

Oznaka poročila: ARRS-RPROJ-ZP-2013/182



ZAKLJUČNO POROČILO RAZISKOVALNEGA PROJEKTA

A. PODATKI O RAZISKOVALNEM PROJEKTU

1.Osnovni podatki o raziskovalnem projektu

Šifra projekta	J2-2353
Naslov projekta	Semantično odkrivanje zakonitosti v okolju spletnih servisov (SemantičniSoKD)
Vodja projekta	8949 Nada Lavrač
Tip projekta	J Temeljni projekt
Obseg raziskovalnih ur	4649
Cenovni razred	B
Trajanje projekta	05.2009 - 04.2012
Nosilna raziskovalna organizacija	106 Institut "Jožef Stefan"
Raziskovalne organizacije - soizvajalke	105 Nacionalni inštitut za biologijo
Raziskovalno področje po šifrantu ARRS	2 TEHNIKA 2.07 Računalništvo in informatika 2.07.07 Inteligentni sistemi - programska oprema
Družbeno-ekonomski cilj	13.01 Naravoslovne vede - RiR financiran iz drugih virov (ne iz SUF)

2.Raziskovalno področje po šifrantu FOS¹

Šifra	1.02
- Veda	1 Naravoslovne vede
- Področje	1.02 Računalništvo in informatika

B. REZULTATI IN DOSEŽKI RAZISKOVALNEGA PROJEKTA

3.Povzetek raziskovalnega projekta²

SLO

Cilj projekta je bil razvoj servisno usmerjenega pristopa k rudarjenju podatkov in odkrivanju znanja. Bolj specifični cilji so bili izboljšati algoritme za indukcijo opisnih pravil iz označenih primerov ter izboljšati rezultate strojnega učenja z uporabo (polavtomatsko zgrajenih) zaporedij algoritmov na osnovi semantične informacije o algoritmih, ki jih vključimo v

ontologijo konceptov rudarjenja podatkov. Te raziskave so bile delno financirane s strani tega projekta, programa Tehnologije znanja (P2-0103) in evropskega projekta 7. OP BISON (2008-2011).

Glavni rezultati projekta so nova platforma Orange 4WS, nova metodologija za semantično analizo mikromrež SegMine in organizacija treh mednarodnih delavnic SoKD-2009, SoKD-2010 in PlanSoKD-2011 s področja servisno-orientiranega odkrivanja zakonitosti v podatkih.

Orange4WS je nova odprtokodna platforma za rudarjenje podatkov, ki omogoča gradnjo delotokov rudarjenja podatkov s kombinacijo algoritmov, ki so na voljo v okviru obstoječega orodja Orange, kot tudi novih algoritmov, razpoložljivih kot servisi na svetovnem spletu. To orodje bistveno poenostavlja gradnjo kompleksnih procesov rudarjenja podatkov, ki v postopkih analize podatkov zahtevajo zaporedje mnogih algoritmov za procesiranje, rudarjenje in vizualizacijo podatkov. Ta dosežek smo objavili v reviji The Computer Journal.

V platformi Orange4WS smo implementirali novo metodologijo SegMine. Metodologija je implementirana kot delotok (ang. workflow), ki vsebuje dva sklopa algoritmov: algoritem SEGS (Searching for Enriched Gene Sets) za odkrivanje skupin diferencialno izraženih genov in sistem Biomine (Knowledge discovery in biological databases) za detekcijo novih povezav med geni. Ta kombinacija algoritmov omogoča semantično analizo izraženosti skupin genov z uporabo predznanja v obliki ontologij. Sistem, implementiran v platformi Orange4WS, smo uporabili za analizo človeških izvornih celic, kjer smo postavili tri nove znanstvene hipoteze. Ta dosežek smo objavili v reviji BMC Bioinformatics.

ANG

The aim of the project was to develop a service-oriented approach to data mining and knowledge discovery. More specific objectives were to improve algorithms for descriptive rule induction from labeled examples and improve the results of machine learning by using sequences of algorithms based on semantic information about the algorithms which is included in the data mining ontology. These studies were supported in part by the Semantic SoKD project, by the Knowledge Technologies research programme and by the EU FP7 FET project BISON (2008-2011).

The main results of the project are the new platform Orange4WS, new methodology for semantic microarray analysis called SegMine and organization of three international workshops for service-oriented discovery in data: SoKD-2009, SoKD-2010 and PlanSoKD-2011.

Orange4WS is a new open source platform for data mining, which enables creating and executing service-oriented data mining workflows, by combining data mining algorithms available within the existing tool Orange, as well as new algorithms, available as web services. This tool significantly simplifies the creation of complex data mining processes which require sequence of algorithms for processing, data mining and visualization. This achievement was published in the Computer Journal.

The new methodology called SegMine was implemented in the Orange4WS platform. It is implemented as a workflow, which contains two sets of algorithms: algorithm SEGS (Searching for Enriched Gene Sets) for detection of differentially expressed gene groups and the Biomine system (Knowledge Discovery and biological databases) for detection of new links between the genes. This combination of algorithms enables to semantically analyze gene expression groups by using the ontological knowledge. The system, implemented in the Orange4WS platform was used for the analysis of human stem cells, where we set up three new scientific hypotheses. This achievement was published in the BMC Bioinformatics Journal.

4. Poročilo o realizaciji predloženega programa dela na raziskovalnem projektu³

Cilj projekta je bil razvoj servisno usmerjenega pristopa k rudarjenju podatkov in odkrivanju znanja. Bolj specifični cilji so bili izboljšati algoritme za indukcijo opisnih pravil iz označenih primerov ter izboljšati rezultate strojnega učenja z uporabo (polavtomatsko zgrajenih) zaporedij algoritmov na osnovi semantične informacije o algoritmih, ki jih vključimo v ontologijo konceptov rudarjenja podatkov. Te raziskave so bile podprte s strani tega projekta, programa Tehnologije znanja (P2-0103) in evropskega projekta BISON (2008-2011).

1. V prvem letu projekta smo razvili ontologijo algoritmov strojnega učenja ter s koncepti te ontologije označili (anotirali) večino algoritmov strojnega učenja, ki so sestavni del orodja Orange, razvitega na Fakulteti za računalništvo in informatiko Univerze v Ljubljani. Razvili smo tudi algoritem za avtomatsko gradnjo delotokov rudarjenja podatkov (data mining workflows)

na osnovi preprostega planerja operacij (fast forward planner). Ta dosežek smo v letu 2011 objavili v reviji IEEE Trans. Autom. Sci. Eng. [COBISS.SI-ID [23993639](#)], zasnovo sistema pa leta 2010 na znanstveni konferenci z objavo v Springer Lect. notes comput. sci. [COBISS.SI-ID [24214311](#)].

2. Že v prvem letu trajanja projekta smo začeli z razvojem sistema Orange4WS, ki omogoča gradnjo delotokov rudarjenja podatkov s kombinacijo algoritmov, ki so na voljo v okviru orodja Orange, kot tudi algoritmov, ki so razpoložljivi kot servisi na svetovnem spletu. To orodje bistveno poenostavlja gradnjo kompleksnih delotokov rudarjenja podatkov, ki zahtevajo zaporedje mnogih algoritmov za procesiranje, rudarjenje in vizualizacijo podatkov. Dosežek smo leta 2012 objavili v reviji The Computer Journal [COBISS.SI-ID [25004071](#)], prej pa v nekaj začetnih objavah na delavnicah SoKD.

3. V platformi Orange4WS smo implementirali novo metodologijo SegMine. Metodologija je implementirana kot delotok (ang. workflow), ki vsebuje dva sklopa algoritmov: algoritem SEGS (Searching for Enriched Gene Sets) za odkrivanje skupin diferencialno izraženih genov in sistem Biomine (Knowledge discovery in biological databases) za detekcijo novih povezav med geni. Ta kombinacija algoritmov omogoča semantično analizo izraženosti skupin genov z uporabo predznanja v obliki ontologij. Sistem, implementiran v platformi Orange4WS, smo uporabili za analizo človeških izvornih celic, kjer smo postavili tri nove znanstvene hipoteze. Ta dosežek smo leta 2012 objavili v reviji BMC Bioinformatics [COBISS.SI-ID [25208871](#)], pred tem pa smo zasnovo sistema objavili leta 2010 v Zborniku 13. mednarodne multikonference Informacijska družba [COBISS.SI-ID [24046375](#)] in v zborniku Int. Conf. Computational Creativity [COBISS.SI-ID [23374375](#)]. Razširjeno metodologijo Segmine je možno uporabiti tudi za analizo rastlinskih patogenov, kar je predmet tekočih raziskav.

4. Organizirali smo tri mednarodne delavnice s področja servisno-orientiranega odkrivanja zakonitosti. Delavnice SoKD-2009 [COBISS.SI-ID [22817575](#)], SoKD-2010 [COBISS.SI-ID [23967271](#)] in PlanSoKD-2011 [COBISS.SI-ID [23967271](#)] smo organizirali v okviru treh zaporednih mednarodnih konferenc ECML/PKDD-2009, 2010 in 2011 na Bledu, Barceloni in v Atenah.

5. Ocena stopnje realizacije programa dela na raziskovalnem projektu in zastavljenih raziskovalnih ciljev⁴

Dosegli smo vse zastavljene cilje, kot smo jih navedli v planu implementacije projekta.

Cilj 1: Razvoj teorije za rudarjenje podatkov tretje generacije.

Razvili smo teoretski okvir za rudarjenje podatkov v obliki ontologije strojnega učenja in z njo anotirali večino algoritmov, ki so sestavni del orodja Orange. Poleg tega smo kot servisno orientirane komponente vključili servise sistema Biomine (naših partnerjev iz Univerze v Helsinkih) za povezovanje različnih bioloških baz.

Cilj 2: Razvoj arhitekture in implementacija prototipa SoKD platforme za odkrivanje znanja.

Razvili smo odprtokodno implementacijo sistema Orange4WS z grafičnim vmesnikom, ki omogoča sestavljanje različnih algoritmov in spletnih servisov za izvajanje kompleksnih delotokov rudarjenja podatkov. Orange4WS je dostopen na naslovu <http://orange4ws.ijs.si/>

Cilj 3: Testiranje sestavin SoKD platforme in aplikacija delotokov SoKD v domenah bioinformatike in biologije.

V okolju Orange4WS smo implementirali delotoke SegMine za analizo podatkov iz mikromrež. SegMine smo uporabili za analizo senescence v človeških izvornih celicah in postavili tri nove raziskovalne hipoteze o možnih mehanizmih in označevalnih genih.

Cilj 4: Razširjanje rezultatov projekta SoKD.

Organizirali smo tri mednarodne delavnice SoKD-2009, SoKD-2010 in PlanSoKD- 2011 na temo planiranja za učenje in servisno-orientiranega odkrivanja znanja. Poleg prispevkov objavljenih na teh delavnicah smo objavili tudi vrsto prispevkov na mednarodnih konferencah ter tri objave v revijah s faktorjem vpliva.

6. Utemeljitev morebitnih sprememb programa raziskovalnega projekta oziroma sprememb, povečanja ali zmanjšanja sestave projektne skupine⁵

Ni bilo sprememb programa.

7. Najpomembnejši znanstveni rezultati projektne skupine⁶

Znanstveni dosežek			
1.	COBISS ID	23993639	Vir: COBISS.SI
	Naslov	SLO	Avtomatizacija izdelave delotokov za odkrivanje znanja s pomočjo ontologije in planerja
		ANG	Automating knowledge discovery workflow composition through ontology-based planning
	Opis	SLO	Članek obravnava izziv avtomatske izdelave delotokov za odkrivanje znanja, ob danih vhodnih in zahtevanih izhodnih podatkih procesa odkrivanja znanja. Naša metodologija je sestavljena iz dveh ključnih delov. Prvi del predstavlja opredelitev formalnega pojmovanja algoritmov podatkovnega rudarjenja s pomočjo ontologije za odkrivanje znanja. Drugi del pa predstavlja sestavljanje delotokov, ki je opredeljeno kot naloga načrtovanja z uporabo domenske ontologije in z opisi nalog. Razvili smo dve vrsti algoritma za načrtovanje z veriženjem naprej. Osnovna različica prikazuje ustreznost ontologije odkrivanje znanja za načrtovanje in uporablja Planning Domain Definition Language (PDDL) za opis algoritmov. Druga pa neposredno uporablja ontologijo preko mehanizma sklepanja. Predlagani pristop je bil preizkušen na dveh primerih uporabe, enem s področja genomike in drugem iz napredne tehnike. Rezultati kažejo možnost avtomatske gradnje delotokov s pomočjo integracije načrtovanja in uporabe ontologij.
		ANG	The problem addressed in this paper is the challenge of automated construction of knowledge discovery workflows, given the types of inputs and the required outputs of the knowledge discovery process. Our methodology consists of two main ingredients. The first one is defining a formal conceptualization of knowledge types and data mining algorithms by means of knowledge discovery ontology. The second one is workflow composition formalized as a planning task using the ontology of domain and task descriptions. Two versions of a forward chaining planning algorithm were developed. The baseline version demonstrates suitability of the knowledge discovery ontology for planning and uses Planning Domain Definition Language (PDDL) descriptions of algorithms. The second directly queries the ontology using a reasoner. The proposed approach was tested in two use cases, one from scientific discovery in genomics and another from advanced engineering. The results show the feasibility of automated workflow construction achieved by tight integration of planning and ontological reasoning.
	Objavljeno v	Institute of Electrical and Electronics Engineers; IEEE transactions on automation science and engineering; 2011; Vol. 8, no. 2; str. 253-264; Impact Factor: 1.461; Srednja vrednost revije / Medium Category Impact Factor: 1.204; WoS: AC; Avtorji / Authors: Žáková Monika, Křemen Petr, Železný Filip, Lavrač Nada	
	Tipologija	1.01 Izvirni znanstveni članek	
2.	COBISS ID	25004071	Vir: COBISS.SI
	Naslov	SLO	Okolje Orange4WS za servisno orientirano rudarjenje podatkov
		ANG	Orange4WS environment for service-oriented data mining
			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. Predstavljeno okolje Orange4WS dodaja naslednje razširitve. (1) Uporaba spletnih servisov kot gradnikov delotokov

	Opis	SLO	za podatkovno rudarjenje. (2) Uporaba algoritmov za relacijsko rudarjenje podatkov v delotokih. (3) Ontologija področja odkrivanja znanja, ki opisuje komponente delotoka (podatke, znanje ter spletne servise za podatkovno rudarjenje) na abstrakten način, ki omogoča strojno interpretacijo. Ontologija je uporabljena tudi za avtomatsko gradnjo delotokov s pomočjo planerja. Vse predstavljene razširitve so prikazane na treh realnih primerih uporabe.	
		ANG	We developed a novel Service-oriented Knowledge Discovery framework and its implementation in a service-oriented data mining environment Orange4WS (Orange for Web Services), based on the existing Orange data mining toolbox and its visual programming environment, which enables manual composition of data mining workflows. The new service-oriented data mining environment Orange4WS includes the following new features: simple use of web services as remote components that can be included into a data mining workflow; simple incorporation of relational data mining algorithms; a knowledge discovery ontology to describe workflow components (data, knowledge and data mining services) in an abstract and machineinterpretable way, and its use by a planner that enables automated composition of data mining workflows. These new features are show-cased in three real-world scenarios.	
	Objavljeno v	Cambridge university press;The journals department; The Computer journal; 2012; Vol. 55, no. 1; str. 82-98; Impact Factor: 0.785;Srednja vrednost revije / Medium Category Impact Factor: 0.96; WoS: ES, ET, EW, EX; Avtorji / Authors: Podpečan Vid, Žáková Monika, Lavrač Nada		
	Tipologija	1.01 Izvirni znanstveni članek		
3.	COBISS ID	25208871	Vir: COBISS.SI	
	Naslov	SLO	SegMine delotoki za semantično analizo podatkov iz mikromrež v okolju Orange4WS	
		ANG	SegMine workflows for semantic microarray data analysis in Orange4WS	
	Opis	SLO	Razvili smo novo metodologijo SegMine za semantično analizo podatkov mikromrež in novo okolje za gradnjo delotokov Orange4WS, ki podpira vključevanje spletnih servisov. SegMine metodologija sestoji iz dveh glavnih korakov. Najprej uporabimo algoritem za semantično odkrivanje podskupin, ki poišče semantično označena pravila za identifikacijo skupin diferencialno izraženih genov. Nato uporabimo servis BioMine za odkrivanje povezav, ki omogoča kreiranje in vizualizacijo novih bioloških hipotez. Uporabnost metodologije SegMine, implementirane z delotoki v Orange4WS, smo prikazali z rezultati analize dveh množic podatkov iz mikromrež. Pri analizi senescence človeških izvornih celic smo z uporabo SegMine formulirali tri nove raziskovalne hipoteze, ki izboljšujejo razumevanje mehanizmov, ki pripeljejo do senescence.	
		ANG	We developed a new methodology SegMine for semantic analysis of microarray data by exploiting general biological knowledge, and a new workflow environment Orange4WS which supports web service integration. The SegMine methodology consists of two main steps. First, a semantic subgroup discovery algorithm is used to construct semantically annotated rules that identify enriched gene sets. Then, link discovery service BioMine is used for the creation and visualization of new biological hypotheses. The utility of SegMine, implemented as a set of workflows in Orange4WS, is demonstrated in two microarray data analysis applications. In the analysis of senescence in human stem cells, the use of SegMine resulted in three novel research hypotheses that can improve the understanding of underlying mechanisms of senescence and the identification of candidate marker genes.	
			BioMed Central; BMC bioinformatics; 2011; Vol. 12, no. 416; str. 416-1-416-16; Impact Factor: 2.751;Srednja vrednost revije / Medium Category	

Objavljeno v	Impact Factor: 1.796; A': 1; WoS: CO, DB, MC; Avtorji / Authors: Podpečan Vid, Lavrač Nada, Mozetič Igor, Kralj Novak Petra, Trajkovski Igor, Langohr Laura, Kulovesi Kimmo, Toivonen Hannu, Petek Marko, Motaln Helena, Gruden Kristina
Tipologija	1.01 Izvirni znanstveni članek

8. Najpomembnejši družbeno-ekonomski rezultati projektne skupine⁷

	Družbeno-ekonomski dosežek		
1.	COBISS ID	26383399	Vir: COBISS.SI
	Naslov	SLO	Vabljen predavanje "Advances in data mining for biomedical research"
		ANG	Invited lecture "Advances in data mining for biomedical research"
	Opis	SLO	Vodja projekta Nada Lavrač je imela vabljen predavanje "Advances in data mining for biomedical research" na konferenci The 25th IEEE International Symposium on Computer-Based Medical System, CBMS 2012, 20.-22.6., Rim, Italija. V svojem predavanju je predstavila tudi novo metodologijo SegMine implementirano v platformi Orange4WS.
		ANG	Prof. Nada Lavrač gave an invited talk titled "Advances in data mining for biomedical research" at The 25th IEEE International Symposium on Computer-Based Medical System, CBMS 2012, June 20-22, Rome, Italy. She presented the new SegMine methodology implemented in the Orange4WS platform.
	Šifra	B.04	Vabljen predavanje
	Objavljeno v	Institute of Electrical and Electronics Engineers = IEEE; CBMS 2012; 2012; 5 str.; Avtorji / Authors: Lavrač Nada	
	Tipologija	1.06 Objavljeni znanstveni prispevek na konferenci (vabljen predavanje)	
2.	COBISS ID	28301785	Vir: COBISS.SI
	Naslov	SLO	SegMine: orodje za semantično analizo podatkov iz mikromrež
		ANG	SegMine: a tool for semantic microarray data analysis
	Opis	SLO	V referatu smo predstavil grafični sistem za izdelavo bioloških delotokov SegMine. Udeleženci specializirane delavnice in naprednega tečaja o sistemski biologiji so bili v glavnem biologi in bioinformatiki, močno zainteresirani za napredna orodja za analizo velikih količin podatkov iz mikromrež.
		ANG	We presented an interactive, visual programming system for complex workflow composition, SegMine. The participants of this specialized workshop and advanced course on system biology were mostly biologists and bioinformaticians, highly interested in advanced software tools for the analysis of large microarray datasets.
	Šifra	B.03	Referat na mednarodni znanstveni konferenci
	Objavljeno v	Federation of European Biochemical Societies (FEBS); Program and Abstracts; 2011; Str. 87; Avtorji / Authors: Gruden Kristina, Podpečan Vid, Lavrač Nada, Mozetič Igor, Kralj Novak Petra, Langohr Laura, Kulovesi Kimmo, Toivonen Hannu, Petek Marko, Motaln Helena	
	Tipologija	1.12 Objavljeni povzetek znanstvenega prispevka na konferenci	
3.	COBISS ID	25433895	Vir: COBISS.SI
	Naslov	SLO	PlanSoKD delavnice: Planiranje za učenje in servisno orientirano odkrivanje znanja

	ANG	PlanSoKD workshops: Planning to Learn and Service-Oriented Knowledge Discovery
Opis	SLO	Cilji delavnic so bili odkrivanje novih možnosti v okviru servisno orientiranih tehnologij, semantičnega spleta, podatkovnih baz, avtomatskega planiranja, gradnje in uporabe delotokov ter načrtovanje in razvoj moderne programske opreme. Poseben poudarek je bil namenjen servisno orientiranemu pristopu k zlivanju informacij iz heterogenih virov ter uporabi ontologij v procesih odkrivanja znanja v podatkih.
	ANG	The workshops gathered contributions supporting third generation data mining and knowledge discovery, elaborating a service-oriented approach to information fusion, for the needs of exploratory data analysis in the framework of inductive databases, enriched with ontology information available from the Web. The workshop gathered researchers in databases, automated planning and in software engineering, for whom data mining is an "application area". The aim of the workshop was to explore the possibilities of this new area, offer a forum for exchanging ideas and experience concerning the state-of-the-art, permit to bring in knowledge gathered in different but related and relevant areas and outline new directions for research.
Šifra	B.01 Organizator znanstvenega srečanja	
Objavljeno v	2011; 64 str.; Avtorji / Authors: Kietz Jörg-Uwe, Fischer Simon, Lavrač Nada, Podpečan Vid	
Tipologija	2.31 Zbornik recenziranih znanstvenih prispevkov na mednarodni ali tuji konferenci	

9. Drugi pomembni rezultati projektne skupine⁸

Najpomembnejši rezultati raziskav v sklopu projekta Semantični SoKD so na področju razvoja novih algoritmov strojnega učenja, na področju razvoja novega orodja za povezovanje postopkov strojnega učenja Orange4WS ter aplikacij na področju bioinformatike. Orange4WS je za vse zainteresirane raziskovalce in uporabnike orodij strojnega učenja prosto dostopen na naslovu <http://orange4ws.ijs.si/>. Orodje za semantično analizo podatkov iz mikromrež SegMine pa je prav tako prosto dostopno na naslovu <http://segmine.ijs.si/>. Že zdaj ga aktivno uporabljajo sodelavci Nacionalnega instituta za biologijo za podporo pri raziskavah na področju bioinformatike.

V času trajanja projekta smo organizirali tri mednarodne delavnice s področja servisno-orientiranega odkrivanja zakonitosti, ki so potekale v sklopu uveljavljene mednarodne konference ECML PKDD. To so bile:

- "Third Generation Data Mining: Towards Service-oriented Knowledge Discovery - SoKD-09", 7.-11.9.2009, Bled, Slovenija. [COBISS.SI-ID 22817575]
- "Third Generation Data Mining: Towards Service-Oriented Knowledge Discovery - SoKD-10", 20.-24.9.2010, Barcelona, Španija. [COBISS.SI-ID 23967271]
- "Workshop on Planning to Learn and Service-Oriented Knowledge Discovery - PlanSoKD-11", 5.-9.9.2011, Atene, Grčija. [COBISS.SI-ID 23967271]

10. Pomen raziskovalnih rezultatov projektne skupine⁹

10.1. Pomen za razvoj znanosti¹⁰

SLO

Pomen projekta Semantični SoKD je razvoj nove paradigme odkrivanja znanja, ki smo ga

implementirali v prototipnem sistemu Orange4WS. Premik paradigme, v primerjavi s sedanjo tehnologijo rudarjenja podatkov, prinašajo naslednji pristopi:

- Orange4WS je privzel servisno usmerjeno paradigmo odkrivanja znanja. Ta paradigma nudi fleksibilnost za orkestracijo različnih servisov, potrebnih za iskanje, filtriranje, rudarjenje ter inteligentno zlitje podatkov, informacij in znanja.
- Orange4WS deluje v distribuiranem okolju svetovnega spleta.
- Orange4WS izvaja uporabo in zlitje vhoda in izhoda (rezultati rudarjenja in zlitja znanja), ki so blizu znanstvenemu načinu razmišljanja.

Z uporabo novih algoritmov semantičnega rudarjenja podatkov in platforme Orange4WS, ki smo jo razvili na IJS, predvidevamo izboljšave rezultatov v več aplikativnih domenah, s poudarkom na medicini in bioinformatiki. Rezultate vrednotijo eksperti z Nacionalnega inštituta za biologijo v Ljubljani.

Gledano širše pa projekt nudi pomoč znanstvenikom pri čedalje težji nalogi odkrivanja in zlitja heterogenega in distribuiranega znanja. Reševanje tega problema zahteva razvoj nove računalniške paradigme, ki integrira ideje iz različnih domen. Ustrezna rešitev tega problema bo omogočila razvoj novih tehnologij, pomembnih za vrsto aplikacij. Semantični SoKD pa sega še dlje saj nudi uporabnikom, zlasti znanstvenikom, pomoč pri odkrivanju znanja iz distribuiranih virov informacij.

Razvoj orodja Orange4WS je omogočil nadaljnje raziskave na tem področju. Kot posledica znanj, pridobljenih pri razvoju platforme Orange4WS, smo v letu 2012 v okviru programa Tehnologije znanja sodelavci IJS začeli razvijati novo platformo ClowdFlows, ki jo lahko uporabljamo za gradnjo in izvajanje delotokov rudarjenja podatkov v vseh sodobnih spletnih brskalnikih.

ANG

The Semantic SoKD project is relevant for science because it has the potential to become a new knowledge discovery paradigm through its prototype implementation in the Orange4WS platform. The Semantic SoKD paradigm shift consists of the following breakthrough approaches, compared to the current data mining technologies:

- Orange4WS adopts a service-oriented knowledge discovery paradigm. This paradigm gives the flexibility for orchestrating different services that are needed for data/information/knowledge finding, filtering, mining, intelligently fusing, etc.
- Orange4WS functions in the distributed setting of the Web.
- Orange4WS uses input and output (results of mining and fusing knowledge) that are close to the scientists' way of thinking.

We foresee improvements of results in several application domains, with an emphasis on medical and bioinformatics domains. These results are being evaluated by domain experts from the National Institute of Biology in Ljubljana.

In broader terms, this project addresses the open problem of assisting scientists with the increasingly daunting task of heterogeneous and distributed information fusion and knowledge discovery. Solving this problem requires the development of a new computational paradigm that integrates ideas from different supporting domains. An adequate solution to this problem will result in new technologies that are relevant to a range of applications. It covers issues such as knowledge management and creation, but goes beyond them in assisting users (particularly scientists).

Knowledge gained from the development of the Orange4WS platform enabled further research in this area. As a result of this in 2012 our research group Knowledge technologies started to develop a new platform called ClowdFlows which can be used for creating and executing data mining workflows in all modern web browsers.

10.2. Pomen za razvoj Slovenije¹¹

SLO

Projekt Semantični SoKD je uspešno povezal raziskovalni skupini (IJS in NIB) z različnih,

multidisciplinarnih področij (tehnologije znanja in biologije). Metodologijo SegMine, implementirano v platformi Orange4WS, ki je prosto dostopna na naslovu <http://segmine.ijs.si/>, aktivno uporabljajo sodelavci Nacionalnega instituta za biologijo za podporo raziskovanju na področju bioinformatike.

Povezal je tudi slovenske raziskovalce s tujimi raziskovalnimi skupinami (Univerza v Helsinkih) ter spodbujal skupno raziskovalno delo na mednarodnih projektih (7.OP projekt BISON). Omogočal je kvalitetno projektno izobraževanje mladih raziskovalcev, njihovo vpetost v tekoče raziskave in aktivno sodelovanje v mednarodnem okolju.

Poleg tega osnovne raziskave prispevajo k splošnemu napredku znanja na področju informacijskih tehnologij. Aplikacije na interdisciplinarnih področjih, v našem primeru bioinformatike, pa prispevajo h kreiranju novih idej na aplikativnih domenah in dvigujejo nivo uporabnosti informacijskih tehnologij.

ANG

The project Semantic SoKD successfully linked two research groups (Jozef Stefan Institute and National Institute of Biology) from different, multidisciplinary fields (knowledge technologies and biology). The SegMine methodology, implemented in the Orange4WS platform, which is publicly available at <http://segmine.ijs.si/>, has already been used by the experts from the National Institute of Biology as a support for their research work in bioinformatics.

It also linked Slovenian researchers with European research groups (University of Helsinki), and fostered joint research work on international projects (EU FP7 project BISON). It enabled high quality, project oriented education of young researchers, their involvement in current state-of-the-art research and active collaboration with international research groups.

Apart from that, the basic research contributes to a higher level of awareness and general knowledge in the area of information technologies. Interdisciplinary applications, in our case bioinformatics, contribute to formation of new ideas and insights in the application domains and raise the level of applicability of information technologies.

11. Samo za aplikativne projekte in podoktorske projekte iz gospodarstva!

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

Cilj		
F.01	Pridobitev novih praktičnih znanj, informacij in veščin	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.02	Pridobitev novih znanstvenih spoznanj	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.03	Večja usposobljenost raziskovalno-razvojnega osebja	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	

F.04	Dvig tehnološke ravni	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.05	Sposobnost za začetek novega tehnološkega razvoja	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.06	Razvoj novega izdelka	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.07	Izboljšanje obstoječega izdelka	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.08	Razvoj in izdelava prototipa	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.09	Razvoj novega tehnološkega procesa oz. tehnologije	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.10	Izboljšanje obstoječega tehnološkega procesa oz. tehnologije	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.11	Razvoj nove storitve	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.12	Izboljšanje obstoječe storitve	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	

	Uporaba rezultatov	
F.13	Razvoj novih proizvodnih metod in instrumentov oz. proizvodnih procesov	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.14	Izboljšanje obstoječih proizvodnih metod in instrumentov oz. proizvodnih procesov	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.15	Razvoj novega informacijskega sistema/podatkovnih baz	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.16	Izboljšanje obstoječega informacijskega sistema/podatkovnih baz	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.17	Prenos obstoječih tehnologij, znanj, metod in postopkov v prakso	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.18	Posredovanje novih znanj neposrednim uporabnikom (seminarji, forumi, konference)	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.19	Znanje, ki vodi k ustanovitvi novega podjetja ("spin off")	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.20	Ustanovitev novega podjetja ("spin off")	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.21	Razvoj novih zdravstvenih/diagnostičnih metod/postopkov	

	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.22	Izboljšanje obstoječih zdravstvenih/diagnostičnih metod/postopkov	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.23	Razvoj novih sistemskih, normativnih, programskih in metodoloških rešitev	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.24	Izboljšanje obstoječih sistemskih, normativnih, programskih in metodoloških rešitev	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.25	Razvoj novih organizacijskih in upravljavskih rešitev	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.26	Izboljšanje obstoječih organizacijskih in upravljavskih rešitev	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.27	Prispevek k ohranjanju/varovanje naravne in kulturne dediščine	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.28	Priprava/organizacija razstave	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.29	Prispevek k razvoju nacionalne kulturne identitete	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	

	Uporaba rezultatov	
F.30	Strokovna ocena stanja	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.31	Razvoj standardov	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.32	Mednarodni patent	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.33	Patent v Sloveniji	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.34	Svetovalna dejavnost	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	
F.35	Drugo	
	Zastavljen cilj	☐ DA ☐ NE
	Rezultat	
	Uporaba rezultatov	

Komentar

12.Samo za aplikativne projekte in podoktorske projekte iz gospodarstva!
Označite potencialne vplive oziroma učinke vaših rezultatov na navedena področja

	Vpliv	Ni vpliva	Majhen vpliv	Srednji vpliv	Velik vpliv	
G.01	Razvoj visokošolskega izobraževanja					
G.01.01.	Razvoj dodiplomskega izobraževanja	☐	☐	☐	☐	
G.01.02.	Razvoj podiplomskega izobraževanja	☐	☐	☐	☐	
G.01.03.	Drugo:	☐	☐	☐	☐	
G.02	Gospodarski razvoj					

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

Komentar

--

13. Pomen raziskovanja za sofinancerje¹²

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

14. Izjemni dosežek v letu 2012¹³**14.1. Izjemni znanstveni dosežek**

Orange4WS je novo servisno orientirano okolje za rudarjenje podatkov ter referenčna implementacija. Temelji na obstoječem programskem paketu za rudarjenje podatkov Orange ter njegovi komponenti za vizualno programiranje, ki omogoča ročno gradnjo delotokov. Dodaja pa naslednje razširitve:

- uporaba spletnih servisov kot gradnikov delotokov za podatkovno rudarjenje,
- uporaba algoritmov za relacijsko rudarjenje podatkov v delotokih,
- ontologija področja odkrivanja znanja, ki opisuje komponente delotoka (podatke, znanje, spletne servise za podatkovno rudarjenje) na abstrakten način, ki omogoča strojno interpretacijo. Ontologija je uporabljena tudi za avtomatsko gradnjo delotokov z uporabo avtomatskega planerja.

To orodje bistveno poenostavi gradnjo kompleksnih procesov rudarjenja podatkov, ki v postopkih analize podatkov zahtevajo zaporedje mnogih algoritmov za procesiranje, rudarjenje in vizualizacijo.

Platforma Orange4WS je prosto dostopna na naslovu <http://orange4ws.ijs.si/>.

14.2. Izjemni družbeno-ekonomski dosežek

Nova metodologija SegMine za semantično analizo podatkov mikromrež sestoji iz dveh glavnih korakov. Najprej uporabimo algoritem za semantično odkrivanje podskupin, ki poišče semantično označena pravila za identifikacijo skupin diferencialno izraženih genov. Nato uporabimo servis BioMine za odkrivanje novih povezav med geni, ki omogoča kreiranje in vizualizacijo novih bioloških hipotez. Uporabnost metodologije SegMine, implementirane z delotoki v Orange4WS, smo prikazali z rezultati analize dveh množic podatkov iz mikromrež. Pri analizi senescence človeških izvornih celic smo formulirali tri nove raziskovalne hipoteze, ki izboljšujejo razumevanje mehanizmov, ki pripeljejo do senescence.

Orodje SegMine je prosto dostopno na naslovu <http://segmine.ijs.si/>. Aktivno ga uporabljajo sodelavci Nacionalnega instituta za biologijo za podporo pri raziskovah na področju bioinformatike.

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 ter obdelavo teh podatkov za evidence ARRS
- so vsi podatki v obrazcu v elektronski obliki identični podatkom v obrazcu v pisni obliki
- so z vsebino zaključnega poročila seznanjeni in se strinjajo vsi soizvajalci projekta

Podpisi:

*zastopnik oz. pooblaščen oseba
raziskovalne organizacije:*

in

vodja raziskovalnega projekta:

Institut "Jožef Stefan"

Nada Lavrač

ŽIG

Kraj in datum: Ljubljana 13.3.2013

Oznaka prijave: ARRS-RPROJ-ZP-2013/182

¹ Opredelite raziskovalno področje po klasifikaciji FOS 2007 (Fields of Science). Prevajalna tabela med raziskovalnimi področji po klasifikaciji ARRS ter po klasifikaciji FOS 2007 (Fields of Science) s kategorijami WOS (Web of Science) kot podpodročji je dostopna na spletni strani agencije (<http://www.arrs.gov.si/sl/gradivo/sifranti/preslik-vpp-fos-wos.asp>). [Nazaj](#)

² Napišite povzetek raziskovalnega projekta (največ 3.000 znakov v slovenskem in angleškem jeziku) [Nazaj](#)

³ Napišite kratko vsebinsko poročilo, kjer boste predstavili raziskovalno hipotezo in opis raziskovanja. Navedite ključne ugotovitve, znanstvena spoznanja, rezultate in učinke raziskovalnega projekta in njihovo uporabo ter sodelovanje s tujimi partnerji. Največ 12.000 znakov vključno s presledki (približno dve strani, velikost pisave 11). [Nazaj](#)

⁴ Realizacija raziskovalne hipoteze. Največ 3.000 znakov vključno s presledki (približno pol strani, velikost pisave 11) [Nazaj](#)

⁵ V primeru bistvenih odstopanj in sprememb od predvidenega programa raziskovalnega projekta, kot je bil zapisan v predlogu raziskovalnega projekta oziroma v primeru sprememb, povečanja ali zmanjšanja sestave projektne skupine v zadnjem letu izvajanja projekta, napišite obrazložitev. V primeru, da sprememb ni bilo, to navedite. Največ 6.000 znakov vključno s presledki (približno ena stran, velikost pisave 11). [Nazaj](#)

⁶ Navedite znanstvene dosežke, ki so nastali v okviru tega projekta. Raziskovalni dosežek iz obdobja izvajanja projekta (do oddaje zaključnega poročila) 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](#)

⁷ Navedite družbeno-ekonomske dosežke, ki so nastali v okviru tega projekta. Družbeno-ekonomski rezultat iz obdobja izvajanja projekta (do oddaje zaključnega poročila) 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 projekta ... - v obeh primerih ni COBISS ID). [Nazaj](#)

⁸ Navedite rezultate raziskovalnega projekta iz obdobja izvajanja projekta (do oddaje zaključnega poročila) v primeru, da katerega od rezultatov ni mogoče navesti v točkah 7 in 8 (npr. ker se ga v sistemu COBISS ne vodi). Največ 2.000 znakov, vključno s presledki. [Nazaj](#)

⁹ Pomen raziskovalnih rezultatov za razvoj znanosti in za razvoj Slovenije bo objavljen na spletni strani: <http://sicris.izum.si/> za posamezen projekt, ki je predmet poročanja [Nazaj](#)

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

¹¹ Največ 4.000 znakov, vključno s presledki [Nazaj](#)

¹² Rubrike izpolnite / prepisite skladno z obrazcem "izjava sofinancerja" <http://www.arrs.gov.si/sl/progproj/rproj/gradivo/>, ki ga mora izpolniti sofinancer. Podpisan obrazec "Izjava sofinancerja" pridobi in hrani nosilna raziskovalna organizacija – izvajalka projekta. [Nazaj](#)

¹³ Navedite en izjemni znanstveni dosežek in/ali en izjemni družbeno-ekonomski dosežek raziskovalnega projekta v letu 2012 (največ 1000 znakov, vključno s presledki). 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/analize/dosez/>. [Nazaj](#)

Obrazec: ARRS-RPROJ-ZP/2013 v1.00

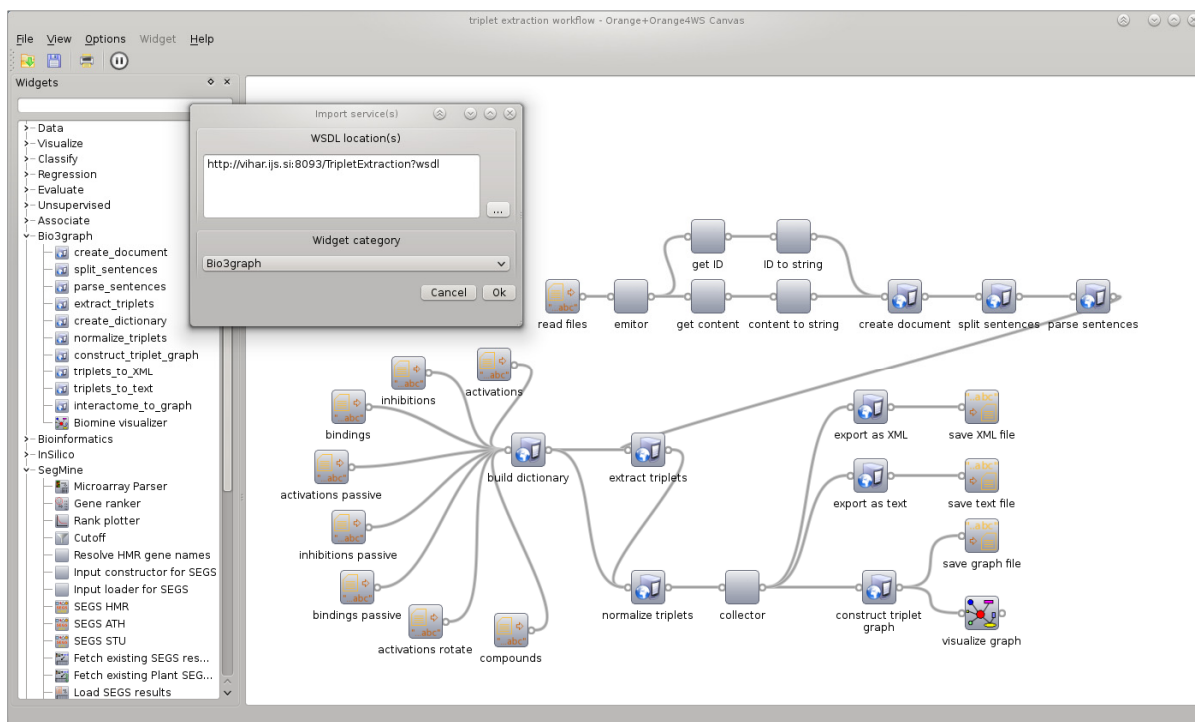
F9-40-E7-E7-E0-D3-03-FC-7A-F2-66-CB-50-57-31-EB-CE-1D-B2-57

TEHNIKA

Področje: 2.07 – Računalništvo in informatika

Dosežek 1: Okolje Orange4WS za servisno orientirano rudarjenje podatkov

Vir: PODPEČAN, Vid, ŽÁKOVÁ, Monika, LAVRAČ, Nada. Orange4WS environment for service-oriented data mining. *Comput. j.*, 2012, vol. 55, no. 1, str. 82-98, doi: 10.1093/comjnl/bxr077. [COBISS.SI-ID 25004071]



Orange4WS je novo servisno orientirano okolje za rudarjenje podatkov ter referenčna implementacija. Temelji na obstoječem programskem paketu za rudarjenje podatkov Orange in njegovi komponenti za vizualno programiranje, ki omogoča ročno gradnjo delotokov. Predstavljeno okolje Orange4WS pa 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 (podatke, znanje ter spletne servise za podatkovno rudarjenje) na abstrakten način, ki omogoča strojno interpretacijo. Ontologija je uporabljena tudi za avtomatsko gradnjo delotokov z uporabo avtomatskega planerja. Vse predstavljene razširitve so bile prikazane na treh realnih primerih uporabe.

Platforma Orange4WS bistveno poenostavlja gradnjo kompleksnih procesov rudarjenja podatkov, ki v postopkih analize podatkov zahtevajo zaporedje več algoritmov za procesiranje, rudarjenje in vizualizacijo podatkov. Poleg tega pa je njena uporaba izboljšala razumljivost rezultatov v več aplikativnih domenah na področju medicine in bioinformatike.

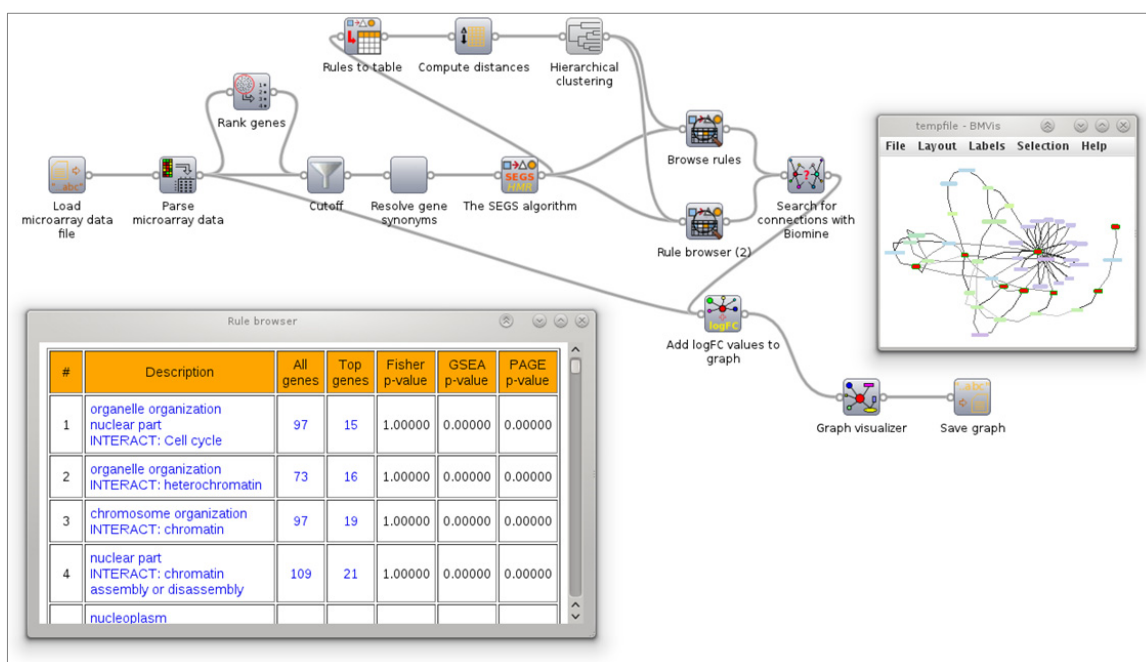
Platforma Orange4WS je prosto dostopna na naslovu <http://orange4ws.ijs.si/>

TEHNIKA

Področje: 2.07 – Računalništvo in informatika

Dosežek 1: **SegMine: orodje za semantično analizo podatkov iz mikromrež**

Vir: PODPEČAN, Vid, LAVRAČ, Nada, MOZETIČ, Igor, KRALJ NOVAK, Petra, TRAJKOVSKI, Igor, LANGOHR, Laura, KULOVESI, Kimmo, TOIVONEN, Hannu, PETEK, Marko, MOTALN, Helena, GRUDEN, Kristina. SegMine workflows for semantic microarray data analysis in Orange4WS. *BMC bioinformatics*, 2011, vol. 12, no. 416, str. 416-1-416-16, doi: 10.1186/1471-2105-12-416. [COBISS.SI-ID 25208871]



SegMine je nova metodologija za semantično analizo podatkov DNA mikromrež. Metodologija je implementirana kot delotok, ki vsebuje dva sklopa algoritmov: algoritem SEGS (Searching for Enriched Gene Sets) za odkrivanje skupin diferencialno izraženih genov in sistem Biominer (Knowledge discovery in biological databases) za odkrivanje novih povezav med geni. Ta kombinacija algoritmov omogoča semantično analizo izraženosti skupin genov z uporabo predznanja v obliki ontologij. SegMine, implementiran v platformi Orange4WS, smo v sklopu projekta »Semantično odkrivanje zakonitosti v okolju spletnih servisov« uporabili za analizo človeških izvornih celic, kjer smo postavili tri nove znanstvene hipoteze za izboljšano razumevanje mehanizmov, ki privedejo do senescence. Orodje SegMine aktivno uporabljajo sodelavci Nacionalnega inštituta za biologijo za podporo pri raziskovah na področju bioinformatike.

Orodje SegMine je prosto dostopno na naslovu <http://segmine.ijs.si/>.