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 an information- and knowledge-based society. Established areas of our work include intelligent data analysis (machine learning, data mining, and knowledge discovery in databases), language technologies and computational linguistics, computational creativity, decision support and knowledge management. In addition to research in knowledge technologies, we are also developing applications in environmental sciences and management, agronomy, medicine, biomedicine and bioinformatics, economics, finance and marketing. The department is also a recognised centre of linguistic research and digital humanities.

Head:
Prof. Nada Lavrač

In 2017 we were involved in fourteen national projects, two EU FP7 and ten Horizon 2020 projects, one COST action, one INTERREG V-A Slovenia-Italy project, one infrastructure project, one smart-specialization project, two bilateral projects, and five industry projects. The department hosted eight junior researchers working towards their PhDs.

In the area of **intelligent data analysis and data mining** we have developed several new methods and used them in a number of application domains. In the area of heterogeneous network analysis we developed new search heuristics based on text mining. In the area of semantic data mining we achieved a one-hundred-times efficiency improvement by network-ranking-assisted data preprocessing and filtering. In redescription mining we developed a new method for redescription construction using predictive clustering trees as well as a new framework and method for redescription set construction. We published methods that enable big-data analysis within the ClowdFlows platform, which is continually developed and whose commercialization potential is being funded within the FET Innovation Launchpad project CF-Web. Within the PD_manager project, which we coordinate at JSI, we developed a method that combines multi-target learning with short-time-series analysis for the identification of groups of similar patients with Parkinson's disease. We started to work on a new H2020 project SAAM (Supporting Active Ageing through Multimodal coaching), where we are developing a Virtual Assistant-Coach that supports the aging population living at home.

Nada Lavrač had an invited lecture "From Relational to Semantic Data Mining" at the 16th International Semantic Web Conference ISWC-2017 in Vienna. The conference is the main world-wide Semantic Web event, with over 600 participants.

In the context of developing methods for the automated modelling of dynamic systems, using both data and domain knowledge, we have proposed a meta-model framework for surrogate-based numerical optimization. The use of surrogate models allows for more efficient parameter optimization in the differential equations used to model dynamic systems.

We successfully completed the coordination of the MAESTRA project (Learning from Massive, Incompletely Annotated, and Structured Data), which received the highest evaluations by the EU's reviewers. The project focused on analysing data that can be complex in several ways, including the semi-supervised setting with partially annotated data. Within the project, we developed many new methods for predicting different types of structured outputs in both a batch and a streaming setting. These include batch methods for semi-supervised classification and multi-target regression, online tree-based methods for multi-target regression, option trees for hierarchical multi-label classification, as well as trees for hierarchical multi-target regression and tree ensembles for multi-label classification using random label subset selections. We have also developed methods for the feature ranking for different types of structured outputs, such as multi-target regression and hierarchical multi-label classification.

We have used the developed methods for image analysis and modelling forest growing stock data. In the area of medicine, we have used them for modelling time-series of glucose measurements from diabetes patients and analysing the influence of pathogens on the composition of human gut microbiota. Finally, we have used them in the ESA Mars Express Power Challenge, where the task was to predict the power consumption of 33 different lines of the thermal subsystem of the Mars Express orbiter.

In the context of the FET Flagship HBP (Human Brain Project) project, we are developing new data-mining methods and applying them to discover the biological signatures of neurodegenerative diseases, such as Alzheimer's. Two novel methods for mining redescription sets were developed based on predictive clustering, and applied to relate the clinical and biological characteristics of cognitively impaired and Alzheimer's disease patients.

In the LANDMARK project we continued with the development of models for the prediction of soil functions. Most of the research activities were focused on data pre-processing, which was a very complex task. Models for the prediction of primary productivity from long-term experimental studies in Austria were developed. The results of this study were submitted to a special issue of the Regional Environmental Change Journal. In addition, more prediction models were developed for other soil functions as well and are in the process of validation with the domain experts. An important part of our research activities in the LANDMARK project was the development of multiple-attribute decision models using the DEX methodology for an assessment of five soil functions. Each of the models was verified by experts and calibrated to three prevailing pedoclimatic European zones. Data from various national and EU databases were extracted and pre-processed. This required a lot of coordination efforts in order to collect the data and to perform the discretisation of their values, which was done after a long process of determining threshold values.

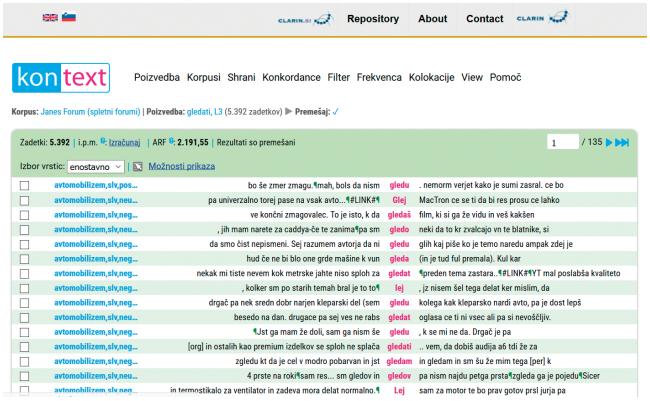

Figure 1: Concordance of verb "gledati" ("to look") in the Janes Forum corpus. The KonText concordancer of the CLARIN.SI research infrastructure offers over 40 corpora, among them several new corpora of user-generated Slovene. Note that the query also returns results with non-standard forms of the verb, such as "gledo".

A large part of the activities in the BioDiv project were connected to data pre-processing. The completed database consists of three large sets of attributes describing: i) taxonomic and functional traits of Syrphide auxiliary species, ii) landscape properties, describing the fields and semi-natural habitats, and iii) environmental conditions during the sampling periods. The abundance of the prevailing Syrphide species and selected functional traits were modelled using model and regression trees.

Two H2020 projects started in 2017: TRUE (TRansition paths to sUstainable legume based systems in Europe), and TomRes (A novel and integrated approach to increase multiple and combined stress tolerance in plants using tomato as a model). Since the majority of our activities are planned for 2018 and 2019, we were mainly focused on initial research activities such as an intensive literature search. Within TRUE we developed the first drafts of the conceptual model for the assessment of the sustainability of legume quality chains. Our activities in the TomRes project were focused on a study of the research domain and on the structuring of the existing knowledge about the prediction models of tomato responses to nutrient and water stress on the eco-physiology level.

In the area of **text and web mining, and heterogeneous information network analysis** we successfully completed the SIMPOL project (Financial Systems Simulation and Policy Modelling). We continued work on an existing H2020 FET project DOLFINS (Distributed Global Financial Systems for Society). We also continued to work on the new national project FORMICA (Influence of formal and informal corporate communications on capital markets). The main emphasis of the research in this area is combining text mining, network analysis and sentiment analysis to reveal and highlight the underlying properties in different domains. The main sources of data that we analyse are social media (Twitter, Facebook), online news, annual reports, and other relevant data (e.g., voting on the Brexit referendum, stock prices, register of lobby organizations, etc.).

We proposed a methodology to properly evaluate Twitter sentiment (or stance) classification models for Twitter-specific time-ordered data. The classification models are language and topic dependent. They are constructed using text-mining methods, from a large set of manually annotated tweets by domain experts, and result in high-quality domain- and language-specific models. We showed that the standard cross-validation approach is appropriate to evaluate such models only if one applies blocked cross-validation, and does not randomly shuffle the examples. We applied the above methodology to a highly relevant European use case: the UK referendum on Brexit. We developed a Brexit-specific Twitter stance classification model and used it to monitor public mood before and during the referendum. We also identified the most influential Twitter users and noted that the proponents of Brexit were considerably more active than the remain camp. We also provided an in-depth analysis of the Twitter volume and stance about the 30 companies in the Dow Jones Industrial Average index (DJIA), over a period of three years. We focused on Earnings Announcements and showed that there is a considerable difference with respect to when the announcements are made: before the market opens or after the market closes. We analysed the differences in terms of the Twitter volumes, cumulative abnormal returns, trade returns, and earnings surprises. We also collected a dataset of annual reports from the DJIA and studied the correlations between the linguistic and financial indicators.

Martin Breskvar, Matej Petkovič and Blaž Škrlj were part of the winning team at the HackElect 2017 hackathon. Using methods such as predictive classification trees and deep neural networks, they were the only ones who successfully predicted the consumption of electricity at the level of individual household appliances for different time scales.

In the area of **decision support** our long-term goal is to develop methods and techniques of decision modelling, support them with software and integrate them with data-mining systems. In 2017, we published a methodological

paper in which – in addition to a formal description of our most commonly used decision-modelling method DEX – we proposed a number of methodological extensions of the approach in terms of using numeric attributes, value distributions and a relational evaluation of alternatives. We also published the results of applying these approaches in the assessment of the long-term sustainability of electric energy production in Slovenia. We continued working on two EU H2020 projects aimed at the development of computer platforms and decision-support systems for the management of severe chronic diseases: the Parkinson's disease (project PD_manager) and congestive heart failure (HeartMan). We also investigated the performance of three methods for the approximation of utility functions in the DEX method, analysed statistical properties of DEX models and carried out preliminary studies for introducing feedback loops (cycles) in DEX models.

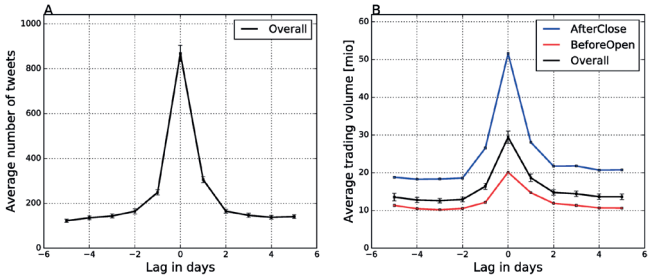

Figure 2: Daily number of tweets (A) and trading volume (B) around the Earnings Announcements. The overall average number of tweets per trading day is 200. The trading volume (B) shows the overall average across all EAs (black line), the average trading for the AfterClose (blue line), and for the BeforeOpen (red line) announcements. Error bars around the black lines denote one standard error.

In the area of **language technologies and digital humanities** we work on producing language resources and methods to annotate text with linguistic information, with a focus on Slovene. In 2017 we helped in producing the digital text-critical edition of the 700-page Poljane Manuscript from 1800, developing a universal dependencies treebank for Serbian, organising the evaluation campaign of VarDial, the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects, and in expanding the Slovene Gigafida reference corpus with internet content. We described the processes and tools necessary for annotation creation, the multilingual MULTEXT-East resources, and the Slovene reference morphological lexicon Sloleks and made a roadmap for its development.

In 2017 we were concluding our work in the scope of the national research project JANES "Resources, Tools and Methods for Research of Non-standard Internet Slovene", where we contributed to an overview of best practices from projects on computer-mediated communication, analysed non-standard language as found in tweets, and developed a system for part-of-speech tagging of non-standard Slovene, Croatian and Serbian. We developed several approaches for gender prediction on social media, including a language-independent system using non-textual information, gender and language variety prediction from text (2nd place in shared task PAN), and a focused study on gender prediction for Slovene. We produced two manually annotated training corpora for non-standard Slovene, the syntactically annotated Janes-Syn and the second version of the Janes-Tag dataset, which adds named entity annotations to the existing normalised words, part-of-speech tags and lemmas. We also produced datasets for Croatian and Serbian non-standard language, which cover the same levels of annotation as Janes-Tag. We compiled manually annotated datasets for linguistic investigations of non-standard language, in particular Janes-Vejica, which annotates (in)correctly placed commas, Janes-Kratko, which annotates shortened words, and Janes-Preklop, which is annotated for code-switching. Finally, we produced the main deliverable of the project, i.e., version 1.0 of the JANES corpus, which contains 250 million tokens or 13 million texts of user-generated internet Slovene, with rich metadata and automatically standardised, part-of-speech tagged and lemmatised words. The complete corpus is available on-line for searching through the CLARIN.SI concordancers, while its constituent parts are also available for download from the CLARIN.SI repository, in particular: Janes-Tweet, Janes-Blog, Janes-Forum, Janes-News, and Janes-Wiki.

Nikola Ljubešić and Yves Scherrer (University of Geneva) won the CLIN2017 shared task on normalising historical text with their CSMtiser tool, which had been developed within the national basic research project JANES and the national research infrastructure CLARIN.SI.

We continued work on our national research project KAS "Slovene scientific texts: resources and description", where we analysed the lexis of the previously produced KAS corpus, containing almost one billion words of scientific texts gathered from the digital libraries of Slovene universities. We started work on a new national research project FRENK "Resources, methods, and tools for the understanding, identification, and classification of various forms of socially unacceptable discourse in the information society", where we developed an annotation schema for socially unacceptable online discourse practices.

In the scope of the industrial project TermIolar for the Slovene language service provider Iolar, we developed a system for semi-automated

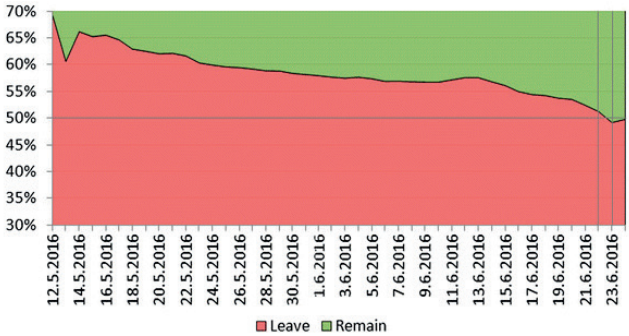

Figure 3: Cumulative stance of Twitter users regarding Brexit. Red color represents the Leave users, and green color represents the Remain users (Neutral users are not included). Remain users were gradually joining as opposed to the Leave users that were present and dominating in the number of tweets most of the time. It was only 2 days before the referendum that the number of Remain users exceeded the number of Leave users.

terminology management. We studied machine-learning approaches to good example extraction and ontology-based translation memory maintenance.

We lead the Slovene research infrastructure CLARIN.SI, which provides easy publication and sustainable access to digital language data for scholars in the humanities and social sciences. In 2017 the number of deposited resources in the CLARIN.SI repository almost doubled, reaching 79 at the end of the year. Here, we participated in the creation of the second version of the Gos VideoLectures spoken corpus, the manually annotated training corpus ssj500k, and the Slovenian parliamentary corpus SlovParl. We installed two concordancers under CLARIN.SI, i.e., KonText and noSketch Engine, and made available over 40 corpora for on-line searching and analysis through these concordancers. We also continued work on integrating tools for the annotation of Slovene into the on-line workflow design and execution platform WebLicht. CLARIN.SI supported the organisation of "TransTech17: 3rd Summer School in Translation Technologies" (September 4–8, 2017) and "ReLDI Seminar: Empirical data in linguistics: From research design to data analysis" (June 21-23, 2017).

We collaborated in the work of the Slovene Institute for Standardization as the Slovene representatives in ISO/TC37/SC4 (Terminology and Other Language and Content Resources / Language Resources Management) by reviewing, translating and approving Slovene standards from this field. We also continue to serve as technical editors for the on-line Slovene Biographical Lexicon.

Some outstanding publications in the past year

1. Pavle Boškoski, Andrej Debenjak, Biljana Mileva-Boshkoska. Rayleigh copula for describing impedance data - with application to condition monitoring of proton exchange membrane fuel cells. *European journal of operational research*, ISSN 0377-2217. [in press] 2017, 21 pp., doi: 10.1016/j.ejor.2017.08.058.
2. Jurica Levatić, Michelangelo Ceci, Dragi Kocev, Sašo Džeroski. Self-training for multi-target regression with tree ensembles. *Knowledge-based systems*, ISSN 0950-7051. 2017, vol. 123, pp. 41-60, doi: 10.1016/j.knosys.2017.02.014.
3. Manja Klemenčič, Martin Žnidaršič, Anže Vavpetič, Matej Martinc. Erasmus students' involvement in quality enhancement of Erasmus+ mobility through digital ethnography and ErasmusShouts. *Studies in higher education*, ISSN 0307-5079, 2017, vol. 42, no. 5, pp. 925-932, doi: 10.1080/03075079.2017.1293879.
4. Tomaž Erjavec. MULTEXT-East. In: Nancy M. Ide, James Pustejovsky (eds.). Handbook of linguistic annotation. Amsterdam: Springer. 2017, pp. 441-462.
5. Grčar, Miha, Čerepnalkoski, Darko, Mozetič, Igor, Kralj Novak, Petra. Stance and influence of Twitter users regarding the Brexit referendum. *Computational social networks*, ISSN 2197-4314, 2017, vol. 4, pp. 6-1-6-25, doi: 10.1186/s40649-017-0042-6.
6. Peter Gabrovšek, Darko Aleksovski, Igor Mozetič, Miha Grčar. Twitter sentiment around the earnings announcement events. *PloS one*, ISSN 1932-6203, 2017, vol. 12, no. 2, pp. e0173151-1-e0173151-21, doi: 10.1371/journal.pone.0173151.
7. Marko Bohanec, Nejc Trdin, Branko Kontić. A qualitative multi-criteria modelling approach to the assessment of electric energy production technologies in Slovenia. *Central European Journal of Operations Research*, ISSN 1435-246X, 2017, vol. 25, no. 3, pp. 611-625, doi: 10.1007/s10100-016-0457-4.
8. Matej Mihelčić, Marko Bohanec. Approximating incompletely defined utility functions of qualitative multi-criteria modeling method DEX. *Central European Journal of Operations Research*, ISSN 1435-246X, 2017, vol. 25, no. 3, pp. 627-649, doi: 10.1007/s10100-016-0451-x.
9. Janez Kranjc, Roman Orač, Vid Podpečan, Nada Lavrač, Marko Robnik Šikonja. ClowdFlows : online workflows for distributed big data mining. *FGCS*, ISSN 0167-739X. 2017, vol. 68, pp. 38-58, doi: 10.1016/j.future.2016.07.018.
10. Matej Mihelčić, Sašo Džeroski, Nada Lavrač, Tomislav Šmuc. A framework for redescription set construction. *Expert systems with applications*, ISSN 0957-4174. 2017, vol. 68, pp. 196-215, doi: 10.1016/j.eswa.2016.10.012.
11. Sabina Horvat, Aleksander Mahnič, Martin Breskvar, Sašo Džeroski, Maja Rupnik. Evaluating the effect of Clostridium difficile conditioned medium on fecal microbiota community structure. *Scientific reports*, ISSN 2045-2322, 2017, vol. 7, 11 pp., doi: 10.1038/s41598-017-15434-1.
12. Dragan Gamberger, Nada Lavrač, et al. Identification of clusters of rapid and slow decliners among subjects at risk for Alzheimer's disease. *Scientific reports*, ISSN 2045-2322, 2017, vol. 7, pp. 1-12, doi: 10.1038/s41598-017-06624-y.

Awards and appointments

1. Marko Bohanec was awarded the "*Fellow and Distinguished Scholar*" by the *International Institute for Applied Knowledge Management*.
2. Martin Breskvar, Matej Petkovič and Blaž Škrlj were in the winning team of the HackElect 2017 hackathon. With methods such as predictive classification trees and deep neural networks they, as the only team, managed to predict electricity consumption for each household appliance in various time frames.
3. Nikola Ljubešić and Yves Scherrer (University of Geneva) won the *CLIN2017 Shared Task on Normalising Historical Text* with their CSMtiser tool, which was developed within the national basic research project JANES and the national research infrastructure CLARIN.SI.
4. The approach developed by Matej Martinc, Iza Škrjanec, Katja Zupan, and Senja Pollak for author profiling has been ranked second among 22 groups in the PAN2017 competition, which this year focused on *Gender and Language Variety Identification on Twitter*.

Organization of conferences, congresses and meetings

1. The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery – ECML PKDD 2017, Skopje, Macedonia, 18.–22. 9. 2017
2. Final meeting of EU project MAESTRA, Ljubljana, 3.–7. 4. 2017
3. Organization of project "Medical Informatics Platform (SP8)", part of EU project HBP, Ljubljana, 25.–27. 1. 2018
4. "The Tenth International Ljubljana-Zagreb Workshop on Knowledge Technologies", Fiesa, Slovenia, 15.–16. 6. 2017

INTERNATIONAL PROJECTS

1. BIODIV - Understanding and Managing Biodiversity in Agricultural Ecosystems by Data Mining and Decision Support; Structured Output Prediction with Applications in Sustainable Agricultural Production
   Prof. Sašo Džeroski
   Arvalis - Institut Du Végétal
2. 7FP - SIMPOL; Financial Systems Simulation and Policy Modelling
   Prof. Igor Mozetič
   European Commission
3. 7FP - MAESTRA; Learning from Massive, Incompletely Annotated, and Structured Data
   Prof. Sašo Džeroski
   European Commission
4. PARSEME: PARSing and Multi-Word Expressions. Towards Linguistic Precision and Computational Efficiency in Natural Language Processing.
   Prof. Tomaž Erjavec
   Cost Office
5. H2020 - DOLFINS; Distributed Global FInancial Systems for Society
   Prof. Igor Mozetič
   European Commission
6. H2020 - HBP SGA1; Human Brain Project Specific Grant Agreement 1 within HBP FPA; Human Brain Framework Partnership Agreement
   Prof. Sašo Džeroski
   European Commission
7. H2020 - TRUE; Transition Paths to Sustainable Legume based Systems in Europe
   Prof. Marko Debeljak
   European Commission
8. H2020 - TomRes; A Novel and Integrated Approach to increase Multiple and Combined Stress Tolerance in Plants Using Tomato as a Model
   Prof. Marko Debeljak
   European Commission
9. H2020 - LANDMARK; LAND Management: Assessment, Research, Knowledge Base
   Prof. Marko Debeljak
   European Commission
10. H2020 - NARSIS; New Approach to Reactor Safety Improvements
    Prof. Marko Bohanec
    European Commission
11. H2020 - SAAM; Supporting Active Ageing through Multimodal Coaching
    Asst. Prof. Bernard Ženko
    European Commission
12. H2020 - PD_manager; Mhealth Platform for Parkinson's Disease Management
    Dr. Dragana Miljković
    European Commission
13. H2020 - HeartMan; Personal Decison Support System for Heart Failure Management
    Prof. Marko Bohanec
    European Commission
14. H2020 - CF-Web; ClowdFlows Data and Text Analytics Marketplace on the Web
    Asst. Prof. Martin Žnidaršič
    European Commission
15. Mining Complex Data in Environmental and Life Sciences
    Prof. Sašo Džeroski
    Slovenian Research Agency
16. Mining Heterogeneous Linked Biomedical Data
    Prof. Nada Lavrač
    Slovenian Research Agency
17. Semantic Role Labeling in Slovene and Croatian
    Prof. Tomaž Erjavec
    Slovenian Research Agency

RESEARCH PROGRAM

1. Knowledge Technologies
   Prof. Nada Lavrač

R & D GRANTS AND CONTRACTS

1. Integrative research of sexual dimorphism evolution
   Prof. Sašo Džeroski
2. Resources, Tools and Methods for the Research of Nonstandard Internet Slovene
   Prof. Tomaž Erjavec
3. Molecular bases of interactions among the grapevine and phytoplasmal causing agents of the grapevine yellows diseases
   Prof. Nada Lavrač
4. Development of a multimethod approach to study wildlife behavior: investigating humanbear conflicts in contrasting landscapes of Europe
   Prof. Sašo Džeroski
5. Influence of formal and informal corporate communications on capital markets
   Dr. Senja Pollak
6. Collocation as a basis for language description: semantic and temporal perspectives
   Dr. Nikola Ljubešić
7. Slovene scientific texts: resources and description
   Prof. Tomaž Erjavec
8. Analysis of heterogeneous information networks for knowledge discovery in life-sciences
   Prof. Nada Lavrač
9. Machine Learning for Systems Sciences
   Prof. Sašo Džeroski
10. Resources, methods and tools for the understanding, identification and classification of various forms of socially unacceptable discourse in the information society
    Prof. Tomaž Erjavec
11. Fellowship to visit ERC Grantee
    Dr. Vedrana Vidulin

12. Forbidden Books in the Slovenian Lands in the Early Modern Period
    Prof. Tomaž Erjavec
13. Structured output prediction with applications in sustainable agricultural production
    Prof. Sašo Džeroski
14. Food for future - F4F
    Asst. Prof. Bernard Ženko
    Ministry of Education, Science and Sport
15. TRAIN: Big Data and Disease Models: A Cross-border Platform for Validated Biotech Industry Kits
    Prof. Sašo Džeroski
    Regione Autonoma Friuli Venezia Giulia
16. Data Mining and Decision support in Sustainable Food Production
    Dr. Vladimir Kuzmanovski
    Ministry of Education, Science and Sport
17. Learning models of diseases and treatments for systems and personalized medicine
    Dr. Jovan Tanevski
    Ministry of Education, Science and Sport
18. ReLDI - Regional Linguistic Data Initiative
    Dr. Nikola Ljubešić
    University Of Zurich, Urpp Language and Space
19. ECML PKDD 2017 - The 28th European Conference on Machine Learning and The 21st European Conference on Principles and Practice of Knowledge Discovery in Databases; Skopje, Macedonia, 18.-22.09.2017
    Prof. Sašo Džeroski
20. CLARIN Project: Multilingual Corpus Annotation Tools: Development and Integration
    Prof. Tomaž Erjavec
    Clarin Eric

NEW CONTRACTS

1. Conducting a seminar „Data Mining in ClowdFlows and TextFlows“
   Prof. Nada Lavrač
   Comtrade d. o. o.
2. TermIolar: Development of a Prototype program solution for support of semi-automatic extraction and management of monolingual and multilingual corpora
   Dr. Senja Pollak
   Iolar d. o. o.
3. TermIOLAR2: Prototype program solution for extraction and alignment of terminology from parallel corpora of translation memories
   Dr. Senja Pollak
   Iolar d. o. o.

VISITORS FROM ABROAD

1. prof. dr. Donato Malerba, University of Bari “Aldo Moro”, Bari, Italy, 13.–15. 3. 2017.
2. prof. dr. Michelangelo Ceci, University of Bari “Aldo Moro”, Bari, Italy 2.–8. 4. 2017 and 1.–2. 6. 2017
3. dr. Ivica Dimitrovski, Ss. Cyril and Methodius, Faculty of Computer Science and Engineering, University of Skopje, Skopje, Macedonia, 2.–8. 4. 2017
4. dr. Tomislav Šmuc, Institut Ruđer Bošković, Zagreb, Croatia, 3.–5. 4. 2017
5. prof. dr. Hendrik Blockeel, KU Leuven, Department of Computer Science, Heverlee, Belgium, 31. 5.–3. 6. 2017
6. Franziska Schütze, Global Climate Forum, Berlin, Germany, 16.–21. 5. 2017
7. Dan Davis, SHARK company, Boston, USA, 5. 6. 2017
8. dr. Vesna Andova, Faculty of Electrical Engineering and Information Technologies, Skopje, Macedonia, 17.–25. 6. 2017
9. Ilin Tolovski, Faculty of Electrical Engineering and Information Technologies, Skopje, Macedonia, 17.–25. 6. 2017
10. Ana Kostovska, Faculty of Electrical Engineering and Information Technologies, Skopje, Macedonia, 17.–25. 6. 2017
11. dr. Maja Miličević, University of Belgrade, Faculty of Philology, Department of General Linguistics, Belgrade, Serbia, 20.–25. 6. 2017
12. dr. Tanja Samardić, University of Zurich, URPP Language and Space, Zurich, Switzerland, 20.–25. 6. 2017
13. dr. Saturnino Luz, Usher Institute of Population Health Sciences and Informatics, Edinburgh Medical School: Molecular, Genetic and Population Health Sciences, Edinburgh, Great Britain, 25.–27. 6. 2018 and 21.–22. 9. 2017
14. Franklin Parrales Bravo, Complutense University of Madrid, Madrid, Spain, 24. 9.-6. 12. 2017
15. prof. dr. Geoff Squire, James Hutton Institute, Dundee, Scotland, 4.–6. 9. 2017
16. Nataša Terzić, Centre for Health System Development, Institute of Public Health, Podgorica, Montenegro, 16.–19. 10. 2017
17. Jihed Khiari, NEC Europe Ltd, Heidelberg, Germany, 22.–27. 10. 2017
18. Jozef Misutka, Charles University in Prague, Czech Republic, 26.–27. 10. 2017
19. Frank van der Velde, University of Twente, Enschede, Centre for Telematics and Information Technology (CTIT), Enschede, The Netherlands, 8.–10. 11. 2017
20. prof. dr. Geraint Wiggins, Queen Mary University of London, London Great Britain, 15.–26. 11. 2017
21. dr. Maximilian Moser, Medical University of Graz, Austria, 20. 11. 2017
22. mag. Thomas Hassler, Human Research Institute, Graz, Austria, 20. 11. 2017
23. prof. dr. Henrik Bostrom, KTH Royal Institute of Technology, Stockholm, Sweden, 23.–26. 11. 2017
24. dr. Luca Leonardo Bologna, Italian National Research Council (CNR), University of Palermo, Italy, 26.–30. 11. 2017
25. Luke Lucas, European Space Agency (ESA), Darmstadt, Germany, 10.– 14. 12. 2017
26. Redouane Boumghar, European Space Agency (ESA), Darmstadt, Germany 10.–14. 12. 2017

STAFF

Researchers

1. Prof. Marko Bohanec
2. Prof. Bojan Cestnik*
3. Prof. Marko Debeljak
4. Prof. Sašo Džeroski
5. Prof. Tomaž Erjavec
6. Dr. Darja Fišer*
7. Dr. Dragi Kocev
8. **Prof. Nada Lavrač, Head**
9. Prof. Zoran Levnajić*
10. Dr. Nikola Ljubešić
11. Prof. Igor Mozetič
12. Prof. Ljupčo Todorovski*
13. Prof. Tanja Urbančič*
14. Prof. Špela Vintar*
15. Asst. Prof. Bernard Ženko
16. Asst. Prof. Martin Žnidaršič

Postdoctoral associates

17. *Dr. Darko Aleksovski, left 01.09.17*
18. *Dr. Miha Grčar, left 01.12.17*
19. Dr. Petra Kralj Novak
20. Dr. Janez Kranjc
21. Dr. Vladimir Kuzmanovski
22. *Dr. Jurica Levatić, left 01.08.17*
23. Asst. Prof. Biljana Mileva Boshkoska
24. Dr. Dragana Miljković
25. Dr. Blaž Mramor
26. Asst. Prof. Panče Panov
27. Dr. Vid Podpečan
28. Dr. Senja Pollak
29. Dr. Nikola Simidjievski
30. Dr. Jasmina Smailović
31. Dr. Jovan Tanevski
32. Dr. Aneta Trajanov
33. Dr. Anže Vavpetič
34. Dr. Vedrana Vidulin*

Postgraduates

35. Martin Breskvar, B. Sc.
36. Dr. Jan Kralj
37. Matej Martinc*, B. Sc.
38. Aljaž Osojnik, B. Sc.
39. Matej Petković, B. Sc.
40. Tomaž Stepišnik Perdih, B. Sc.
41. Tadej Škvorc, B. Sc.
42. Anita Valmarska, B. Sc.
43. Katja Zupan, B. Sc.

Technical officer

44. Milica Bauer, B. Sc.

Technical and administrative staff

46. Tina Anžič, B. Sc.

Note:
* part-time JSI member

BIBLIOGRAPHY

Original Article

1. Kristina Ban, Matjaž Perc, Zoran Levnajić, "Robust clustering of languages across Wikipedia growth", *Royal Society Open Science*, **4**, 10, 1-12, 2017.
2. Marko Bohanec, Nejc Trdin, Branko Kontić, "A qualitative multi-criteria modelling approach to the assessment of electric energy production technologies in Slovenia", *Cent. Eur. j. oper. res.*, **25**, 3, 611-625, 2017.
3. Bojan Cestnik, Elsa Fabbretti, Donatella Gubiani, Tanja Urbančič, Nada Lavrač, "Reducing the search space in literature-based discovery by exploring outlier documents: a case study in finding links between gut microbiome and Alzheimer's disease", *Genomics and computational biology*, **3**, 3, e58, 2017.
4. Dagmar Divjak, Serge Sharoff, Tomaž Erjavec, "Slavic corpus and computational linguistics", *J. Slav. linguist.*, **25**, 2, 171-198, 2017.
5. Peter Gabrovšek, Darko Aleksovski, Igor Mozetič, Miha Grčar, "Twitter sentiment around the earnings announcement events", *PloS one*, **12**, 2, e0173151, 2017.
6. Dragan Gamberger *et al.* (5 authors), "Identification of clusters of rapid and slow decliners among subjects at risk for Alzheimer's disease", *Sci. rep.*, **7**, 6763, 2017.
7. Miha Grčar, Darko Čerepnalkoski, Igor Mozetič, Petra Kralj Novak, "Stance and influence of Twitter users regarding the Brexit referendum", *Comput. soc. networks*, **4**, 6, 2017.
8. Donatella Gubiani, Elsa Fabbretti, Bojan Cestnik, Nada Lavrač, Tanja Urbančič, "Outlier based literature exploration for cross-domain linking of Alzheimer's disease and gut microbiota", *Expert syst. appl.*, **85**, 386-396, 2017.
9. Sabina Horvat, Aleksander Mahnič, Martin Breskvar, Sašo Džeroski, Maja Rupnik, "Evaluating the effect of Clostridium difficile conditioned medium on fecal microbiota community structure", *Sci. rep.*, **7**, 16448, 2017.
10. Jernej Jevšenak, Sašo Džeroski, Tom Levanič, "Uporaba metod strojnega učenja za preučevanje odnosov med značilnostmi branik in okoljem", *Acta silvae et ligni*, [Št.] 114, 21-29, dec. 2017.
11. Manja Klemenčič, Martin Žnidaršič, Anže Vavpetič, Matej Martinc, "Erasmus students' involvement in quality enhancement of Erasmus+ mobility through digital ethnography and ErasmusShouts", *Stud. high. educ. (Dorchester-on-Thames)*, **42**, 5, 925-932, 2017.
12. Janez Kranjc, Roman Orač, Vid Podpečan, Nada Lavrač, Marko Robnik Šikonja, "ClowdFlows: online workflows for distributed big data mining", *FGCS, Future gener. comput. syst.*, **68**, 38-58, 2017.
13. Marc Grau Leguia, Ralph G Andrzejak, Zoran Levnajić, "Evolutionary optimization of network reconstruction from derivative-variable correlations", *J. phys., A, Math. theor.*, **50**, 33, 334001, 2017.
14. Jurica Levatić, Michelangelo Ceci, Dragi Kocev, Sašo Džeroski, "Self-training for multi-target regression with tree ensembles", *Knowl.-based syst.*, **123**, 41-60, 2017.
15. Jurica Levatić, Michelangelo Ceci, Dragi Kocev, Sašo Džeroski, "Semi-supervised classification trees", *Journal of intelligent information systems*, **49**, 3, 461-486, 2017.
16. Matej Mihelčič, Goran Šimić, Mirjana Babić Leko, Nada Lavrač, Sašo Džeroski, Tomislav Šmuc, "Using redescription mining to relate clinical and biological characteristics of cognitively impaired and Alzheimer's disease patients", *PloS one*, **12**, 10, 0187364, 2017.
17. Matej Mihelčič, Marko Bohanec, "Approximating incompletely defined utility functions of qualitative multi-criteria modeling method DEX", *Cent. Eur. j. oper. res.*, **25**, 3, 627-649, 2017.
18. Matej Mihelčič, Sašo Džeroski, Nada Lavrač, Tomislav Šmuc, "A framework for redescription set construction", *Expert syst. appl.*, **68**, 196-215, 2017.
19. Dragana Miljković, Nada Lavrač, Marko Bohanec, Biljana Mileva-Boshkoska, "Discovering dependencies between domains of redox potential and plant defence through triplet extraction and copulas", *Int. j. intell. eng. inform.*, 1-17, 2017.
20. Jasmina Smailović, Martin Žnidaršič, Aljoša Valentinčič, Igor Lončarski, Marko Pahor, Pedro Martins, Senja Pollak, "Automatic analysis of annual financial reports: a case study", *Computación y sistemas*, **21**, 4, 809-818, 2017.
21. Erik Tjong Kim Sang *et al.* (19 authors), "The CLIN27 shared Task: Translating historical text to contemporary language for improving automatic linguistic annotation", *Computational linguistics in the Netherlands journal*, **7**, 53-64, 2017.
22. Antonio Toral, Miquel Esplá-Gomis, Filip Klubička, Nikola Ljubešić, Vassilis Papavassiliou, Prokopis Prokopidis, Raphael Rubino, Andy Way, "Crawl and crowd to bring machine translation to under-resourced languages", *Language resources and evaluation*, **51**, 4, 1019-1051, 2017.
23. Kostas M. Tsiouris *et al.* (11 authors), "PD_Manager: a mHealth platform for Parkinson's disease patient management", *Healthc. technol. lett.*, **4**, 3, 102-108, 2017.
24. Tea Tušar, Klemen Gantar, Valentin Koblar, Bernard Ženko, Bogdan Filipič, "A study of overfitting in optimization of a manufacturing quality control procedure", *Applied soft computing*, **59**, 77-87, 2017.
25. Anita Valmarska, Nada Lavrač, Johannes Fürnkranz, Marko Robnik Šikonja, "Refinement and selection heuristics in subgroup discovery and classification rule learning", *Expert syst. appl.*, **81**, 147-162, 2017.

Published Conference Contribution

1. Michael Beißwenger *et al.* (11 authors), "Closing a gap in the language resources landscape: groundwork and best practices from projects on computer-mediated communication in four European countries", In: *Selected papers from the CLARIN Annual Conference 2016*, (Linköping Electronic Conference Proceedings, **136**), 2017, 1-18.
2. Michael Beißwenger *et al.* (19 authors), "Connecting resources: which issues have to be solved to integrate CMC corpora from heterogeneous sources and for different languages?", In: *Proceedings of the 5th Conference on CMC and Social Media Corpora for the Humanities (cmccorpora17)*, Egon W. Stemle, ed., Ciara R. Wigham, ed., Bolzano, [S. n.], 2017, 10-13.
3. Mate Beštek, Dragi Kocev, Sašo Džeroski, Andrej Brodnik, Rade Iljaž, "Modelling time-series of glucose measurements from diabetes patients using predictive clustering trees", In: *Artificial intelligence in medicine: 16th Conference on Artificial Intelligence in Medicine, AIME 2017, Vienna, Austria, June 21-24, 2017: proceedings*, (Lecture notes in artificial intelligence, **10259**), Annette ten Teije, ed., 2017, 95-104.
4. Marko Bohanec, "Multi-criteria DEX model: an overview and analysis", In: *SOR '17 proceedings*, Lidija Zadnik Stirn, ed., et al, Ljubljana, Slovenian Society Informatika, Section for Operational Research, 2017, 155-160.
5. Marko Bohanec, Erik Dovgan, Pavel Maslov, Aljoša Vodopija, Mitja Luštrek, Paolo Emilio Puddu, Michele Schiariti, Maria Constanza Ciancarelli, Anneleen Baert, Sofie Pardaens, Els Clays, "Designing a personal decision support system for congestive heart failure management", In: *Slovenska konferenca o umetni inteligenci: zbornik 20. mednarodne multikonference Informacijska družba - IS 2017, 9.-13. oktober 2017, [Ljubljana, Slovenia]: zvezek A: proceedings of the 20th International Multiconference Information Society - IS 2017, 9th-13th October, 2017, Ljubljana, Slovenia: volume A*, Mitja Luštrek, ed., Rok Piltaver, ed., Matjaž Gams, ed., Ljubljana, Institut Jožef Stefan, 2017, 67-70.
6. Martin Breskvar, Dragi Kocev, Sašo Džeroski, "Multi-label classification using random label subset selections", In: *Discovery science: 20th International Conference, DS 2017, Kyoto, Japan, October 15-17, 2017: proceedings*, (Lecture notes in artificial intelligence, **10558**), Akihiro Yamamoto, ed., 2017, 108-115.
7. Martin Breskvar, Dragi Kocev, Jurica Levatić, Aljaž Osojnik, Matej Petković, Nikola Simidjievski, Bernard Ženko, Redouane Boumghar, Luke Lucas, "Predicting thermal power consumption of the Mars Express satellite with machine learning", In: *Proceedings*, 6th IEEE International Conference on Space Mission Challenges for Information Technology, SMC-IT 2017, 27-29 September 2017, Alcalá de Henares, Spain, Danvers, IEEE = The Institute of Electrical and Electronics Engineers, 2017, 88-93.
8. Vitor Cerqueira, Luís Torgo, Jasmina Smailović, Igor Mozetič, "A comparative study of performance estimation methods for time series forecasting", In: *IEEE/ACM/ASA DSAA' 2017*, Eric Gaussier, ed., Danvers, IEEE = Institute of Electrical and Electronics Engineers, 2017, 529-538.
9. Marko Debeljak, Vladimir Kuzmanovski, Veronique Tosser, Aneta Trajanov, "Knowledge discovery from complex ecological data: exploring Syrphidae species in agricultural landscapes", In: *Slovenska konferenca o umetni inteligenci: zbornik 20. mednarodne multikonference Informacijska družba - IS 2017, 9.-13. oktober 2017, [Ljubljana, Slovenia]: zvezek A: proceedings of the 20th International Multiconference Information Society - IS 2017, 9th-13th October, 2017,*

Ljubljana, Slovenia: volume A, Mitja Luštrek, ed., Rok Piltaver, ed., Matjaž Gams, ed., Ljubljana, Institut Jožef Stefan, 2017, 55-58.
10. Marko Debeljak, Florence Leprince, Sašo Džeroski, Aneta Trajanov, "Criteria for co-existence of GM and conventional Maize production", In: *Slovenska konferenca o umetni inteligenci: zbornik 20. mednarodne multikonference Informacijska družba - IS 2017, 9.-13. oktober 2017, [Ljubljana, Slovenia]: zvezek A: proceedings of the 20th International Multiconference Information Society - IS 2017, 9th-13th October, 2017, Ljubljana, Slovenia: volume A*, Mitja Luštrek, ed., Rok Piltaver, ed., Matjaž Gams, ed., Ljubljana, Institut Jožef Stefan, 2017, 51-54.
11. Ivica Dimitrovski, Dragi Kocev, Suzana Loskovska, Sašo Džeroski, "Image representation annotation and retrieval with predictive clustering trees", In: *Machine learning and knowledge discovery in databases: European Conference, ECML PKDD 2017 Skopje, Macedonia, September 18-22, 2017. Part III*, (Lecture notes in artificial intelligence, **10536**), Yasemin Altun, ed., Sašo Džeroski, ed., 2017, 363-367.
12. Kaja Dobrovoljc, Tomaž Erjavec, Simon Krek, "The Universal Dependencies Treebank for Slovenian", In: *Proceedings of the EACL workshop*, The 6th Workshop on Balto-Slavic Natural Language Processing, April 4, 2017 Valencia, Spain, Stroudsburg, The Association for Computational Linguistics = ACL, 2017, 33-38.
13. Tomaž Erjavec, Nikola Ljubešić, Darja Fišer, "Adapting a state-of-the-art tagger for South Slavic languages to non-standard text", In: *Proceedings of the EACL workshop*, The 6th Workshop on Balto-Slavic Natural Language Processing, April 4, 2017 Valencia, Spain, Stroudsburg, The Association for Computational Linguistics = ACL, 2017, 60-68.
14. Darja Fišer, Jakob Lenardič, "Parliamentary Corpora in the CLARIN infrastructure", In: *Proceedings*, [S. l., s. n.], 2017, 1-4.
15. Darja Fišer, Nikola Ljubešić, Tomaž Erjavec, "Legal framework, dataset and annotation schema for socially unacceptable online discourse practices in Slovene", In: *Proccedings of the 1st Workshop on Abusive Language Online*, 1st Workshop on Abusive Language Online, ACL 2017, Vancouver, Canada, August 4 2017, Vancouver, ACL, 2017, 46-51.
16. Valentin Gjorgjioski, Dragi Kocev, Andrej Bončina, Sašo Džeroski, Marko Debeljak, "Predictive clustering of multi-dimensional time series applied to forest growing stock data for different tree sizes", In: *Data-driven innovation: 9th International Conference, ICT Innovations 2017 Skopje, Macedonia, September 18-23, 2017: proceedings*, (Communications in computer and information science, **758**), Dimitar Trajanov, ed., Veriva Bakeva, ed., 2017, 186-195.
17. Taja Kuzman, Darja Fišer, "Corpus-based analysis of demonyms in Slovene Twitter", In: *Proceedings of the 5th Conference on CMC and Social Media Corpora for the Humanities (cmccorpora17)*, Egon W. Stemle, ed., Ciara R. Wigham, ed., Bolzano, [S. n.], 2017, 29-33.
18. Jurica Levatić, Marija Brbić, Tomaž Stepišnik Perdih, Dragi Kocev, Vedrana Vidulin, Tomislav Šmuc, Fran Supek, Sašo Džeroski, "Phenotype prediction with semi-supervised learning", In: *New frontiers in mining complex patterns: Sixth edition of the International Workshop NFMCP 2017 in conjunction with ECML-PKDD 2017, September 18-22, 2017, Skopje, Macedonia*, 2017.
19. Nikola Ljubešić, Tomaž Erjavec, Darja Fišer, Enhard Hinrichs, Marie Hinrichs, Cyprian Laskowski, Filip Petkovski, Wei Qui, "Multilingual text annotation of Slovenian, Croatian and Serbian with WebLicht", In: *Proceedings*, CLARIN Annual Conference 2017, 18-20 September, Budapest, Hungary, 2017, 18.
20. Nikola Ljubešić, Darja Fišer, Tomaž Erjavec, "Language-independent gender prediction on Twitter", In: *Proceedings of the Second Workshop on Natural Language Processing and Computational Social Science*, The 2nd Workshop on Natural Language Processing and Computational Social Science, August 4, Vancouver, Canada, Vancouver, Association for Computational Linguistics, 2017, 1-6.
21. Žiga Lukšič, Jovan Tanevski, Sašo Džeroski, Ljupčo Todorovski, "General meta-model framework for surrogate-based numerical optimization", In: *Discovery science: 20th International Conference, DS 2017, Kyoto, Japan, October 15-17, 2017: proceedings*, (Lecture notes in artificial intelligence, **10558**), Akihiro Yamamoto, ed., 2017, 51-66.
22. Matej Martinc, Iza Škrjanec, Katja Zupan, Senja Pollak, "PAN 2017: author profiling - gender and language variety prediction: notebook for PAN at CLEF 2017", In: *CLEF 2017: CLEF 2017 Working notes, Conference and Labs of the Evaluation Forum, Dublin, Ireland, September 11-14, 2017*, (CEUR workshop proceedings, **1866**), Linda Cappellato, ed., 2017, 78.
23. Vanja Mileski, Sašo Džeroski, Dragi Kocev, "Predictive clustering trees for hierarchical multi-target regression", In: *Advances in intelligent data analysis XVI: 16th International Symposium on Intelligent Data Analysis, IDA 2017, London, UK, October 26 28, 2017: proceedings*, (Lecture notes in computer science, **10584**), Niall M. Adams, ed., Allan Tucker, ed., David Weston, ed., 2017, 223-234.
24. Biljana Mileva-Boshkoska, Dragana Miljković, Anita Valmarska, Dimitros Gatsios, George Rigas, Spyros Konitsiotis, Kostas M. Tsiouris, Dimitrios I. Fotiadis, Marko Bohanec, "A state-transition decision support model for medication Change of Parkinson's disease patients", In: *Slovenska konferenca o umetni inteligenci: zbornik 20. mednarodne multikonference Informacijska družba - IS 2017, 9.-13. oktober 2017, [Ljubljana, Slovenia]: zvezek A: proceedings of the 20th International Multiconference Information Society - IS 2017, 9th-13th October, 2017, Ljubljana, Slovenia: volume A*, Mitja Luštrek, ed., Rok Piltaver, ed., Matjaž Gams, ed., Ljubljana, Institut Jožef Stefan, 2017, 63-66.
25. Stevanche Nikoloski, Dragi Kocev, Sašo Džeroski, "Decomposition of the output space in multi-label classification using feature ranking", In: *New frontiers in mining complex patterns: Sixth edition of the International Workshop NFMCP 2017 in conjunction with ECML-PKDD 2017, September 18-22, 2017, Skopje, Macedonia*, [S. l., s. n.], 2017.
26. Gjorgi Peev, Nikola Simidjievski, Sašo Džeroski, "Modeling of dynamical systems: a survey of tools and a case study", In: *Slovenska konferenca o umetni inteligenci: zbornik 20. mednarodne multikonference Informacijska družba - IS 2017, 9.-13. oktober 2017, [Ljubljana, Slovenia]: zvezek A: proceedings of the 20th International Multiconference Information Society - IS 2017, 9th-13th October, 2017, Ljubljana, Slovenia: volume A*, Mitja Luštrek, ed., Rok Piltaver, ed., Matjaž Gams, ed., Ljubljana, Institut Jožef Stefan, 2017, 15-18.
27. Matej Petković, Sašo Džeroski, Dragi Kocev, "Feature ranking for multi-target regression with tree ensemble methods", In: *Discovery science: 20th International Conference, DS 2017, Kyoto, Japan, October 15-17, 2017: proceedings*, (Lecture notes in artificial intelligence, **10558**), Akihiro Yamamoto, ed., 2017, 171-185.
28. Matej Petković, Jovan Tanevski, Aleš Maver, Lovro Vidmar, Borut Peterlin, Sašo Džeroski, "Comparison of feature ranking approaches for discovery of rare genetic variants related to multiple sclerosis", In: *Slovenska konferenca o umetni inteligenci: zbornik 20. mednarodne multikonference Informacijska družba - IS 2017, 9.-13. oktober 2017, [Ljubljana, Slovenia]: zvezek A: proceedings of the 20th International Multiconference Information Society - IS 2017, 9th-13th October, 2017, Ljubljana, Slovenia: volume A*, Mitja Luštrek, ed., Rok Piltaver, ed., Matjaž Gams, ed., Ljubljana, Institut Jožef Stefan, 2017, 11-14.
29. Senja Pollak, Ana Marija Sobočan, "Corpus-assisted discourse analysis of family legislation changes in Slovene media", In: *9th International Corpus Linguistics Conference, 25 - 28 July 2017, Birmingham, GB: extended abstracts*, Birmingham, University of Birmingham, 2017, 765-769.
30. Damjan Popič, Darja Fišer, "Fear and loathing on Twitter: attitudes towards language", In: *Proceedings of the 5th Conference on CMC and Social Media Corpora for the Humanities (cmccorpora17)*, Egon W. Stemle, ed., Ciara R. Wigham, ed., Bolzano, [S. n.], 2017, 61-64.
31. Andraž Repar, Senja Pollak, "Good examples for terminology databases in translation industry", In: *Electronic lexicography in the 21st century: proceedings of eLex 2017 Conference, 19-21 September 2017, Leiden, The Netherlands*, Iztok Kosem, ed., et al, Brno, Lexical Computing, 2017, 650-661.
32. Andraž Repar, Senja Pollak, "Ontology-based translation memory maintenance", In: *Slovenska konferenca o umetni inteligenci: zbornik 20. mednarodne multikonference Informacijska družba - IS 2017, 9. - 13. oktober 2017, Ljubljana, Slovenia: zvezek C: proceedings of the 20th International Multiconference Information Society - IS 2017, 9-13 October, 2017, Ljubljana, Slovenia: volume C*, Dunja Mladenić, ed., Marko Grobelnik, ed., Ljubljana, Institut Jožef Stefan, 2017, 19-22.
33. Tanja Samardžić, Mirjana Starović, Željko Agić, Nikola Ljubešić, "Universal dependencies for Serbian in comparison with Croatian and other slavic languages", In: *Proceedings of the EACL workshop*, The 6th Workshop on Balto-Slavic Natural Language Processing, April 4, 2017 Valencia, Spain, Stroudsburg, The Association for Computational Linguistics = ACL, 2017, 39-44.
34. Tomaž Stepišnik Perdih, Aljaž Osojnik, Sašo Džeroski, Dragi Kocev, "Option predictive clustering trees for hierarchical multi-label classification", In: *Discovery science: 20th International Conference, DS 2017, Kyoto, Japan, October 15-17, 2017: proceedings*, (Lecture notes in artificial intelligence, **10558**), Akihiro Yamamoto, ed., 2017, 116-123.
35. Jovan Tanevski, Nikola Simidjievski, Ljupčo Todorovski, Sašo Džeroski, "Process-based modeling and design of dynamical systems", In: *Machine learning and knowledge discovery in databases: European Conference, ECML PKDD 2017 Skopje, Macedonia, September 18 - 22, 2017: proceedings. Part I*, (Lecture notes in artificial intelligence, **10534**), Michelangelo Ceci, ed., Ljupčo Todorovski, ed., 2017, 78-382.
36. Jovan Tanevski, Nikola Simidjievski, Ljupčo Todorovski, Sašo Džeroski, "Process-based modeling and design of dynamical systems", In: *Machine learning and knowledge discovery in databases: European*

Conference, ECML PKDD 2017 Skopje, Macedonia, September 18-22, 2017. Part III, (Lecture notes in artificial intelligence, **10536**), Yasemin Altun, ed., Sašo Džeroski, ed., 2017, 378-382.
37. Anita Valmarska, Dragana Miljković, Spyros Konitsiotis, Dimitros Gatsios, Nada Lavrač, Marko Robnik Šikonja, "Combining multitask learning and short time series analysis in Parkinson's disease patients stratiscation", In: *Artificial intelligence in medicine: 16th Conference on Artificial Intelligence in Medicine, AIME 2017, Vienna, Austria, June 21-24, 2017: proceedings*, (Lecture notes in artificial intelligence, **10259**), Annette ten Teije, ed., 2017, 116-125.
38. Anita Valmarska, Dragana Miljković, Marko Robnik Šikonja, Nada Lavrač, "Multi-view approach to Parkinson's disease quality of life data analysis", In: *New frontiers in mining complex patterns: 5th International Workshop NFMCP 2016, held in conjunction with ECML-PKDD 2016, Riva del Garda, Italy, September 19, 2016: revised selected papers*, Annalisa Appice, ed., Cham, Springer, 2017, 163-178.
39. Ben Verhoeven, Iza Škrjanec, Senja Pollak, "Gender profiling for slovene twitter communication: the influence of gender marking, content and style", In: *Proceedings of the EACL workshop*, The 6th Workshop on Balto-Slavic Natural Language Processing, April 4, 2017 Valencia, Spain, Stroudsburg, The Association for Computational Linguistics = ACL, 2017, 119-125.
40. Marcos Zampieri, Shervin Malmasi, Nikola Ljubešić, Preslav Nakov, Ahmed Ali, Jörg Tiedemann, Yves Scherrer, Noëmi Aepli, "Findings of the VarDial evaluation campaign 2017", In: *Proceedings of the Workshop*, Fourth Workshop on NLP for Similar Languages, Varieties and Dialects, (VarDial2017), April 3, 2017, Valencia, Spain, Stroudsburg, Association for Computational Linguistics, 2017, 1-15.

Independent Component Part or a Chapter in a Monograph

1. Kaja Dobrovoljc, Simon Krek, Tomaž Erjavec, "The Sloleks morphological lexicon and its future development", In: *Dictionary of modern Slovene: problems and solutions*, (Book series Prevodoslovje in uporabno jezikoslovje), Vojko Gorjanc, ed., et al, 1st ed., e-ed., Ljubljana, Ljubljana University Press, Faculty of Arts, 2017, 42-63.
2. Tomaž Erjavec, "MULTEXT-East", In: *Handbook of linguistic annotation*, Nancy M. Ide, ed., James Pustejovsky, ed., James Pustejovsky, Amsterdam, Springer, 2017, 441-462.
3. Tomaž Erjavec, Darja Fišer, Nikola Ljubešić, Nataša Logar, Vesna Mikolič, "The expansion of the Gigafida corpus: internet content", In: *Dictionary of modern Slovene: problems and solutions*, (Book series Prevodoslovje in uporabno jezikoslovje), Vojko Gorjanc, ed., et al, 1st ed., e-ed., Ljubljana, Ljubljana University Press, Faculty of Arts, 2017, 120-138.
4. Tomaž Erjavec, Peter Holozan, Nikola Ljubešić, "Language technologies and corpus encoding", In: *Dictionary of modern Slovene: problems and solutions*, (Book series Prevodoslovje in uporabno jezikoslovje), Vojko Gorjanc, ed., et al, 1st ed., e-ed., Ljubljana, Ljubljana University Press, Faculty of Arts, 2017, 140-154.
5. Mark A. Finlayson, Tomaž Erjavec, "Overview of annotation creation: processes and tools", In: *Handbook of linguistic annotation*, Nancy M. Ide, ed., James Pustejovsky, ed., James Pustejovsky, Amsterdam, Springer, 2017, 167-192.
6. Darja Fišer, Jaka Čibej, "Crowdsourcing workflows in lexicography", In: *Dictionary of modern Slovene: problems and solutions*, (Book series Prevodoslovje in uporabno jezikoslovje), Vojko Gorjanc, ed., et al, 1st ed., e-ed., Ljubljana, Ljubljana University Press, Faculty of Arts, 2017, 230-244.
7. Darja Fišer, Jaka Čibej, "The potential of crowdsourcing in modern lexicography", In: *Dictionary of modern Slovene: problems and solutions*, (Book series Prevodoslovje in uporabno jezikoslovje), Vojko Gorjanc, ed., et al, 1st ed., e-ed., Ljubljana, Ljubljana University Press, Faculty of Arts, 2017, 212-228.
8. Darja Fišer, Tomaž Erjavec, Nikola Ljubešić, "The compilation, processing and analysis of the JANES corpus of Slovene user-generated content", In: *Corpus de communication médiée par les réseaux: construction, structuration, analyse*, (Collection Humanités numériques), Ciara R. Wigham, ed., Gudrun Ledegen, ed., Paris, L'Harmattan, cop. 2017, 125-138.
9. Maja Miličević, Nikola Ljubešić, Darja Fišer, "Birds of a feather don't quite tweet together: an analysis of spelling variation in Slovene, Croatian and Serbian twitterese", In: *Investigating computer-mediated communication: corpus-based approaches to language in the digital world*, (Book series Translation studies and applied linguistics), Darja Fišer, ed., Michael Beißwenger, ed., 1st ed., Ljubljana, University Press, Faculty of Arts, 2017, 14-43.
10. Friedrich Recknagel, Dragi Kocev, Cao Hongqin, Christina Castelo Branco, Ricardo Minoti, Sašo Džeroski, "Inferential modelling of population dynamics", In: *Ecological informatics: data management and knowledge discovery*, Friedrich Recknagel, ed., William K. Michener, ed., 3rd ed., Cham, Springer, 2017, 161-188.
11. Andrejka Žejn, Tomaž Erjavec, Matija Ogrin, "Poljanski rokopis: Baročno besedilno izročilo o Jezusovem življenju v elektronski izdaji", In: *Rokopisi slovenskega slovstva od srednjega veka do moderne*, (Obdobja, **36**), Aleksander Bjelčevič, ed., Matija Ogrin, ed., Urška Perenič, ed., 1. natis, Ljubljana, Znanstvena založba Filozofske fakultete, 2017, 111-118.

Scientific Monograph

1. Pavle Boškoski, Andrej Debenjak, Biljana Mileva-Boshkoska, *Fast electrochemical impedance spectroscopy: as a statistical condition monitoring tool*, (Springer briefs in applied sciences and technology), Cham, Springer, 2017.
2. Blaž Podpečan, Vid Podpečan, *Govorica tišine: nagrobnik kot vir podatkov*, (Archaeologia historica Slovenica, **7**), 1. izd., Ljubljana, Znanstvena založba Filozofske fakultete, 2017.

Mentoring

1. Jože Bučar, *Sentiment based classification of the web texts:* doctoral dissertation, Novo Mesto, 2017 (mentor Janez Povh; co-mentor Martin Žnidaršič).
2. Jan Kralj, *Heterogeneous information network analysis for semantic data mining:* doctoral dissertation, Ljubljana, 2017 (mentor Nada Lavrač; co-mentor Marko Robnik Šikonja).
3. Janez Kranjc, *Web workflows for data mining in the cloud:* doctoral dissertation, Ljubljana, 2017 (mentor Nada Lavrač; co-mentor Marko Robnik Šikonja).
4. Jurica Levatić, *Semi-supervised learning for structred output prediction:* doctoral dissertation, Ljubljana, 2017 (mentor Sašo Džeroski; co-mentor Dragi Kocev).