012; Vol. 7, no. 12; str. e51822-1e51822-18; Impact Factor: 3.730; Srednja vrednost revije / Medium Category Impact Factor: 2.514; A': 1; WoS: RO; Avtorji / Authors: Miljković Dragana, Stare Tjaša, Mozetič Igor, Podpečan Vid, Petek Marko, Witek Kamil, Dermastia Marina, Lavrač Nada, Gruden Kristina	
	Tipologija		1.01	Izvirni znanstveni članek

8. Drugi pomembni rezultati programske skupine²

Člani programske skupine so prejeli naslednje nagrade:

- Metoda "NoiseRank" avtorja Boruta Slubana je bila s strani Znanstvenoraziskovalnega sveta za tehniške vede ARRS prepoznana kot izjemen znanstveni dosežek in je bila predstavljena na dogodku "Odlični v znanosti 2013" v sklopu 9. Slovenskega foruma inovacij 2014.
- Nada Lavrač je v 2013 prejela Zoisovo priznanje za delo na področju inteligentne analize podatkov.
- Marko Bohanec in soavtorji so na 26. e-konferenci na Bledu 2013 prejeli nagrado za izjemen prispevek »Hot stock or not? A qualitative multi-attribute model to detect financial market manipulation«.
- Marko Grobelnik in soavtorji so prejeli drugo nagrado za najboljši demo prispevek "Exploring History through Newspaper Archives" na mednarodni konferenci Extended Semantic Web Conference 2012 – ESWC, maj 2012, Grčija
- Elena Ikonomovska je prejela nagrado za najboljši prispevek na doktorskem simpoziju konference EDBT/ICDT 2011.
- Blaž Fortuna in soavtorji so prejeli nagrado za tretji najboljši demo prispevek na konferenci ESWC 2011.
- Nenad Tomašev in soavtorji so prejeli nagrado za najboljši prispevek na konferenci MLDM 2011 in za drugi najboljši prispevek na konferenci PAKDD 2011.
- Tomaž Erjavec je skupaj z Matijo Ogrinom (ZRC SAZU) je v letu 2010 in ponovno v 2011 prejel nagrado »Google Digital Humanities Research Award«, ki je podprla njune raziskave na jezikoslovnih modelih starejše slovenščine.
- Orodje OntoGen avtorjev Blaža Fortune, Dunje Mladenič in Marka Grobelnika je prejelo 3. nagrado za inovativnost na Konferenci za prenos tehnologij 2010, ki sta jo organizirala Institut Jožef Stefan in TechnoCenter Univerze v Mariboru.
- Spletna storitev Searchpoint avtorja Boštjana Pajntara pod mentorstvom Marka Grobelnika in Dunje Mladenič je na Konferenci za prenos tehnologij na Institutu Jožef Stefan 2010 prejela nagrado za najbolj inventivne in inovativne ideje, ki so s svojo aplikativno vrednostjo zanimive in uporabne tudi v gospodarstvu.

9. Pomen raziskovalnih rezultatov programske skupine⁸

9.1. Pomen za razvoj znanosti²

SLO

Z raziskavami na področju tehnologij znanja smo si prizadevali za znanstveno odličnost in za uvajanje novih raziskovalnih smeri, skupaj s partnerji v številnih evropskih projektih, v katerih sodelujemo. Z znanstveno odličnostjo in s prenosom raziskovalnih rezultatov v prakso smo prispevali k povečanju znanstvene kompetitivnosti Slovenije v evropskem in svetovnem merilu. Delovali smo kot ključni slovenski in mednarodni akter na znanstvenih področjih analize kompleksnih podatkov, velikih podatkov, analize besedil in multimedijskih podatkov. Našo pomembno vlogo potrjuje uspešno sodelovanje v številnih mednarodnih projektih, vključno z 52 evropskimi projekti v obdobju 2009-2014, od katerih trenutno koordiniramo štiri (projekti 7.OP MAESTRA, NRG4Cast, TOPOSYS, XLike).

Objavljali smo v revijah z najvišjimi faktorji vpliva na področjih strojnega učenja, rudarjenja podatkov in upravljanja znanja (JMLR, MLJ, DMKD, Knowledge-Based Systems, itd.) in drugih visoko uvrščenih znanstvenih revijah (Nature Methods, PLOS Computational Biology, PLOS One itd.) ter na prestižnih mednarodnih konferencah (ECML/PKDD, IDA, ICDM, ILP, DS, NIPS itd.). Delovali smo kot predsedniki programskih odborov in organizator mednarodnih konferenc, delavnic, seminarjev in poletnih šol (ECML/PKDD-09, AIME 2011, ILP 2012, ICCV 2014, DS/ALT 2014 itd.) ter kot člani uredniških odborov znanstvenih revij (MLJ, DAMI, ACM TKDD, AIM, AAI, JMLR, JAIR, AI Com, IJCL, JLRE, EcoMod, EcoInf). Aktivni smo bili v programskih odborih konferenc (ICML, ECML/PKDD, ICDM, ILP, KDD, IDA, ACL) ter kot vabljeni predavatelji na konferencah. Sodelovali smo s prestižnimi znanstvenimi (Stanford, MIT, Tokyo Institute of Technology, itd.) in industrijskimi partnerji (Microsoft Research Cambridge, British Telecom, New York Times).

ANG

Our research in the area of knowledge technologies aimed at research excellence and at setting new research trends in collaboration with partners from the numerous European projects in which we participated. Through research excellence and transfer of scientific results into practice we contributed to Slovene competitiveness in the European and international research arena. We were a key international player in the areas of complex data analysis, big data, text and multi-media analysis. Our key role is exemplified by our successful participation in numerous international projects, including 52 EU projects in the 2009-2014 program funding period, four of which we currently coordinate (FP7 projects MAESTRA, NRG4Cast, TOPOSYS, XLike).

We have published papers in journals with the highest impact factors in the field of machine learning, data mining and knowledge management (JMLR, MLJ, DMKD, Knowledge-Based Systems, etc.) and in other top scientific journals (Nature Methods, PLOS Computational Biology, PLOS One, etc.) and presented our work at prestigious international conferences (ECML/PKDD, IDA, ICDM, ILP, DS, NIPS, etc.). We were active as chairs of program committees and organizers of numerous international conferences, workshops, tutorials, seminars and summer schools (ECML/PKDD-09, AIME 2011, ILP 2012, ICCV 2014, DS/ALT 2014), as members of 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, etc.), and gave invited lectures at conferences. We collaborated with distinguished scientific (Stanford, MIT, Tokyo Institute of Technology, etc.) and industrial institutions (Microsoft Research Cambridge, British Telecom, New York Times).

9.2. Pomen za razvoj Slovenije¹⁰

SLO

Neposreden učinek naših temeljnih raziskav je bil v napredku informacijskih tehnologij, potrebnih za razvoj slovenske družbe znanja. Aktivno smo prispevali k ekonomskemu razvoju Slovenije z aktivnim sodelovanjem v evropskih projektih, v katere smo aktivno vključevali slovenske industrijske partnerje. Temeljne raziskave so prispevale tudi k napredku informacijske infrastrukture na drugih področjih, kot so varstvo okolja, bioinformatika, medicina, zdravstvo, ekonomija, management in jezikoslovje oz. slovenistika, v katerih smo razvijali napredne aplikacije.

Druga, posredna vloga programa se je odražala v pomenu naprednih informacijskih tehnologij kot osnove za razvoj in podporo drugih področij znanja za družbenoekonomski in kulturni razvoj Slovenije. (a) V sodelovanju s slovenskimi industrijskimi partnerji smo metode za analizo kompleksnih podatkov uporabili v aplikacijah na področjih varstva okolja, varovanja zdravja, financ, socialnih omrežij in analize medijev. (b) Prispevki programa so doprinesli k družbenemu razvoju in razvoju storitev, kjer smo sisteme za večparametrsko podporo odločanja uporabili v aplikacijah na področju izobraževanja, zdravstva ter vrednotenja investicij in projektov, kar je omogočilo večjo transparentnost delovanja javnih zavodov in javne uprave. (c) Pomen za kulturni razvoj Slovenije temelji na tesni povezavi kulturne identitete Slovencev s slovenskim jezikom. Raziskave na področju jezikovnih tehnologij in digitalne humanistike nas postavljajo ob bok državam, ki vzorno skrbijo za svojo "digitalno dediščino". Z razvojem digitalnih izdaj starejših slovenskih besedil s podporo dveh nagrad Google in razvojem jezikovnih virov sodobne slovenščine smo omogočili boljše poučevanje slovenskega jezika, podprli sodobno slovaropisje ter razvoj drugih sodobnih računalniških metod za obdelavo slovenskega jezika.

Raziskovalni program je prispeval h gospodarskemu razvoju in izobraževanju tudi s prenosom znanja preko aplikativnih projektov in seminarjev organiziranih v sodelovanju s Centrom za prenos znanja IJS. Naše nagrajeno spletišče VideoLectures.net, na katerem je že preko 16.000 videoposnetkov predavanj, nudi profesionalno izobraževanje na daljavo. Smo profesorji in mentorji na Univerzah v Ljubljani, Mariboru in Novi Gorici in vrsti tujih univerz. Sodelujemo pri oblikovanju in izvajanju študijskega modula Tehnologije znanja na podiplomskem programu IKT, ki smo ga leta 2004 ustanovili na Mednarodni podiplomski šoli Jožefa Stefana, v okviru katerega letno izvajamo šest predmetov in mentoriramo večje število podiplomskih študentov.

ANG

Through basic research, we directly contributed to the advances in information technologies,

needed for the Slovene knowledge-based society. We contributed to advances through continued participation in EU projects, further enabling Slovenian industrial partners to become members of European consortia. Basic research in knowledge technologies has become increasingly important also as a provider of advanced ICT infrastructures to other disciplines, including environment protection, bioinformatics, medicine, health care, economy and management, where we contributed to application development.

The second role of the program was indirect, through the impact on socio-economic and cultural development of Slovenia. (a) Complex data analytics, text and multimedia mining methods were used (in continued collaboration with Slovene industrial partners) in practical solutions in the areas of environmental protection, health care, finance, social network and media analysis. (b) By using multiparameter decision support, positive impact on societal and service sectors was achieved via applications in education, health care and investments/projects assessment, leading to greater efficiency and operational transparency of public and governmental institutions. (c) Contribution to Slovene cultural development is based on the fact that the cultural identity of Slovenes is intimately tied to our language. Our language technologies and digital heritage research contributed to the accessibility of written Slovene cultural heritage through open digital editions of historical (for which we were twice granted Google awards) and contemporary Slovene language resources, which are used in teaching of the Slovene language, in lexicography and in the development of modern computational methods for processing of Slovene, thus empowering Slovene to fully function in the information society.

The program significantly impacted businesses and education by continued know-how transfer through applied projects and seminars organized by the JSI Center for IT Knowledge Transfer, while distance learning is enabled through our award-winning portal VideoLectures.net, already providing over 16,000 video lectures. We contributed to postgraduate education as professors and supervisors at the Universities of Ljubljana, Maribor, and Nova Gorica and several foreign universities. In 2004 we have founded the Knowledge Technologies module of the ICT postgraduate study program at the Jožef Stefan International Postgraduate School, in which we are yearly responsible for six courses and for supervising many PhD/MSc students.

10. Zaključena mentorstva članov programske skupine pri vzgoji kadrov v obdobju 1.1.2009-31.12.2014¹¹

10.1. Diplome¹²

vrsta usposabljanja	število diplom
bolonjski program - I. stopnja	13
bolonjski program - II. stopnja	18
univerzitetni (stari) program	6

10.2. Magisterij znanosti in doktorat znanosti¹³

Šifra raziskovalca	Ime in priimek	Mag.	Dr.	MR	
32152	Matjaž Juršič	☐	☒	☒	
29482	Mojca Stubelj Ars	☐	☒	☒	
32253	Borut Sluban	☐	☒	☐	
36836	Biljana Mileva Boshkoska	☐	☒	☐	
29539	Vid Podpečan	☐	☒	☒	
29890	Darko Cherepnalkoski	☐	☒	☒	
36912	Dragana Miljković	☐	☒	☐	
32284	Elena Ikonomovska	☐	☒	☒	
31571	Daniela Stojanova	☐	☒	☐	

32908	Ivica Slavkov	☐	☒	☐	
27759	Panče Panov	☐	☒	☐	
28488	Katerina Tashkova	☐	☒	☒	
31050	Dragi Kocev	☐	☒	☐	
32282	Aneta Trajanov	☐	☒	☐	
28291	Petra Kralj Novak	☐	☒	☐	
29306	Ingrid Petrič	☐	☒	☐	
0	Aleksandar Pečkov	☐	☒	☐	
0	Aleksander Pur	☐	☒	☐	
0	Vladimir Kuzmanovski	☒	☐	☐	
0	Xiaobin Li	☒	☐	☐	
0	Janez Kralj	☒	☐	☐	
33582	Bojan Blažica	☐	☒	☐	
0	Nenad Tomašev	☐	☒	☐	
34646	Inna Novalija	☐	☒	☐	
28015	Blaž Fortuna	☐	☒	☒	
0	Daniela Stojanova	☒	☐	☐	
0	Delia Rusu	☐	☒	☐	
37652	Darko Aleksovski	☐	☒	☐	

Legenda:

Mag. - Znanstveni magisterij**Dr.** - Doktorat znanosti**MR** - mladi raziskovalec**11. Pretok mladih raziskovalcev – zaposlitev po zaključenem usposabljanju¹⁴**

Šifra raziskovalca	Ime in priimek	Mag.	Dr.	Zaposlitev	
32152	Matjaž Juršič	☐	☒	A - raziskovalni zavodi	
29539	Vid Podpečan	☐	☒	A - raziskovalni zavodi	
32284	Elena Ikonomovska	☐	☒	E - Tujina	
28488	Katerina Tashkova	☐	☒	E - Tujina	
28015	Blaž Fortuna	☐	☒	A - raziskovalni zavodi	
29482	Mojca Stubelj Ars	☐	☒	F - Drugo	
29890	Darko Cherepnalkoski	☐	☒	A - raziskovalni zavodi	

Legenda zaposlitev:

A - visokošolski in javni raziskovalni zavodi**B** - gospodarstvo**C** - javna uprava**D** - družbene dejavnosti**E** - tujina**F** - drugo

12.Vključenost raziskovalcev iz podjetij in gostovanje raziskovalcev, podoktorandov ter študentov iz tujine, daljše od enega meseca, v obdobju 1.1.2009-31.12.2014

Šifra raziskovalca	Ime in priimek	Sodelovanje v programski skupini	Število mesecev	
0	Timo Aho	C - študent – doktorand ▼	6	
0	Monika Žakova	C - študent – doktorand ▼	6	
0	Lise Getoor	C - študent – doktorand ▼	3	
0	Nenad Stojanović	A - raziskovalec/strokovnjak ▼	9	
0	Baptiste Lecroart	A - raziskovalec/strokovnjak ▼	1	
0	Laura Langohr	C - študent – doktorand ▼	5	
0	Andreea Bizau	C - študent – doktorand ▼	2	
0	Larisa Soldatova	B - uveljavljeni raziskovalec ▼	2	
0	Martin Saveski	C - študent – doktorand ▼	10	
0	Michelangelo Ceci	B - uveljavljeni raziskovalec ▼	4	
0	Don Hodges	B - uveljavljeni raziskovalec ▼	6	
0	Suzana Loškoska	B - uveljavljeni raziskovalec ▼	7	
0	Bogdan Okreša Djurić	C - študent – doktorand ▼	2	
0	Sebastian Dumančić	C - študent – doktorand ▼	1	
0	Katarina Trojchanec	C - študent – doktorand ▼	1	
0	Ivan Kitanovski	C - študent – doktorand ▼	1	
0	Mario Karlovčec	C - študent – doktorand ▼	12	
0	Sabrina Guettes	D - podoktorand ▼	3	
0	Calin Raielan	C - študent – doktorand ▼	3	
0	Gianvito Pio	C - študent – doktorand ▼	4	
0	Ivica Dimitrovski	C - študent – doktorand ▼	12	
0	Gjorgji Madjarov	C - študent – doktorand ▼	12	
0	Ruslan Miniakhmetov	C - študent – doktorand ▼	3	
36871	Nikola Ljubešić	B - uveljavljeni raziskovalec ▼	3	

Legenda sodelovanja v programski skupini:

- A** - raziskovalec/strokovnjak iz podjetja
B - uveljavljeni raziskovalec iz tujine
C - študent – doktorand iz tujine
D - podoktorand iz tujine

13.Vključevanje v raziskovalne programe Evropske unije in v druge mednarodne raziskovalne in razvojne programe ter drugo mednarodno sodelovanje v obdobju 1.1.2009-31.12.2014¹⁵

SLO

Zaradi prostorske omejitve na tem mestu podajamo samo seznam kratic projektov in njihovih nosilcev. Celoten seznam s polnimi imeni projektov je v priponki "EU-Projects.pdf".

- ACTIVE; D. Mladenić, M. Grobelnik.
- ALERT; D. Mladenić.
- BISON; N. Lavrač.
- Co-Extra; M. Bohanec.
- COIN; D. Mladenić, M. Grobelnik.
- ConCrete; N. Lavrač, M. Žnidaršič.
- DECATHLON; M. Bohanec.
- EET-Pipeline; S. Džeroski.
- E-LICO; N. Lavrač, M. Žnidaršič.
- ENVISION; D. Mladenić, N. Lavrač.
- ESC; D. Mladenić, M. Grobelnik.
- EURIDICE; D. Mladenić, M. Grobelnik.
- FIRST; N. Lavrač.
- FlaReNet; T. Erjavec.
- FOC II; I. Mozetič.
- GENDERA; D. Mladenić.
- HBP; S. Džeroski, N. Lavrač.
- Healththreats; N. Lavrač.
- IMAGINATION; D. Mladenić.
- IMPACT; T. Erjavec.
- LT-Web; D. Mladenić.
- MAESTRA; S. Džeroski, D. Kocev; **koordinatorji**.
- MEDIAMIXER; D. Mladenić.
- META Net; M. Grobelnik.
- MOBIS; D. Mladenić.
- MONDILEX; T. Erjavec.
- MULTILINGUALWEB; M. Grobelnik, D. Mladenić.
- MULTIPLEX; I. Mozetič, P. Kralj Novak.
- MUSE; N. Lavrač.
- NeOn; M. Grobelnik, D. Mladenić.
- NRG4Cast; D. Mladenić; **koordinatorji**.
- PASCAL2; M. Grobelnik.
- PHAGOSYS; S. Džeroski.
- PLANET DATA; M. Grobelnik, D. Mladenić.
- ProaSense; M. Grobelnik.
- PROSECCO; N. Lavrač.
- RENDER; M. Grobelnik, D. Mladenić.
- REWIRE; S. Džeroski, B. Ženko.
- SIMPOL; I. Mozetič.
- SiS CATALYST; D. Mladenić.
- SMART; M. Grobelnik.
- SOPHOCLES; M. Grobelnik.
- SUMO; S. Džeroski.
- SWING; M. Grobelnik, D. Mladenić.
- SYMPHONY; D. Mladenić, M. Grobelnik.
- TAO; M. Grobelnik, D. Mladenić.
- TOPOSYS; M. Grobelnik; **koordinatorji**.
- Translectures; D. Mladenić.
- VIDI; M. Grobelnik, D. Mladenić.
- WHIM; N. Lavrač, M. Žnidaršič.
- XLike; M. Grobelnik; **koordinatorji**.
- XLime; M. Grobelnik.

14. Vključenost v projekte za uporabnike, ki so v obdobju trajanja raziskovalnega programa (1.1.2009–31.12.2014) potekali izven financiranja ARRS¹⁶

SLO

Projekti izven financiranja ARRS:

- EVADIFF: Vrednotenje obstoječih modelov in razvoj novih odločitvenih orodij za

preprečevanje razpršenega onesnaževanja, ki ga povzročajo fitofarmacevtska sredstva; naročnik ARVALIS, Francija; M. Debeljak.

- GOOGLE: skupaj z ZRC SAZU smo bili dvakrat (2011, 2012) prejemniki Googlove nagrade "Developing Language Models of Historical Slovene"; T. Erjavec.
- LAND-EME: Trajnostni sistem posevkov na Škotskem; naročnik SCRI, Living Technology, Škotska; M. Debeljak.
- Evropski koncepti in pristopi za proizvodnjo gensko spremenjene koruze; naročnik ARVALIS, Francija; M. Debeljak.
- Regionalni sistemi svetovanja za zagotavljanje soobstoja koruze; naročnik ARVALIS, Francija; M. Debeljak.
- VIZIPIN: Varna infrastruktura za izvajanje poveljevanja in nadzora; naročnik Iskra Zaščite; M. Grobelnik.
- GMOtrack - prototip sistema za podporo pri odločanju; naročnik Nacionalni Inštitut za biologijo; N. Lavrač, P. Kralj Novak.
- SM-RIS: Mladinska mreža razvoja raziskovalnih vrednot mladih; naročnik MIZŠ; D. Mladenčič.
- Vpliv prenosa genov, genske raznovrstnosti in raznolikosti ter tehnologije pridelovanja oljne ogrščice na soobstoj in izpolnjevanje pogojev trajnostne pridelave ter razvoj metod za sledljivost; naročnik Ministrstvo za kmetijstvo; M. Debeljak.
- Analiza in scenarij razvoja in rabe gozdov v Sloveniji; naročnik Ministrstvo za kmetijstvo; M. Debeljak.
- Sodelovanje pri razvoju produkta "Virtualni spletni asistent"; naročnik Antiveb d.o.o.; D. Mladenčič.
- Raziskovalni vavčer: Razvoj semantičnih analiz za evropske jezike; naročnik GAMA SYSTEM d.o.o.; I. Mozetič.
- Raziskovalni vavčer: Vzpostavitev platforme naprednih storitev za upravljanje z energijo pri gospodinjstvih odjemalcih; naročnik SOLVERA LYNX d.d.; M. Grobelnik.
- Spodbujanje raziskovalcev na začetku kariere: Delotoki v oblaku; naročnik MIZŠ, D. Cherepnalkoski.
- Spodbujanje raziskovalcev na začetku kariere: Analiza sentimenta; naročnik MIZŠ; M. Juršič.
- OVJE - Ocena vzdržnosti za razvoj energetike v Sloveniji do leta 2030 s poudarkom na jedrski tehnologiji; naročnik GEN energija d.o.o.; M. Bohanec.
- Kompetenčni center: KC Odprta komunikacijska platforma za integracijo storitev (KC OPCOMM); M. Grobelnik.
- Kompetenčni center: KC Storitve podprte z računalništvom v oblaku (KC CLASS); M. Grobelnik.
- Smer+ : Tehnologije za izvedbo inteligentnega avtodoma naslednje generacije; naročnik Adria Mobil d.o.o.; M. Grobelnik.

Družbene in kulturne dejavnosti:

- Sodelovanje pri dejavnostih Slovenskega društva za umetno inteligenco SLAIS in Slovenskega društva za jezikovne tehnologije SDJT.
- Sodelovanje pri pripravi Nacionalnega programa za jezikovno politiko 2014-2018 in Akcijskega načrta za izvajanje programa.
- Sodelovanje pri pripravi Digitalne agende za Evropo na področju kulture v Sloveniji.
- Sodelovanje pri pripravi Akcijskega načrta za vzpostavitev sistema odprtega dostopa do raziskovalnih podatkov financiranih z javnimi sredstvi.

15. Ocena tehnološke zrelosti rezultatov raziskovalnega programa in možnosti za njihovo implementacijo v praksi (točka ni namenjena raziskovalnim programom s področij humanističnih ved)¹⁷

SLO

Nekaj programskih rešitev programske skupine, razvitih v obdobju financiranja programske skupine (2009-2014), je zrelih za trženje in se redno uporabljajo v praksi. Mednje sodijo:

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

Programa GMOTrack in GMOSeek za sledenje in identifikacijo gensko spremenjenih organizmov (GSO) v hrani, krmi in semenih, sta bila razvita v sodelovanju z Nacionalnim institutom za biologijo. Programa bistveno znižata ceno testiranja prisotnosti in identifikacijo GSO v vzorcih. Ker prisotnost GSO na trgu narašča, Evropska zakonodaja pa predpisuje označevanje izdelkov, ki vsebujejo GSO, je potreba po učinkovitem testiranju zelo velika v Evropski Uniji, zaradi tržne niše ekoloških izdelkov pa tudi izven Evropske Unije. Programa sta javno dostopna na spletni strani <http://kt.ijs.si/software/GMOTrack/> in v redni rabi na Nacionalnem institutu za biologijo.

Računalniški program za večparametrsko odločanje DEXi (zadnja verzija je 4.00) je dostopen na <http://kt.ijs.si/MarkoBohanec/dexi.html>. Program omogoča razvoj, evalvacijo in uporabo kvalitativnih večparametrskih modelov za podporo odločanju pri zahtevnih odločitvah. DEXi se uporablja na številnih področjih, npr. za vrednotenje projektov, investicij in ocenjevanje ugleda institucij.

Orange4WS je odprta platforma za sestavljanje in izvajanje delotokov podatkovnega rudarjenja. Zapleteni procesi podatkovnega rudarjenja in strojnega učenja so preko vizualnega programiranja v njej predstavljeni kot enostavni delotoki, razumljivi tako strokovnjakom kot tudi laikom. S platformo, ki je dostopna na <http://orange4WS.ijs.si>, bi bilo možno trženje podpornih storitev in razvoj dodatnih gradnikov.

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

16. 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	69.000 EUR
ocena potrebne infrastrukture in opreme ¹⁸	Spin-off podjetje bi lahko ustanovili za trženje programske opreme Orange4WS in s tem povezane storitve kot so: razvoj programskih komponent na zahtevo stranke, razvoj rešitev na področju analize podatkov, servisna podpora in usposabljanje. Predvsem na področju usposabljanja so mogoče sinergije z raziskovalnimi in pedagoškimi aktivnostmi programske skupine. Potrebna oprema obsega dva večprocesorska strežnika (2 x 3.000 EUR) in dva osebna računalnika (2 x 1.500 EUR). Skupni stroški opreme in infrastrukture v prvem letu delovanja bi tako znašali približno 9.000 EUR. Stroški dela in oglaševanja so prilagodljivi; ob pričakovani aktivnosti jih ocenjujemo na 60.000 EUR letno.

17. Izjemni dosežek v letu 2014¹⁹

17.1. Izjemni znanstveni dosežek

Pri podatkovnem rudarjenju se vedno pogosteje srečujemo s kompleksnimi in strukturiranimi podatki, pri čemer njihovo enotno obravnavanje predstavlja velik raziskovalni izziv. Teoretični okvir, ki bi enotno obravnaval tako različne kompleksne tipe podatkov, kot tudi različne naloge

in metode podatkovnega rudarjenja, bi pomenil velik prispevek v smeri avtomatizacije odkrivanja znanja v podatkih. Razvili smo ontologijo OntoDM, ki formalno opredeljuje vse glavne entitete podatkovnega rudarjenja, vključno s podatki ter nalogami in metodami podatkovnega rudarjenja. Ontologija je razvita v skladu s sodobnimi standardi in načeli načrtovanja ontologij in smiselno vključuje najsodobnejše obstoječe znanstvene ontologije kot so Basic Formal Ontology (BFO), Relational Ontology (RO) in druge. To zagotavlja skladnost in povezljivost z drugimi ontologijami in sklepanje prek več domen. OntoDM je javno dostopna na naslovu <http://www.ontodm.com>.

17.2. Izjemni družbeno-ekonomski dosežek

Razvili smo metodologijo iterativne optimizacije parametrov modela, ki je bila uporabljena za izgradnjo kompleksnega matematičnega modela obrambnega odziva rastlin na virusne okužbe. Predlagana metodologija omogoča izgradnjo simulacijskega modela, ki ga je zdaj mogoče zgraditi kljub pomanjkanju kinetičnih podatkov iz katerih bi se lahko določilo parametre takega modela. Domenski strokovnjaki z Nacionalnega instituta za biologijo so svoje domensko znanje formalizirali v obliki omejitev, ki morajo biti zadoščene pri simulaciji modela. Predlagani omejitveno usmerjeni optimizacijski pristop omogoča učinkovito preučevanje dinamičnega obnašanja bioloških modelov, hkrati pa povečuje njihovo zanesljivost. Razvito metodologijo je odslej mogoče uporabiti za izdelavo dinamičnih modelov poljubnih bioloških mehanizmov, kjer je kljub pomanjkanju eksperimentalnih rezultatov na voljo dovolj domenskega znanja.

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 in obdelavo teh podatkov za evidence ARRS;
- so vsi podatki v obrazcu v elektronski obliki identični podatkom v obrazcu v papirnati obliki;
- so z vsebino poročila seznanjeni in se strinjajo vsi izvajalci raziskovalnega programa.

Podpisi:

*zastopnik oz. pooblaščen oseba
matične RO (JRO in/ali RO s
koncesijo):*

in

vodja raziskovalnega programa:

Institut "Jožef Stefan"

Nada Lavrač

ŽIG

Kraj in datum:

Ljubljana

13.3.2015

Oznaka poročila: ARRS-RPROG-ZP-2015/1

¹ Napišite povzetek raziskovalnega programa v slovenskem jeziku (največ 3.000 znakov vključno s presledki – približno pol strani, velikost pisave 11) in angleškem jeziku (največ 3.000 znakov vključno s presledki – približno pol strani, velikost pisave 11). [Nazaj](#)

² Napišite kratko vsebinsko poročilo, v katerem predstavite raziskovalno hipotezo in opis raziskovanja. Navedite ključne ugotovitve, znanstvena spoznanja, rezultate in učinke raziskovalnega programa in njihovo uporabo ter sodelovanje s tujimi partnerji. V primeru odobrenega povečanja obsega financiranja raziskovalnega programa v letu 2014 mora poročilo o realizaciji programa dela zajemati predložen program dela ob prijavi in predložen dopolnjen program dela v letu 2014. Največ 12.000 znakov vključno s presledki (približno dve strani, velikosti pisave 11). [Nazaj](#)

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

V primeru bistvenih odstopanj in sprememb od predvidenega programa dela raziskovalnega programa, kot je bil zapisan v predlogu raziskovalnega programa oziroma v primeru sprememb, povečanja ali zmanjšanja sestave programske skupine v zadnjem letu izvajanja raziskovalnega programa, napišite obrazložitev. V primeru, da sprememb ni bilo, navedite: "Ni bilo sprememb.". Največ 6.000 znakov vključno s presledki (približno ena stran, velikosti pisave 11). [Nazaj](#)

⁵ Navedite znanstvene dosežke (največ pet), ki so nastali v okviru izvajanja raziskovalnega programa. Raziskovalni dosežek iz obdobja izvajanja programa vpišete tako, da izpolnite COBISS kodo dosežka – sistem nato sam izpolni naslov objave, naziv, IF in srednjo vrednost revije, naziv FOS področja ter podatek, ali je dosežek uvrščen v A'' ali A'. [Nazaj](#)

⁶ Navedite družbeno-ekonomske dosežke (največ pet), ki so nastali v okviru izvajanja raziskovalnega programa. Družbeno-ekonomski dosežek iz obdobja izvajanja programa vpišete tako, da izpolnite COBISS kodo dosežka – sistem nato sam izpolni naslov objave, naziv, IF in srednjo vrednost revije, naziv FOS področja ter podatek, ali je dosežek uvrščen v A'' ali A'.

Družbeno-ekonomski dosežek je po svoji strukturi drugačen kot znanstveni dosežek. Povzetek znanstvenega dosežka je praviloma povzetek bibliografske enote (članka, knjige), v kateri je dosežek objavljen.

Povzetek družbeno-ekonomskega dosežka praviloma ni povzetek bibliografske enote, ki ta dosežek dokumentira, ker je dosežek sklop več rezultatov raziskovanja, ki je lahko dokumentiran v različnih bibliografskih enotah. COBISS ID zato ni enoznačen, izjemoma pa ga lahko tudi ni (npr. prehod mlajših sodelavcev v gospodarstvo na pomembnih raziskovalnih nalogah, ali ustanovitev podjetja kot rezultat programa ... - v obeh primerih ni COBISS ID). [Nazaj](#)

⁷ Navedite rezultate raziskovalnega programa iz obdobja izvajanja programa v primeru, da katerega od rezultatov ni mogoče navesti v točkah 6 in 7 (npr. ker se ga v sistemu COBISS ne vodi). Največ 2.000 znakov vključno s presledki (približno 1/3 strani, velikost pisave 11). [Nazaj](#)

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

⁹ Največ 4.000 znakov vključno s presledki (približno 2/3 strani, velikost pisave 11). [Nazaj](#)

¹⁰ Največ 4.000 znakov vključno s presledki (približno 2/3 strani, velikost pisave 11). [Nazaj](#)

¹¹ Upoštevajo se le tiste diplome, magisteriji znanosti in doktorati znanosti (zaključene/i v obdobju 1.1.2009–31.12.2014), pri katerih so kot mentorji sodelovali člani programske skupine. [Nazaj](#)

¹² Vpišite število opravljenih diplom v času izvajanja raziskovalnega programa glede na vrsto usposabljanja. [Nazaj](#)

¹³ Vpišite šifro raziskovalca in/ali ime in priimek osebe, ki je v času izvajanja raziskovalnega programa pridobila naziv magister znanosti in/ali doktor znanosti ter označite doseženo izobrazbo. V primeru, da se je oseba usposabljala po programu Mladi raziskovalci, označite "MR". [Nazaj](#)

¹⁴ Za mlade raziskovalce, ki ste jih navedli v tabeli 11.2. točke (usposabljanje so uspešno zaključili v obdobju od 1.1.2009 do 31.12.2014), izberite oz. označite, kje so se zaposlili po zaključenem usposabljanju. [Nazaj](#)

¹⁵ Navedite naslove projektov in ime člana programske skupine, ki je bil vodja/koordinator navedenega projekta. Največ 6.000 znakov vključno s presledki (približno ena stran, velikosti pisave 11). [Nazaj](#)

¹⁶ Navedite naslove projektov, ki ne sodijo v okvir financiranja ARRS (npr: industrijski projekti, projekti za druge naročnike, državno upravo, občine idr.) in ime člana programske skupine, ki je bil vodja/koordinator navedenega projekta. Največ 6.000 znakov vključno s presledki (približno ena stran, velikosti pisave 11). [Nazaj](#)

¹⁷ Opišite 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, razvito v okviru programa oziroma dodano vrednost na zaposlenega, če jo je mogoče oceniti (npr. v primerih, ko je rezultat izboljšava obstoječih tehnologij oziroma izdelkov). Največ 3.000 znakov vključno s presledki (približno pol strani, velikosti pisave 11). [Nazaj](#)

¹⁸ Največ 1.000 znakov vključno s presledki (približno 1/6 strani, velikost pisave 11) [Nazaj](#)

¹⁹ Navedite en izjemni znanstveni dosežek in/ali en izjemni družbeno-ekonomski dosežek raziskovalnega programa v letu 2014 (največ 1000 znakov, vključno s presledki, velikost pisave 11). Za dosežek pripravite diapozitiv, ki vsebuje sliko ali drugo slikovno gradivo v zvezi z izjemnim dosežkom (velikost pisave najmanj 16, približno pol strani) in opis izjemnega dosežka (velikost pisave 12, približno pol strani). Diapozitiv/-a priložite kot priponko/-i k temu poročilu. Vzorec diapozitiva je objavljen na spletni strani ARRS <http://www.arrs.gov.si/sl/gradivo/>, predstavitev dosežkov za pretekla leta pa so objavljena na spletni strani <http://www.arrs.gov.si/sl/analyze/dosez/>. [Nazaj](#)

Obrazec: ARRS-RPROG-ZP/2015 v1.00b

B8-6B-EE-2F-FF-9E-15-3A-A7-78-2A-A8-45-CA-CD-7B-A9-5A-96-C4

Priloga 1

Complete list of EU funded projects; celoten seznam evropskih projektov

52 projects in total, including 4 coordinated by JSI; skupaj 52 projektov, od tega 4 koordinatorstva:

- ACTIVE: Enabling the Knowledge Powered Enterprise; Razvoj in implementacije tehnologij znanj v znanjsko-intenzivnih podjetjih; D. Mladenič, M. Grobelnik.
- ALERT: Active support and real-time coordination based on event processing in open source software development; Aktivna podpora in koordinacija odprtokodnih projektov v realnem času, osnovana na procesiranju dogodkov; D. Mladenič.
- BISON: Bisociation Networks for Creative Information Discovery; Bisociativne mreže za kreativno odkrivanje znanja; N. Lavrač.
- Co-Extra: GM and non-GM supply chains: their CO-Existence and Traceability; Sožitje in sledljivost gensko spremenjenih in nespremenjenih proizvodov v preskrbovalnih verigah; M. Bohanec.
- COIN: Collaboration and Interoperability for networked enterprises; Sodelovanje in interoperabilnost v mrežnih organizacijah; D. Mladenič, M. Grobelnik.
- ConCrete: Concept Creation Technology; Tehnologija kreiranja konceptov; N. Lavrač, M. Žnidaršič.
- DECATHLON: Development of Cost efficient Advanced DNA-based methods for specific Traceability issues and High Level On-site applications; Razvoj naprednih cenovno učinkovitih metod, osnovanih na analizi DNK za posebna vprašanja sledljivosti in visokonivojske aplikacije na kraju samem; M. Bohanec.
- EET-Pipeline: European Embryonal Tumor Pipeline; Evropska platforma za raziskave tumorjev zarodkov; S. Džeroski.
- E-LICO: eLaboratory for Interdisciplinary Collaborative Research in Data Mining and Data-Intensive Sciences; e-Laboratorij za skupinske interdisciplinarne raziskave na področju rudarjenja podatkov in podatkovno intenzivnih znanosti; N. Lavrač, M. Žnidaršič.
- ENVISION: Environmental Services Infrastructures with Ontologies; Semantična infrastruktura za izvedbo okoljskih informacijskih sistemov; D. Mladenič, N. Lavrač.
- ESC: European Security Challenge; Evropski varnostni izziv; D. Mladenič, M. Grobelnik.
- EURIDICE: European Inter-Disciplinary Research on Intelligent Cargo for Efficient, Safe and Environment-friendly Logistics; Evropske interdisciplinarne raziskave inteligentnih sistemov za učinkovito, varno ter okolju prijazno upravljanje tovornega prometa; D. Mladenič, M. Grobelnik.
- FIRST: Large scale information extraction and integration infrastructure for supporting financial decision making; Integracija in analiza informacijskih tokov za podporo odločanja v domeni finančnih trgov; N. Lavrač.
- FlaReNet: Fostering Language Resources Network; Razvoj mreže jezikovnih virov; T. Erjavec.
- FOC: Forecasting Financial Crises; Napovedovanje finančnih kriz; I. Mozetič.
- GENDERA: Gender Debate in the European Research Area; Razprava o enakosti spolov v evropskem raziskovalnem prostoru; D. Mladenič.
- HBP: Human Brain Project; Projekt človeški možgani; S. Džeroski, N. Lavrač.
- Healthreats: Integrated Decision Support System for HEALTH THREATS and crises Management; Integrirani sistem podpore odločanja v primerih zdravstvenih groženj in upravljanja kriznih situacij; N. Lavrač.
- IMAGINATION: Image-based Navigation in Multimedia Archives; Pregledovanje multimedijskih arhivov na podlagi slik; D. Mladenič.

- IMPACT: Improving Access to Text; Izboljšanje dostopa do besedil; T. Erjavec.
- LT-Web: Language Technologies in the Web; Jezikovne tehnologije na svetovnem spletu; D. Mladenić.
- MAESTRA: Learning from Massive, Incompletely annotated, and Structured Data; Učenje iz obsežnih, nepopolno označenih in strukturiranih podatkov; S. Džeroski, D. Kocev; **coordinators / koordinatorji**.
- MEDIAMIXER: Community set-up and networking for the reMIXing of online MEDIA fragments; Oblikovanje skupnosti in povezovanje za ponovno uporabo in mešanje spletnih medijskih fragmentov; D. Mladenić.
- META-NET: Technologies for the Multilingual European Information Society; Tehnologije za večjezično evropsko informacijsko družbo; M. Grobelnik.
- MOBIS: Personalized Mobility Services for energy efficiency and security through advanced Artificial Intelligence techniques; Personalizirane mobilne storitve za prilagojeno in energetske učinkovito potovanje, podprto s tehnikami umetne inteligence; D. Mladenić.
- MONDILEX: Conceptual Modelling of Networking of Centres for High-Quality Research in Slavic Lexicography and their Digital Resources; Konceptualno modeliranje povezovanja centrov za visokokvalitene raziskave slovanske leksikografije in njihovih digitalnih virov; T. Erjavec.
- MULTILINGUALWEB: Advancing the Multilingual Web, Thematic Network; Izboljšajmo večjezični splet; M. Grobelnik, D. Mladenić.
- MULTIPLEX: Foundational Research on Multilevel Complex Networks and Systems; Osnovne raziskave večnivojskih kompleksnih omrežij in sistemov; I. Mozetič, P. Kralj Novak.
- MUSE: Machine Understanding for interactive Storytelling; Računalniško razumevanje jezika za interaktivno pripovedništvo; N. Lavrač.
- NeOn: Lifecycle Support for Networked Ontologies; Podpora življenjskega cikla mrežno povezanih ontologij; M. Grobelnik, D. Mladenić.
- NRG4Cast: Energy Forecasting; Napovedovanje na področju energije; D. Mladenić; **coordinators / koordinatorji**.
- PASCAL2: Pattern Analysis, Statistical Modelling and Computational Learning 2; Analiza vzorcev, statistično modeliranje in računalniško učenje 2; M. Grobelnik.
- PHAGOSYS: Systems biology of phagosome formation and maturation - modulation by intracellular pathogens; Sistemska biologija formacije in dozorevanja fagosomov - modulacija z znotrajceličnimi patogeni; S. Džeroski.
- PLANET DATA; Inteligentni informacijski management; M. Grobelnik, D. Mladenić.
- ProaSense: The Proactive Sensing Enterprise; Podjetje za proaktivno zaznavanje; M. Grobelnik.
- PROSECCO: Promoting the Scientific Exploration of Computational Creativity; Promoviranje znanstvenega raziskovanja računalniške kreativnosti; N. Lavrač, S. Pollak.
- RENDER: Reflecting Knowledge Diversity; Prikazovanje raznolikosti znanja; M. Grobelnik, D. Mladenić.
- REWIRE: Rehabilitative Wayout in Responsive home Environments; Rehabilitacija v interaktivnem domačem okolju; S. Džeroski, B. Ženko.
- SIMPOL: Financial Systems Simulation and Policy Modelling; Simulacija finančnih sistemov in modeliranje politik; I. Mozetič.
- SiS CATALYST: Children as Change Agents for Science in Society; Otroci kot nosilci sprememb za prihodnost znanosti v družbi; D. Mladenić.
- SMART: Statistical Multilingual Analysis for Retrieval and Translation; Statistična večjezična analiza za zajemanje in prevajanje besedil; M. Grobelnik.

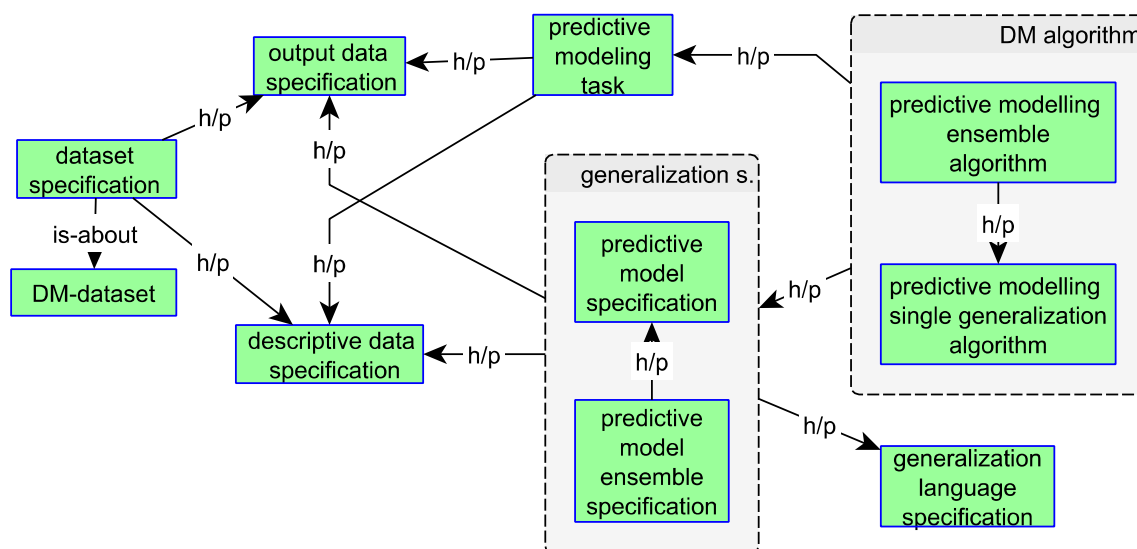
- SOPHOCLES: Self-Organised information PRocessing, Criticality and Emergence in multilevel Systems; Samoorganizirano informacijsko procesiranje, kritičnost in nujnost v večnivojskih sistemih; M. Grobelnik.
- SUMO: Supermodeling by combining imperfect models; Supermodeliranje s kombiniranjem nepopolnih modelov; S. Džeroski.
- SWING: Semantic Web Services Interoperability for Geospatial Decision Making; Interoperabilnost semantičnih spletnih servisov za podporo odločanja v geografskih domenah; M. Grobelnik, D. Mladenić.
- SYMPHONY: Orchestrating Information Technologies and Global Systems Science for Policy Design and Regulation of a Resilient and Sustainable Global Economy; Organiziranje informacijskih tehnologij in globalnih sistemov znanosti za izdelovanje politik in regulative prožne in trajnostne globalne ekonomije; D. Mladenić, M. Grobelnik.
- TAO: Transitioning Applications to Ontologies; Podpiranje aplikacij z ontologijami; M. Grobelnik, D. Mladenić.
- TOPOSYS: Topological Complex Systems; Topološki kompleksni sistemi; M. Grobelnik; **coordinators / koordinatorji**.
- Translectures: Transcription and Translation of Video Lectures; Prepis in prevod videopredavanj; D. Mladenić.
- VIDi: Visualising the impact of the legislation by analysing public discussions using statistical means; Vizualizacija vpliva zakonodaje s pomočjo statističnih metod; M. Grobelnik, D. Mladenić.
- WHIM: The What-If Machine; Kaj-če stroj; N. Lavrač, M. Žnidaršič.
- XLike: Cross Lingual Knowledge Extraction; Večjezično izločanje znanja; M. Grobelnik; **coordinators / koordinatorji**.
- XLime: crossLingual crossMedia knowledge extraction; Medjezično medmedijsko izločanje znanja; M. Grobelnik.

Priloga 2

TEHNIKA

Področje: 2.07 – Računalništvo in informatika

Dosežek: Ontologija rudarjenja podatkov OntoDM, Vir: PANOV, Panče, SOLDATOVA, Larisa N., DŽEROSKI, Sašo. Ontology of core data mining entities. Data mining and knowledge discovery, ISSN 1384-5810, Vol. 28, no. 5/6, 2014, vol. 28, no. 5-6, str. 1222-1265.



Opis splošne naloge napovednega modeliranja s samostojnimi in ansambelskimi modeli v ontologiji OntoDM.

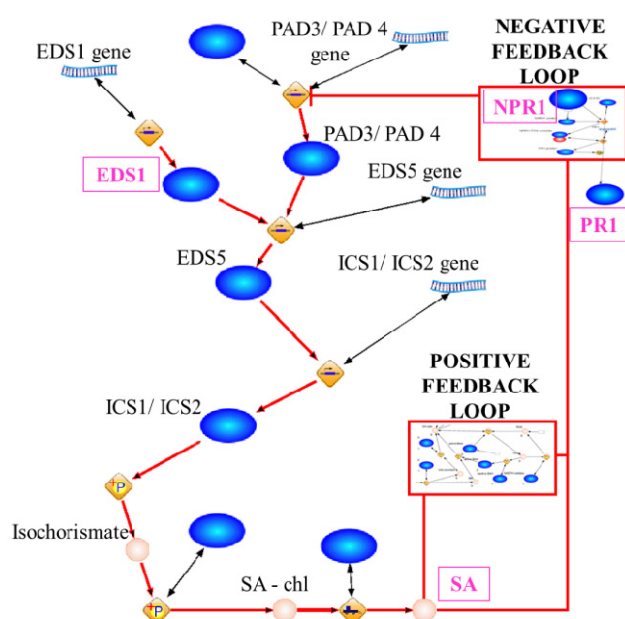
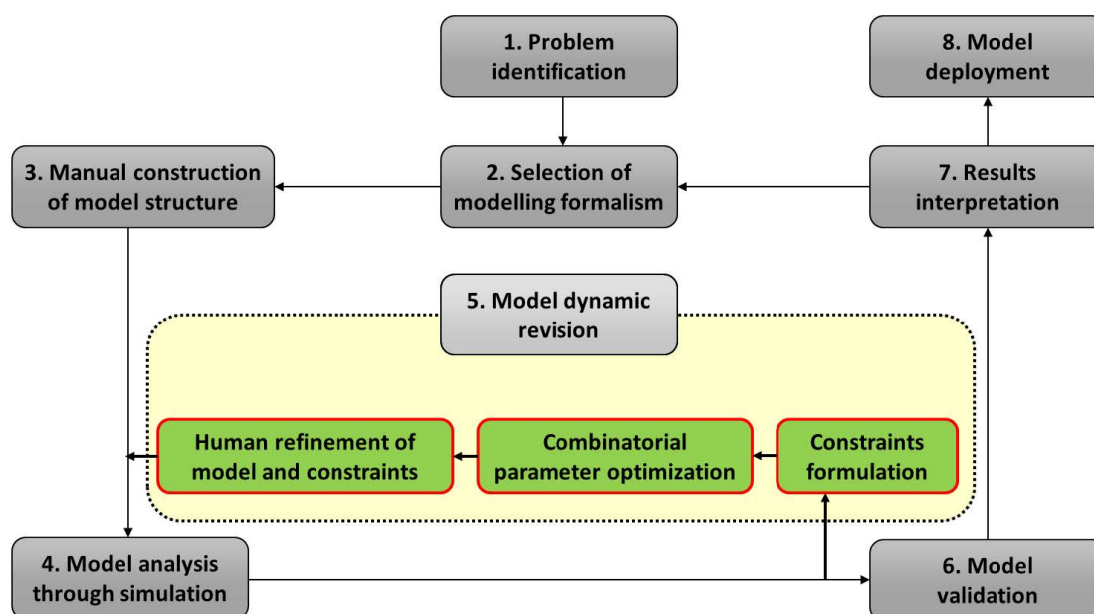
Pri podatkovnem rudarjenju se srečujemo z različnimi vrstami podatkov. Ti so običajno podani v obliki ene same tabele s primitivnimi podatkovnimi tipi, vendar se vedno pogostejše pojavljajo tudi bolj kompleksni in strukturirani podatki, kot so zaporedja, omrežja, besedila, slike in druge multimedijske vsebine. Enotno obravnavanje takšnih podatkov predstavlja velik raziskovalni izziv. Po drugi strani se pri podatkovnem rudarjenju srečujemo tudi z množico različnih nalog, ki jih običajno rešujemo z metodami, razvitimi za točno določeno nalogo in tip podatkov. Takšna razdrobljenost podatkovnega rudarjenja predstavlja oviro pri nadaljnjem razvoju področja in še posebej pri razvoju popolnoma avtomatizirane analize podatkov. Razvili smo teoretični okvir, ki enotno obravnava tako različne kompleksne tipe podatkov kot tudi različne naloge in metode podatkovnega rudarjenja, kar pomeni velik prispevek v smeri avtomatizacije odkrivanja znanja v podatkih. Razvita ontologija OntoDM formalno opredeljuje vse glavne entitete podatkovnega rudarjenja, vključno s podatki ter nalogami in metodami podatkovnega rudarjenja. Ontologija je razvita v skladu s sodobnimi standardi in načeli načrtovanja ontologij in smiselno vključuje najsodobnejše obstoječe znanstvene ontologije, kot so Basic Formal Ontology (BFO), Relational Ontology (RO) in druge. To zagotavlja skladnost in povezljivost z drugimi ontologijami in sklepanje prek več domen. OntoDM je javno dostopna na naslovu <http://www.ontodm.com>.

Priloga 3

TEHNIKA

Področje: 2.07 – Računalništvo in informatika

Dosežek: Iterativno izboljševanje bioloških modelov na osnovi domenskega znanja, Vir: MILJKOVIĆ, Dragana, DEPOLLI, Matjaž, STARE, Tjaša, MOZETIČ, Igor, PETEK, Marko, GRUDEN, Kristina, LAVRAČ, Nada. Plant defence model revisions through iterative minimisation of constraint violations. *International journal of computational biology and drug design*, ISSN 1756-0756, 2014, vol. 7, no. 1, str. 61-79.



Ena od ovir pri razvoju dinamičnega modela obrambnega odziva rastlin na virusne okužbe je pomanjkanje kinetičnih podatkov, iz katerih bi lahko določili parametre modela. Za reševanje tega problema smo razvili metodologijo, ki iterativno izboljšuje vrednosti parametrov modela v skladu z domenskim znanjem strokovnjakov, formaliziranim v obliki omejitev, ki jih morajo simulacije modela izpolnjevati. Predlagana metodologija je bila uspešno uporabljena za gradnjo modela salicilne kisline, ki je najpomembnejša komponenta obrambnega odziva rastlin. Razvito metodologijo je odslej mogoče uporabiti za izdelavo dinamičnih modelov poljubnega biološkega mehanizma, kjer je kljub pomanjkanju eksperimentalnih podatkov na voljo dovolj domenskega znanja.