DEPARTMENT OF KNOWLEDGE TECHNOLOGIES E-8

The Department of 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. Established areas of knowledge technologies include intelligent data analysis (machine learning, data mining, and knowledge discovery in databases), semantic data mining and semantic web, language technologies and computational linguistics, decision support and knowledge management. In addition to knowledge technologies research, we also develop applications in environmental sciences and ecology, medicine and health care, biomedicine and bioinformatics, economy and marketing.

Head:
Prof. Nada Lavrač

In 2011 we were involved in 11 national and 12 European projects, most of them founded by the seventh framework programme (7FP). In terms of our collaboration in the EU projects, we are the most successful programme group in Slovenia.

In the area of **intelligent data analysis and data mining** we have developed several new methods. (a) The new approach for creating and implementing data mining workflows in our service-oriented knowledge discovery platform Orange4WS was published in the Computer Journal. (b) We implemented the SegMine methodology in the Orange4WS platform that, together with the BioMine system for detection of new links between genes, allows a semantic gene expression analysis by using bio-ontologies as background knowledge; this work was published in the BMC Bioinformatics journal. (c) We have developed a new system for bisociative knowledge discovery called CrossBee for discovering new connections between different medical domains. (d) We have developed new methods for learning regression trees that take into account spatial or network autocorrelation in the values of the target variable. (e) We have developed new methods for learning from data streams, in particular the methods for learning different kinds of ensembles (including the option trees for regression). (f) We have developed new methods for structured output prediction, based on the nearest-neighbour principle. We have successfully concluded our work in the 7FP project BISON (Bisociation Networks for Creative Information Discovery), where we developed the methods for discovering bisociative links between different contexts and domains. We have collaborated in the 7FP project on data mining called e-LICO (e-Laboratory for Interdisciplinary Collaborative Research in Data Mining and Data-Intensive Sciences), where we developed new web services for a subgroup discovery and co-organized a successful Videolectures.net Challenge aimed at developing improved recommendation system for video lectures. We continued successful collaboration within the 7FP project PHAGOSYS (Systems biology of phagosome formation and maturation – modulation by intracellular pathogens), where the emphasis was on the use of meta-heuristic methods for parameter estimation in models of dynamic systems, in particular for the task of modelling endocytosis, an important process of immune response. We became partners in two new 7FP projects: SUMO (Supermodelling by combining imperfect models) and REWIRE (Rehabilitative Wayout In Responsive home Environments). The SUMO project deals with the problem of learning supermodels (ensemble models of dynamic systems that combine the existing models), and with the use of supermodels in climate modelling. The REWIRE project focuses on the development of a new approach to rehabilitation, which uses virtual-reality-based games for rehabilitation at home and wearable sensors (as well as machine learning) to monitor and adaptively plan rehabilitation.

We successfully organized the "13th Conference on Artificial Intelligence in Medicine" (AIME 2011) in Bled, Slovenia.

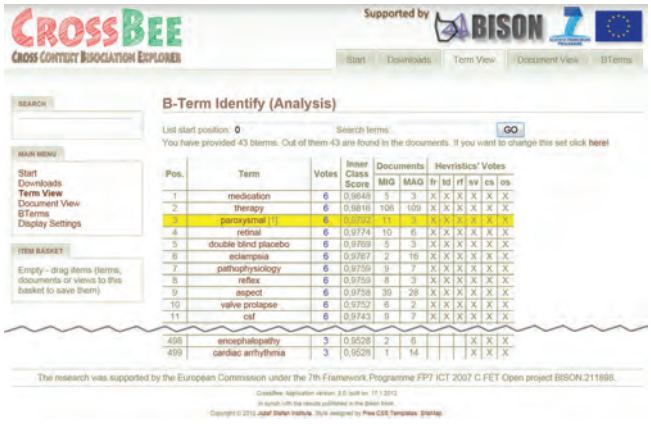

Figure 1: CrossBee is a web application that implements an algorithm for discovering hidden links in literatures from different scientific domains. It provides the user with a list of key terms (shown in the figure) which have the potential for cross-domain scientific discovery.

In the area of **text and web mining and heterogeneous information networks analysis** we participated in three European projects: FIRST (large scale information extraction and integration infrastructure for supporting financial decision making), ENVISION (ENVIronmental Services Infrastructure with ONtologies), and FOC (Forecasting Financial Crises).We are one of

Tomaž Erjavec, together with Matija Ogrin from ZRC SAZU, again received the Google Digital Humanities Research Award for his work on language models for historical Slovene.

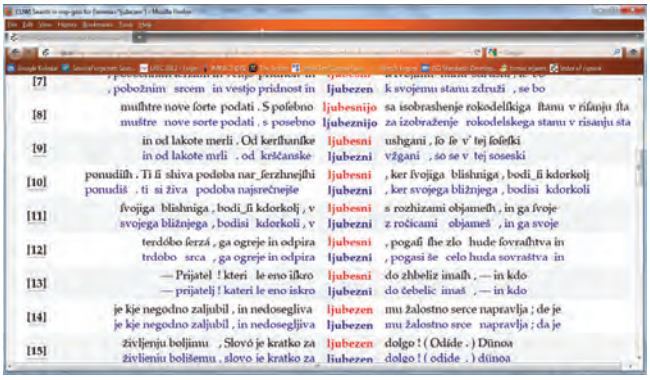

Figure 2: Concordances over a corpus of historical Slovene with modernised forms of words. The corpus and associated programs have been developed in the scope of the EU project IMPACT and the Google Award "Developing language models for historical Slovene".

the key partners in the 7FP project FIRST where we focus on analyzing large amounts of dynamic and heterogeneous sources of financial information. The daily work and the business success of all decision makers in this industry depend on the availability of such highly trustable, easily acquirable information. Our focus in this project is online data mining (financial news, blogs and tweets) in near-real time over vast amounts of constantly evolving data. In 2011 we designed the architecture and implemented a software system for capturing and annotating online texts in real time. We daily monitor 200 financial web sources and capture about 40,000 pieces of news per day. In the future we will focus on data analysis and solving practical problems for end users (financial-product sentiment analysis, analysis of financial institutions' reputation, online deceit detection). In the ENVISION project we developed a methodology to support the semantic description of geographic data and services. We specifically focused on supporting multilinguality in ontology management that enables end users to search for the basic building blocks of semantic descriptions. This methodology was implemented as a web application and will represent one of the modules of the on-line environmental decision support portal. In September 2011 we started working on the7FP project FOC that focuses on forecasting financial crises. Our role in this project is to analyze non-financial crisis indicators based on the sentiment analysis of large streams of textual data.

Elena Ikonomovska received Best Submission Award at the doctoral symposium of the EDBT/ ICDT 2011 Joint Conference.

In the area of **language technologies** we applied the methodology developed for contemporary Slovene to the development of annotated corpora, computational lexica and an annotation tool for historical Slovene. These resources are important for a better accessibility of the Slovenian written cultural heritage on the dLib.si digital library portal of the National and University Library of Slovenia (NUL), as well as for enabling empirically supported diachronic studies of the Slovene language. The work on the development of the language resources for historical Slovene was undertaken within the scope of the 7FP project IMPACT (Improving Access to Text), in which we collaborate with the NUL, and as part of the Google awarded project "Developing language models for historical Slovene", in which we collaborate with the Scientific Research Centre of the Slovenian Academy of Sciences and Arts (SRC SASA). In the scope of the national project "Unknown 17th and 18th century manuscripts of Slovenian literature: information-technology aided register, scholarly editions and analyses", we continued joint work with SRC SASA on complex digital editions of older Slovenian literature, where we implemented a Fedora Commons platform for manuscript viewing and searching. The platform was also used to host a digital edition of the Slovenian Biographic Lexicon, developed by SASA. In the second half of 2011 we also started joint work with SRC SASA on the national project "Leading Slovenian humanists from the 16th to the mid-19th century". We continued work on linguistic annotation of parallel bi-lingual corpora, performed in the scope of the national project "Slovenian translation studies – resources and research" under the lead of the Department for Translation Studies at the Arts Faculty of the University of Ljubljana. The corpora are to be used to enable linguistic studies of translation processes, while also being useful for the development of multilingual language technologies. In cooperation with the same department we also continued work on enlarging the scope and variety of information in the Slovenian semantic lexicon sloWNet. In the field of access to and standardisation of language resources we worked on the EU project FlareNet "Fostering Language Resources Network". As one of the founding members we joined, at the end of 2010, the COST project MUMIA (Multilingual and Multifaceted Interactive Information Access) and the European ESF network NetWordS (The European Network on Word Structure). We intensely collaborated in the work of the Slovenian Institute of Standardisation, as the Slovenian delegates of ISO/TC37/SC4 »Terminology and Other Language and Content Resources/Language Resources Management«, by taking part in the meetings of ISO TC 37 and by reviewing, translating and approving Slovenian standards from this field. At the Slovenian Ministry of Culture we are active in the preparation of the National Programme for

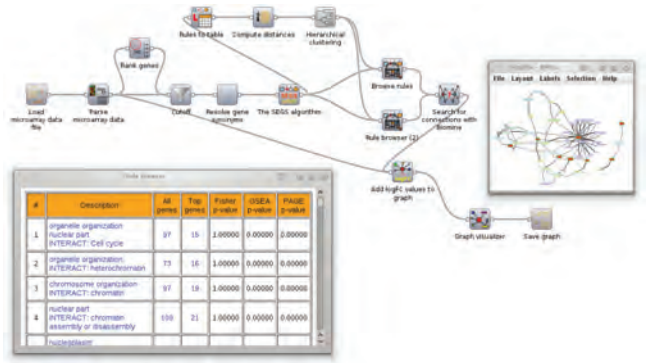

Figure 3: Screenshot of a workflow in the Orange4WS environment that implements the SegMine methodology. SegMine enables the construction of new biological hypotheses by linking the results of a semantic analysis of experimental microarray data to the knowledge available in the public biological databases.

Sašo Džeroski gave an invited lecture at the 9th International Conference on Formal Concept Analysis, held in Nicosia, Cyprus, in May 2011.

the Language Policy 2012-2016, and in taking the steps necessary for Slovenia to join the research infrastructure CLARIN (Common Language Resources and Technology Infrastructure).

In the area of **decision support** our long-term goal is to develop the methods and techniques of decision modelling, support them with software and integrate them with the data mining systems. In 2011 we implemented a method for qualitative multi-attribute decision making DEX on the Decision Deck platform. Decision Deck is an open platform that facilitates a uniform incorporation and connection of various multi-attribute modelling methods and algorithms. We developed a new version (3.03) of the computer program for multi-attribute decision making DEXi by introducing better facilities for the utility functions' definition and data interchange. We also developed new methods for ranking the alternatives in qualitative multi-attribute models, based on copulas, which improve the sensitivity of decision models and alleviate some drawbacks of the existing methods.

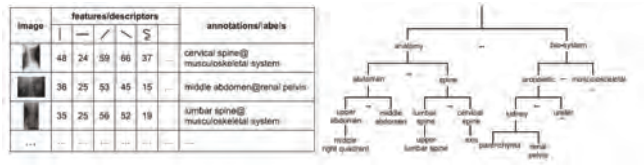

Figure 4: Hierarchical annotation of medical images, where a set of images is given with their visual descriptors and annotations from a hierarchy of labels.

Some outstanding publications in the last year

1. Vid Podpečan, Nada Lavrač, Igor Mozetič, Petra Kralj Novak, Igor Trajkovski, Laura Langohr, Kimmo Kulovesi, Hannu Toivonen, Marko Petek, Helena Motaln, Kristina Gruden. SegMine workflows for semantic microarray data analysis in Orange4WS. BMC Bioinformatics, 2011, vol. 12, no. 416, pp. 1-16, doi: 10.1186/1471-2105-12-416
2. Vid Podpečan, Monika Zemenova, Nada Lavrač. Orange4WS environment for service-oriented data mining. The Computer Journal., [in press] 2011, 17 pp., doi:10.1093/comjnl/bxr077.
3. Daniela Stojanova, Andrej Kobler, Peter Ogrinc, Bernard Ženko, Sašo Džeroski. Estimating the risk of fire outbreaks in the natural environment. Data Mining and Knowledge Discovery, [in press] 2011, 10 pp., doi: 10.1007/s10618-011-0213-2.
4. Tomaž Erjavec. MULTEXT-East : Morphosyntactic resources for Central and Eastern European languages. Language Resources and Evaluation, [in press] 2011, 12 pp., doi: 10.1007/s10579-011-9174-8.
5. Ivica Dimitrovski, Dragi Kocev, Suzana Loskovska, Sašo Džeroski. Hierarchical annotation of medical images. Pattern Recognition, 2011, vol. 44, no. 10/11, pp. 2436-2449.

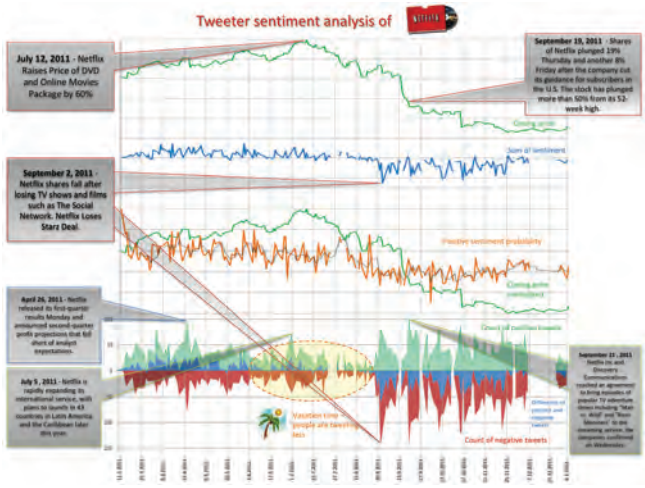

Figure 5: Real-time sentiment analysis of financial tweets, performed within project FIRST. In the case of the Netflix stock, the graphs show that the sentiment dropped before the big price plunge on 19 September 2011.

Organization of conferences, congress and meetings

1. Subconference: Intelligent system, Information Systems 2011, Ljubljana, Slovenia, 10–14 Oct. 2011
2. 13th Conference on Artificial Intelligence in Medicine AIME'11, Bled, Slovenia, 2–6 Jul. 2011
3. Project meeting of European project ENVISION, Kranjska Gora, Slovenia, 25–28 Jan. 2011
4. Project meeting of European project FIRST, Bled, Slovenia, 16–18 May 2011

INTERNATIONAL PROJECTS

1. Rehabilitative Wayout In Responsive home Environments
   REWIRE
   7. FP, 287713
   EC; Prof. Alberto Borghese, Universita degli Studi di Milano, Milano, Italy
   Prof. Sašo Džeroski
2. Supermodeling by Combining Imperfect Models - Enlarged EU
   SUMO
   7. FP, 266722
   EC; Prof. Ljupco Kocarev, Macedonian Academy of Sciences and Arts, Skopje, Macedonia
   Prof. Sašo Džeroski
3. Forecasting Financial Crises - Enlarged EU
   FOC II
   7. FP, 255987
   EC; Dr. Guido Caldarelli, Consiglio Nazionale delle Ricerche, Rome, Italy
   Dr. Igor Mozetič, Miha Grčar, B. Sc.
4. e-Laboratory for Collaborative Interdisciplinary Research in Data Mining and Data Intensive Sciences
   e-LICO
   7. FP, 257680
   EC; Dr. Mélanie Hilario, Universite de Geneve, Carouge, Switzerland
   Prof. Nada Lavrač, Asst. Prof. Martin Žnidaršič, Mitja Jermol, M. Sc.
5. Large Scale Information Extraction and Integration Infrastructure for Supporting Financial Decision Making
   FIRST
   7. FP, 257928
   EC; Tomas Pariente Lobo, Atos Origin Sociedad Anónima Espanola, Madrid, Spain
   Miha Grčar, B. Sc., Prof. Nada Lavrač
6. Improving Access to Text
   IMPACT
   7. FP, 215064
   EC; Hildelies Balk, Koninklijke Biblioteek, Haag, The Netherlands
   Asst. Prof. Tomaž Erjavec, Jan Jona Javoršek, B. Sc.
7. Systems Biology of Phagosome Formation and Maturation, Modulation by Intracellula Pathogens

PHAGOSYS
7. FP, 223451, HEALTH-F4-2008-223451
EC; Dr. Brian D. Robertson, Imperial College London, Centre for Molecular Microbiology and Infection, London, Great Britain
Prof. Sašo Džeroski

8. Bisociation Networks for Creative Information Discovery
BISON
7. FP, 211898
EC; Prof. Michael Berthold, Universität Konstanz, Konstanz, Germany
Prof. Nada Lavrač

9. Environmental Services Infrastructures with Ontologies
ENVISION
7. FP, 249120
EC; Bjorn Skjellaug, Arne J. Berre, Stiftelsen Sintef, Trondheiim, Norway
Miha Grčar, B. Sc., Prof. Nada Lavrač, Prof. Dunja Mladenić, Mitja Jermol, M. Sc.

10. Fostering Language Resources Network
FLaReNet
e-Contentplus
ECP-2007-LANG-617001
EC; Dr. Nicoletta Calzolari, Consiglio Nazionale delle Ricerche (CNR-ILC), Rome, Italy
Asst. Prof. Tomaž Erjavec

11. Multilingual and Multifaceted Interactive Information Access
MUMIA
COST IC1002
EC; Dr. Michail Salampasis, Alexander Technological Educational Institute (T.E.I.) of Thessaloniki, Thessaloniki, Greece
Dr. Igor Mozetič, Prof. Nada Lavrač

12. The European Network on Word Structure
NetWordS
RNP Ref. Number 09-RNP-08
ESF; Vito Pirrelli, Institute for Computational Linguistics, Italian National Research Council, Pisa, Italy
Asst. Prof. Tomaž Erjavec

13. Google Digital Humanities Research Award
Award dtd. 13. 12. 2011
Google Inc, Mountain View, CA, USA
Asst. Prof. Tomaž Erjavec

14. Identifying Optimal Management Strategies for Biodiversity and Related Ecosystem Services on Private Forestes
BI-US/11-12-048
Prof. Donald Glenn Hodges, University of Tennessee-Knoxville College of Agricultural Sciences and Natural Resources, Department of Forestry, Wildlife and Fisheries, Knoxville, USA
Prof. Marko Debeljak

R & D GRANTS AND CONTRACTS

1. Advanced ML methods for automated modeling of dynamic systems
Prof. Sašo Džeroski
2. Data mining for integrative data analysis in systemic biology
Prof. Sašo Džeroski
3. Semantic rule discovery in the context of Web services
Prof. Nada Lavrač
4. Systemic biology approaches to analyzing interactions between pathogens and plants
Prof. Nada Lavrač
5. Growth and defense trade-offs in multitrophic interaction between potato and its two major pests
Prof. Nada Lavrač
6. Slovene translation studies - resources and research
Asst. Prof. Tomaž Erjavec
7. The leading humanists in the Slovenian territory between the 16th and mid-19th centuries and their social and cultural environment
Asst. Prof. Tomaž Erjavec
8. Ecological restoration of natural disturbances in forests
Prof. Marko Debeljak
9. Unknown 17th and 18th century manuscripts of Slovenian literature: information-technology aided register, scholarly editions and analyses
Asst. Prof. Tomaž Erjavec
10. Analysis and Scenario of Development and Exploration of Forests in Slovenia
Prof. Marko Debeljak

RESEARCH PROGRAM

1. Knowledge Technologies
Prof. Nada Lavrač

MENTORING

Ph. D. Theses

1. Ivica Dimitrovski, Generic system for content - based image retrieval (mentor Suzana Loškovska; co-mentor Sašo Džeroski).
2. Andrej Kobler, New methods of processing aerial laser scanner data for forest ecosystem monitoring (mentor Krištof Oštir; co-mentor Sašo Džeroski).
3. Dragi Kocev, Ensembles for predicting structured outputs (mentor Sašo Džeroski).
4. Viljem Pavlovič, Modelling of alpha-acid content early prediction by hops (Humulus lupulus L.) with machine learning models (mentor Črtomir Rozman; co-mentor Marko Bohanec).
5. Aleksander Pur, Model for monitoring and assessment of a public health care network (mentor Marko Bohanec).

M. Sc. Theses

1. Robert Čebron, Design of an information system to support education in the secondary school (mentor Bojan Cestnik).
2. Vitko Črep, Information system for the management of multi-residential buildings and market opportunities for housing managers (mentor Bojan Cestnik).
3. Silvester Jeršič, The use of UML in procedures of standards ISO 9001 and 14001 (mentor Bojan Cestnik).
4. Marijan Merljak, Implementing CRM into Business Operations (mentor Bojan Cestnik).
5. Janko Skok, Impacts of forest management on biodiversity : small mammals in the fir-beech forest of Mt. Snežnik as a model group (mentor Boris Kryštufek; co-mentor Marko Debeljak).
6. Li Xiaobin, Implementing DEXi evaluation models in Decision Deck platform (mentor Marko Bohanec).
7. Iztok Zajc, Renewal of an information system for logistic support of firefighting interventions (mentor Bojan Cestnik).

Bologna M. Sc. Thesis

1. Petra Barber, Use of multi-attribute decision support model in public procurement (mentor Marko Bohanec; co-mentor Ljupčo Todorovski).

VISITORS FROM ABROAD

1. Martin Saveski, Skopje, Macedonija, 1 Jan.–1 Nov. 2011
2. Dr. Michelangelo Ceci, Dipartimento di Informatica, Universita degli Studi di Bari, Bari, Italy, 9 Jan.–4 Apr. 2011
3. Prof. Dr. Don Hodges, Univerza v Tennesseeju, Institute of Agroculture, Knoxwille, USA, 15 Jan.–1 Jul. 2011
4. Mag. Ivica Dimitrovski, Faculty of Electrical Engineering and Information Technologies, University Ss. Cyril and Methodius, Skopje, Macedonia, 16.–23. 4. 2011 in 9.–31. 1. 2011
5. Jovan Tanevski, University Ss. Cyril and Methodius, Skopje, Makeconia, 5. 3.–3. 4. 2011
6. Dr. Werner Dubitsky, University of Ulster, Coleranine, Ireland, 31 Mar.–1 Apr. 2012
7. Oliver Schmidt, University of Ulster, Coleranine, Ireland, 31 Mar.–1 Apr. 2012
8. Dr. Cesare Furlanello, Fodacione Bruno Kessler, Trento, Italy, 16–23 Apr. 2011
9. Dr. Christopher Barnes, Imperial College London, London, Great Britain, 19–22 Jun. 2011
10. Dr. Boris Delibašić, University of Belgrade, Faculty of Organizational Sciences, Beograd, Republic of Serbia, 6–8 Jul. 2011
11. Dr. Miloš Jovanović, University of Belgrade, Faculty of Organizational Sciences, Beograd, Republic of Serbia, 6–8 Jul. 2011
12. Dr. Milan Vukičević, University of Belgrade, Faculty of Organizational Sciences, Beograd, Republic of Serbia, 6–8 Jul. 2011
13. Prof. Dr. Barry Hardy, Douglas Connect, Zeiningen, Switzerland, 29.–31. 8. 2011
14. Prof. Dr. Joost Kok, Leiden Institute of Advanced Computer Science, Leiden University, Leiden, Netherlands, 11–13 Sept. 2011
15. Dr. William Klement, Ottawa–Carleton Institute for Computer Science, School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Ontario, Canada, 26–30 Sept. 2011
16. Dr. Richard Wheeler, University of Edinburgh, Edinburgh, Scotland, 23 May–12 Jun. 2011 and 10–17 Oct. 2011
17. Dr. Gregory Duane, Macedonian Academy of Science and Arts, Skopje, Macedonia, 26 Oct. 2011
18. Dr. Stijn Meganck, The Institute for the encouragement of Scientific Research and Innovation, Vrije Universiteit, Brussels, Belgium, 30 Nov.–1 Dec. 2011
19. Dr. Taminau Jonatan, The Institute for the encouragement of Scientific Research and Innovation, Brussels, Belgium, 30 Nov.–1 Dec. 2011
20. Dr. Dragan Gamberger, Institut Rudjer Bošković, Zagreb, Croatia, 30 Nov. 2011
21. Dr. Larisa Soldatova, Aberystwyth University, Department of Computer Science, Aberystwyth, Great Britain, 2–17 Dec. 2011

STAFF

Researchers
1. Prof. Marko Bohanec
2. Prof. Bojan Cestnik*
3. Prof. Marko Debeljak
4. Prof. Sašo Džeroski
5. Asst. Prof. Tomaž Erjavec
6. **Prof. Nada Lavrač, Head**
7. Prof. Tanja Urbančič*

Postdoctorial associates
8. Dr. Dragi Kocev
9. Dr. Petra Kralj Novak
10. Dr. Aneta Trajanov
11. Asst. Prof. Bernard Ženko
12. Asst. Prof. Martin Žnidaršič

Postgraduates
13. Darko Čerepnalkoski, B. Sc.
14. *Valentin Gjorgjioski, B. Sc., left 01.06.11*
15. Miha Grčar, B. Sc.
16. Elena Ikonomovska, M. Sc.
17. Matjaž Juršič, B. Sc.
18. Janez Kranjc, B. Sc.
19. Panče Panov, B. Sc.
20. Vid Podpečan, B. Sc.
21. Nikola Simidjievski, B. Sc.
22. Ivica Slavkov, B. Sc.
23. Borut Sluban, B. Sc.
24. Nejc Trdin, B. Sc.
25. Anže Vavpetič, B. Sc.

Technical officers
26. Marko Brakus, B. Sc.
27. Dr. Igor Mozetič

Technical and administrative staff
28. Živa Antauer, B.Sc., left 01.07.11
29. Tina Anžič, B. Sc.
30. Milica Bauer, B. Sc.
31. Dr. France Dacar
32. Dr. Damjan Demšar, left 01.04.11

Note:
* part-time JSI member

BIBLIOGRAPHY

Original Articles

1. Nataša Atanasova, Sašo Džeroski, Boris Kompare, Ljupčo Todorovski, Gideon Gal, "Automated discovery of a model for dinoflagellate dynamics", *Environ. model. softw.*, vol. 26, no. 5, pp. 658-668, 2011.
2. Jérome Cortet, Dragi Kocev, Caroline Ducobu, Sašo Džeroski, Marko Debeljak, Christophe Schwartz, "Using data mining to predict soil quality after application of biosolids in agriculture", *J. environ. qual.*, vol. 40, no. 6, pp. 1972-1982, 2011.
3. Ivica Dimitrovski, Dragi Kocev, Suzana Loskovska, Sašo Džeroski, "Hierarchical annotation of medical images", *Pattern recogn.*, vol. 44, no. 10/11, pp. 2436-2449, 2011.
4. Elena Ikonomovska, João Gama, Sašo Džeroski, "Learning model trees from evolving data streams", *Data mining and knowledge discovery*, vol. 23, no. 1, pp. 128-168, 2011.
5. Reuben P. Keller, Dragi Kocev, Sašo Džeroski, "Trait-based risk assessment for invasive species: high performance across diverse taxonomic groups, geographic ranges and machine learning/statistical tools", *Divers. distrib.*, vol. 17, no. 3, pp. 451-461, 2011.
6. Lado Kutnar, Andrej Kobler, Sašo Džeroski, "Kakšni bi lahko bili učinki segrevanja ozračja na bukove gozdove v prihodnosti", *Les (Ljubl.)*, let. 63, no. 5, pp. 203-207, 2011.
7. Nada Lavrač, Anže Vavpetič, Larisa N. Soldatova, Igor Trajkovski, Petra Kralj Novak, "Using ontologies in semantic data mining with SEGS and g-SEGS", In: Discovery science: 14th International Conference, Ds 2011, Espoo, Finland: proceedings, *Lect. notes comput. sci.,* vol. 6926, pp. 165-178, 2011.
8. Nikola Ljubešić, Tomaž Erjavec, "hrWac and siWac: compiling web corpora for Croatian and Slovene", In: Text, speech and dialogue: proceedings, *Lect. notes comput. sci.,* vol. 6836, pp. 395-402, 2011.
9. Marta Macedoni-Lukšič, Ingrid Petrič, Bojan Cestnik, Tanja Urbančič, "Developing a deeper understanding of autism: connecting knowledge through literature mining", *autism res. treat.*, vol. 2011, 8 pp., 2011.
10. Martin Pavlovič *et al.* (6 authors), "Development of DEX-HOP multi-attribute decision model for preliminary hop hybrids assessment", *Comput. electron. agric.*, vol. 75, no. 1, pp. 181-189, 2011.
11. Vid Podpečan, Nada Lavrač, Igor Mozetič, Petra Kralj Novak, Igor Trajkovski, Laura Langohr, Kimmo Kulovesi, Hannu Toivonen, Marko Petek, Helena Motaln, Kristina Gruden, "SegMine workflows for semantic microarray data analysis in Orange4WS", *BMC bioinformatics*, vol. 12, no. 416, pp. 416-1-416-16, 2011.
12. Senja Pollak, Roel Coesemans, Walter Daelemans, Nada Lavrač, "Detecting contrast patterns in newspaper articles by combining discourse analysis and text mining", *Pragmatics (Wilrijk)*, vol. 21, no. 4, pp. 647-683, 2011.
13. Mitja Pugelj, Sašo Džeroski, "Predicting structured outputs k-Nearest Neighbours Method", In: Discovery science: 14th International Conference, Ds 2011, Espoo, Finland: proceedings, *Lect. notes comput. sci.,* vol. 6926, pp. 262-276, 2011.
14. Daniela Stojanova, Michelangelo Ceci, Annalisa Appice, Sašo Džeroski, "Network regression with predictive clustering trees", In: Machine learning and knowledge discovery in databases. Part III: European Conference, ECML PKDD 2011, Athens, Greece, *Lect. notes comput. sci.,* vol. 6913, pp. 333-348, 2011.
15. Daniela Stojanova, Michelangelo Ceci, Annalisa Appice, Donato Malerba, Sašo Džeroski, "Global and local spatial autocorrelation in predictive clustering trees", In: Discovery science: 14th International Conference, Ds 2011, Espoo, Finland: proceedings, *Lect. notes comput. sci.,* vol. 6926, pp. 307-322, 2011.
16. Katerina Taškova, Peter Korošec, Jurij Šilc, Ljupčo Todorovski, Sašo Džeroski, "Parameter estimation with bio-inspired meta-heuristic optimization: modeling the dynamics of endocytosis", *BMC systems biology*, vol. 5, pp. 159-1-159-52, 2011.
17. Aneta Trajanov, "Analysis of results of ecological simulation models with machine learning", *Informatica (Ljubl.)*, vol. 35, no. 2, pp. 285-286, 2011.
18. Monika Žáková, Petr Křemen, Filip Železný, Nada Lavrač, "Automating knowledge discovery workflow composition through ontology-based planning", *IEEE trans. autom. sci. eng.*, vol. 8, no. 2, pp. 253-264, 2011.

Review Articles and Chapters in Books

1. Marko Debeljak, Sašo Džeroski, "Decision trees in ecological modelling", In: *Modeling complex ecological dynamics: an introduction into ecological modelling for students, teachers & scientists*, Fred Jopp, ed., Berlin, Heidelberg [etc.], Springer, cop. 2011, pp. 197-209.
2. Tomaž Erjavec, "Slovenska prevodna književnost 1848-1918: digitalna knjižnica in korpus AHLIB", In: *Meddisciplinarnost v slovenistiki*, (Obdobja, Simpozij, = Symposium, 30), Simona Kranjc, ed., 1. natis, Ljubljana, Znanstvena založba Filozofske fakultete, 2011, pp. 33-40.
3. Tomaž Erjavec, Jan Jona Javoršek, Matija Ogrin, Petra Vide Ogrin, "Od biografskega leksikona do znanstvenokritične izdaje: vprašanje trajnosti elektronskih besedil", *Knjižnica (Tisk. izd.)*, vol. 55, no. 1, pp. 103-114, 2011.
4. Tomaž Erjavec, Ines Jerele, Maša Kodrič, "Izdelava korpusa starejših slovenskih besedil v okviru projekta IMPACT", In: *Meddisciplinarnost v slovenistiki*, (Obdobja, Simpozij, = Symposium, 30), Simona Kranjc, ed., 1. natis, Ljubljana, Znanstvena založba Filozofske fakultete, 2011, pp. 41-47.
5. Petra Kralj Novak, Nada Lavrač, Geoffrey I. Webb, "Supervised descriptive rule induction", In: *Encyclopedia of machine learning: with 293 figures and 78 tables*, Claude Sammut, ed., Geoffrey I. Webb, ed., New York, Springer, 2011, pp. 938-941.

6. Matija Ogrin, Jan Jona Javoršek, Tomaž Erjavec, "Rokopisi slovenskega slovstva baročne in razsvetljenske dobe: predmet več ved in metod", In: *Meddisciplinarnost v slovenistiki*, (Obdobja, Simpozij, = Symposium, 30), Simona Kranjc, ed., 1. natis, Ljubljana, Znanstvena založba Filozofske fakultete, 2011, pp. 337-341.
7. Ingrid Petrič, Tanja Urbančič, Bojan Cestnik, "Ontological representation of virtual business communities: how to find right business partners", In: *Innovations in smes and conducting e-business: technologies, trends and solutions*, (Premier reference source), Maria Manuela Cruz-Cunha, ed., Joao Varajao, ed., Hershey, New York, London, Information Science Reference, Igi Global, 2011, pp. 263-277.

Published Conference Papers

Invited Papers

1. Marko Debeljak, "Gozdovi v svetu in Sloveniji", In: *Povezanost procesov: zbornik prispevkov: proceedings*, Mednarodni posvet Biološka znanost in družba, Ljubljana, 6. in 7. oktober 2011 = Conference on Bioscience and Society, October 6-7, 2010, Ljubljana, Slovenia, Minka Vičar, ed., Saša Kregar, ed., Frances M. Ashcroft, 1. izd., Ljubljana, Zavod RS za šolstvo, 2011, pp. 86-94.
2. Sašo Džeroski, "Inductive databases and constraint-based data mining", In: *Formal concept analysis: proceedings*, (Lecture note in computer science, Lecture notes in artificial intelligence, vol. 6628), 9th International Conference, ICFCA 2011, May 2-6, 2011, Nicosia, Cyprus, Petko Valtchev, ed., Robert Jäschke, ed., Heidelberg [etc.], Springer, cop. 2011, pp. 1-17.

Regular papers

1. Maja Bračič Lotrič, Bojan Cestnik, "Literature mining of blood flow regulatory mechanisms for finding cross context relations", In: *Computer systems and technology: 12th International Conference, CompSysTech'11, June 16-17, 2011, Vienna, Austria: proceedings*, (ACM international conference proceedings series, vol. 578), Boris Rachev, ed., Angel Smrikarov, ed., New York, ACM, = Association for Computing Machinery, 2011, pp. 353-358.
2. Bojan Cestnik, Tanja Urbančič, Ingrid Petrič, "Ontological representations for supporting learning in business communities", In: *e-Learning'11: proceedigs of the International Conference on e-Learning and the Knowledge Society, 25-26 August 2011, Bucharest, Romania*, Angel Smrikarov, ed., Bucharest, ASE Publishing House, 2011, pp. 260-265.
3. Christian Chiarcos, Tomaž Erjavec, "OWL/DL formalization of the MULTEXT-East morphosyntactic specifications", In: *LAW V*, 5th Linguistic Annotation Workshop, June 23, 2011, Portland, Oregon, Portland, Association for Computational Linguistics, 2011, pp. 11-20.
4. Marko Debeljak, Bernard Ženko, Aleš Poljanec, "Napovedni modeli razvoja gozdov v Sloveniji", In: *Priprava gozdnogospodarskih in lovsko upravljalskih načrtov območij za obdobje 2011-2020: zbornik prispevkov*, (Gospodarjenje z gozdovi in načrtovanje, 5), Andrej Bončina, ed., Ljubljana, Oddelek za gozdarstvo in obnovljive gozdne vire - Biotehniška fakulteta, 2011, pp. 11-14.
5. Marko Debeljak, Bernard Ženko, Aleš Poljanec, "Predictive models of forest development in Slovenia", In: *Innovations in sharing environmental observations and informations: proceedings of the 25th EnviroInfo International Conference, October 5-7, 2011, Ispra, Italy*, Werner Pillmann, ed., S. Schade, ed., Paul Smits, ed., Aachen, Shaker Verlag, cop. 2011, pp. 97-100.
6. Tomaž Erjavec, "Automatic linguistic annotation of historical language: ToTrTaLe and XIX century Slovene", In: *LaTeCH 2011*, 5th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities [and] 49th Annual Meeting of the Association for Computational Linguistics: human language technologies (ACL/HLT), June 19-24, 2011, Portland, Oregon, USA, Kalliopi Zervanou, ed., Piroska Lendvai, ed., Portland, Association for Computational Linguistics, 2011, pp. 33-38.
7. Tomaž Erjavec, Christoph Ringlstetter, Maja Žorga Dulmin, Annette Gotscharek, "A lexicon for processing archaic language: the case of XIXth century Slovene", In: *An ESSLLI 2011 Workshop*, First International Workshop on Lexical Resources, An ESSLLI 2011 Workshop, Ljubljana, Slovenia - August 1-5, 2011, [S. l., s. n., 2011], pp. 24-30.
8. Valentin Gjorgjioski, Dragi Kocev, Sašo Džeroski, "Comparison of distances for multi-label classification with PCTs", In: *Zbornik 14. mednarodne multikonference Informacijska družba - IS 2011, 10.-14. oktober 2011: zvezek A: volume A*, (Informacijska družba), Marko Bohanec, ed., Matjaž Gams, ed., Dunja Mladenić, ed., Marko Grobelnik, ed., Marjan Heričko, ed., Urban Kordeš, ed., Olga Markič, ed., Jadran Lenarčič, ed., Leon Žlajpah, ed., Andrej Gams, ed., Vladimir Fomichov, ed., Olga S. Fomichova, ed., Andrej Brodnik, ed., Rok Sosič, ed., Vladislav Rajkovič, ed., Tanja Urbančič, ed., Mojca Bernik, ed., Ljubljana, Institut Jožef Stefan, 2011, pp. 125-128.
9. Elena Ikonomovska, "Regression on evolving multi-relational data streams", In: *Electronic conference proceedings*, Uppsala, ACM, 2011, 7 pp.
10. Elena Ikonomovska, João Gama, Sašo Džeroski, "Incremental multi-target model trees for data streams", In: *Proceedings of the 26th Annual ACM Symposium on Applied Computing 2011, Taichung, Taiwan, March 21 - 24, 2011*, [S. l.], ACM, cop. 2011, pp. 988-993.
11. Elena Ikonomovska, João Gama, Bernard Ženko, Sašo Džeroski, "Speeding up Hoeffding-based regression trees with options: a", In: *Proceedings of the Twenty-eighth International Conference on Machine Learning: [Bellevue, Washington, USA, June 28 - July 2, 2011]*, Lise Getoor, ed., Tobias Scheffer, ed., [Madison, International Machine Learning Society], 2011, pp. 537-552.
12. Stojan Košti, Bojan Cestnik, "Using business analysis to select appropriate e-learning system", In: *e-Learning'11: proceedigs of the International Conference on e-Learning and the Knowledge Society, 25-26 August 2011, Bucharest, Romania*, Angel Smrikarov, ed., Bucharest, ASE Publishing House, 2011, pp. 165-176.
13. Biljana Mileva-Boshkoska, Marko Bohanec, "IPSSC: Copula regression based ranking of non-linear decision options", In: *Zbornik prispevkov*, 3. študentska konferenca Mednarodne podiplomske šole Jožefa Stefana = 3rd Jožef Stefan International Postgraduate School Students Conference, 25. maj 2011, Ljubljana, Slovenija, Dejan Petelin, ed., Aleš Tavčar, ed., Brigita Rožič, ed., Bogdan Pogorelc, ed., Ljubljana, Mednarodna podiplomska šola Jožefa Stefana, 2011, pp. 91-97.
14. Borut Sluban, Dragan Gamberger, Nada Lavrač, "Performance analysis of class noise detection algorithms", In: *Stairs 2010: proceedings of the Fifth Starting AI Researchers' Symposium in conjunction with the 19th European Conference on Artificial Intelligence (ECAI 2010), 16-20 August 2010, Lisbon, Portugal*, (Frontiers in artificial intelligence and applications, vol. 222), Thomas Ågotnes, ed., Amsterdam ... [et al.], IOS Press, pp. 303-314.
15. Borut Sluban, Matjaž Juršič, Bojan Cestnik, Nada Lavrač, "Evaluating outliers for cross-context link discovery", In: *Artificial intelligence in medicine: proceedings*, (Lecture notes in computer science, Lecture notes in artificial intelligence, 6747), 13th Conference on Artificial Intelligence in Medicine, AIME 2011, Bled, Slovenia, July 2-6, 2011, Mor Peleg, ed., Nada Lavrač, ed., Carlo Combi, ed., Berlin [etc.], Springer, cop. 2011, pp. 343-347.
16. Jasmina Smailović, Martin Žnidaršič, Miha Grčar, "Web-based experimental platform for sentiment analysis", In: *The proceedings of the 3nd International Conference on Information Society and Information Technologies - ISIT 2011, [Dolenjske Toplice, 9-11 November 2011]*, Matej Mertik, ed., Novo mesto, Faculty of Information Studies, 2011, 6 pp.
17. Anže Vavpetič, Igor Trajkovski, Petra Kralj Novak, Nada Lavrač, "Semantic data mining system g-SEGS", In: *PlanSoKD'11, Planning to Learn and Service-Oriented Knowledge Discovery September 9, 2011, Athens, Greece*, Jörg-Uwe Kietz, ed., Simon Fischer, ed., Nada Lavrač, ed., Vid Podpečan, ed., [S. l., s. n.], 2011, pp. 17-29.

Research Monograph

1. Aneta Trajanov, *Machine learning in agroecology: from simulation models to co-existence rules*, Saarbrücken, LAP Lambert, cop. 2011.

Theses

Ph. D. Thesis

1. Dragi Kocev, *Ensembles for predicting structured outputs:* doctoral dissertation, Ljubljana, [D. Kocev], 2011.

B. Sc. Theses

1. Nejc Trdin, *Decision support model for management of water sources:* undergraduate thesis, Ljubljana, [N. Trdin], 2011.
2. Anže Vavpetič, *The use of ontologies as background knowledge in data mining:* undergraduate thesis, Ljubljana, [A. Vavpetič], 2011.