

Slavko Žitnik

slavko.zitnik@fri.uni-lj.si



Vrbičje 4
SI-1290 Grosuplje
+386 31 543 547
skype: slavkozitnik
sicris: 34156

<http://zitnik.si>
<http://zitnik.si/research>
<http://www.linkedin.com/in/szitnik>
<https://bitbucket.org/szitnik>
<https://github.com/szitnik>

Employment

- 2014–present **University of Ljubljana, Faculty for computer and information science**
Assistant Professor
I am a faculty member of the [Laboratory for Data Technologies](#) and I am mainly researching in the fields of text information extraction and retrieval. I am also teaching courses of natural language processing, web information extraction and retrieval, database design and management and information systems and modeling. I also work on projects in the fields of Internet of Things (*EkoSMART*), information extraction (*Ministry of Public Administration*), text search in NoSQL databases (*Iskratel*), data warehouse design for large real-time streams processing (*DRI*) and digitalization of a company processes for optimization and assets numbering and tracking (*Kolektor d.o.o.*). Internationally I actively cooperate with the [Université Paris 1, Sorbonne](#), the [School of Electrical Engineering, University of Belgrade](#), [College of Education, University of South Florida](#) and [Harvard University](#)
- 2014–2014 **Microsoft Development Center Norway, Oslo**
Software Development Engineer in Test Intern
The research and development center [MDCN](#) in Oslo focuses on information retrieval technologies within the enterprises. Their services are used by Microsoft's products such as Office, Sharepoint, Exchange and others. As an intern I participated in the feeding team, which goal is to extract appropriate data from a document and feed the index with it. My task was to enable the embedded documents parsing, calculate expected index growth, monitor the three-phase deployment process across all the farms in the world and provide reports that show the correct behaviour of the feature. With a group of four buddies we attended the internal Microsoft hackathon, where we proposed an algorithm and implemented a prototype that enables the expert search in Microsoft Delve. My role was to design an algorithm for expertise extraction and experts ranking. The local management liked the idea of our approach, so we also presented it in Redmond, where they decided to apply it for a patent.
- 2011–2014 **Optilab d.o.o. & University of Ljubljana**
Junior Researcher from Industry
[Optilab](#) is a company, which focuses mainly on fraud management systems - [Admiral](#). As a Junior Researcher I try to merge data from arbitrary data sources and extract structured data from textual sources, which can help in anomaly detection. I am also trying to transfer research results into usable prototypes or end products. Furthermore, I am responsible for [science and tech talks](#) organization at Optilab. My research is conducted at the [Faculty of Computer and Information Science](#), in [Laboratory for Data Technologies](#), under supervision of [Prof. Dr. Marko Bajec](#). My field of research is focused on information extraction (i.e., entity extraction, relation extraction and coreference resolution) from textual data, for which most of my research-based source code is [publicly available](#).
- 2009–2010 **Mikro Graf Art d.o.o.**
Software developer
Fully independent design and development of [Sociogram v4.01](#). Java-based software with automatic licensing, which is *currently used by more than 50 Slovenian primary schools*.
- 2008–2009 **Avtenta.si**
Software developer
Integration of call log data streams between three different call center systems. Development of a meeting minutes application on Sharepoint 2007 server for Telekom Slovenije d.d.. Development of internal phonebook and reports application in ASP.NET and Cisco Call Manager (CDR) for brokerage house Argonos.
- 2007–2007 **Hermes Softlab d.o.o.**, Vendors department

Software developer

Development of a Timecard application for The SEZ Group in Java SE and Jasper Reports. Development of a wafer control server for The SEZ Group in C++.

Education

2010–2014 | **PhD** University of Ljubljana, Faculty of Computer and Information Science, *Ljubljana*, (*GPA of 9.0 on scale 1-10*), Operation financed by Slovene Ministry of Economic Development and Technology and European Social Fund.

2006–2010 | **BSc** University of Ljubljana, Faculty of Computer and Information Science and Faculty of Mathematics and Physics, *Ljubljana*, (*GPA of 9.0 on scale 1-10*), Interdisciplinary University Program of Computer Science and Mathematics

2002–2006 | Upper Secondary School of Electrical and Computer Engineering, Vegova, *Ljubljana*, *summa cum laude*, Technical Gymnasium

Research projects

2021–2022 | **DE4A - Digital Europe for All EU Project**

Investigator

Design of ontology-based schemas and system architecture for semantic interoperability between EU Member States. Definition of authorization controls and exact legal proofs to be sent electronically. [□](#)

2021–2022 | **CLARIN.SI project - “SloBENCH”**

PI

Implementation of an independent framework for evaluation of Slovene language technologies. SloBENCH tool is available to anyone to submit results of his language tool to a specific leaderboard (e.g., named entity recognition, question answering). Evaluation scripts are publicly available in a Github repository, while reference annotations for test dataset are hidden.

2020–2023 | **Development of Slovene in a Digital Environment, National project**

Semantic technologies co-PI

Development of language resources and tools to equip Slovene with advanced Natural Language Processing technologies. In WP3 we develop NER, REL, COREF, knowledge base, question answering, summarization, tools for semantic shift detection, etc. [□](#)

2019–2022 | **COST project - “Nexus Ligarium”**

Research collaborator, Task leader

Goal of the COST action is to connect researcher working in the field of linguistics, natural language processing and semantic Web. My role is related to WG4 where we want to apply NLP algorithms to different domains.

2019–2021 | **ARRS National bilateral project - “Tools for automatic embedding and text-based interpretation of biomedical knowledge graphs”**

PI

Research collaboration with prof. Marinka Žitnik from the Harvard University. Idea of the project is to build text embeddings for information retrieval and interpretation. Project code: BI-US/19-21-013

2019–2021 | **Press clipping d.o.o.**

PI

Processing of daily news from electronic and printed media. Design, implementation and installation of scalable architecture for processing and searching over the data.

2019 | **Clarín.si - “Corpus for Slovene coreference resolution and aspect-based sentiment analysis–SentiCoref 1.0”**

PI

Preparation of the corpus that includes coreference chains and sentiment for each chain.. [□](#)

2018–2019 | **ARRS National bilateral project - “Web-based eBooks with activities: internationalization of Natural Language Processing”**

Co-PI

Research collaboration with prof. Glenn Smith, University of South Florida in the field of reading comprehension. Automatic grading of open-ended questions in IMapBook project. Project code: BI-US/18-19-043

Professional projects

2020–2022	Ineor d.o.o. Establishment of a company's internal <i>Ineor Business AI</i> , design of an internal staff certification scheme. The official launch of the AI centre happened at Expo 2020 in Dubai, December 16, 2021. Training a group of employees to lead and develop for the AI technologies - related to MS Azure technologies. Organization of the Ineor Talent Hackathon.
2021–2022	Inden d.o.o. Design and implementation of search services based on Apache Solr. Services are used in production by the Slovene Legal Information system. □
2018–2020	DARS d.d. Preparation of general technical specifications for the Intelligent Transportation Systems in traffic control centres. Application level architecture design.
2018	Kolektor d.o.o. Digitalization of business processes for optimization and assets numbering and tracking. Guidelines design and a prototype implementation.
2017–2018	Ministry of Public Administration Design and implementation of a layered ontology to support semantic mapping between national registries into interoperable data services.
2016–2019	EkoSMART project Within the programme of projects I am responsible to develop and design an IoT platform for remote patient monitoring.
2016	Iskratel d.d. Implementation and optimization of full-text search and indexing in Cassandra NoSQL database.
2014–2016	Krimar.si and Mikro Graf Art d.o.o. Independed design and implementation of the new application Sociogram 5.0. The new version supports end-to-end sociometric analysis for schools, teams and industry. Due to existing customers excitements about Sociogram v4.01 we added new functionalities and also supported additional languages.
2013	Lajkbuk.si Idea and development of a Facebook like platform using Spring (MVC, Social, Security) Framework and Hibernate. The platform enables the automatic arrangement of Facebook prize competitions for companies that would like to get more likes for their Facebook pages.
2013	Institute of Clinical Neurophysiology, Ljubljana Development of a register of extrapyramidal diseases in Spring (WebFlow, Security) Framework, JSF 2.0 and Hibernate.
2010	Abito d.o.o. Co-authoring Computer networks multimedia e-learning materials for Slovenian Vocational Colleges, Sharepoint Server administration.
2008	Mikro Graf Art d.o.o. Fully independent design and development of CrossManager 1.0 - a sports timing software, written in C# and WPF, that supports competitors import, barcode scanning at finish line and automatic report and diploma generation. The software is still used to organize cross competitions in Ljubljana primary schools.
2007	Hermes Softlab d.o.o Summer school participant, development of a web application in Java EE, Hibernate and ZKOSS UI framework.

Language skills

Slovene	Native speaker (Mother tongue)
English	Fluent (C2 - Proficient user)
German, Croatian, Serbian	Intermediate (B1 - Independent user)

Honors, awards & fellowships

2019	Award of the University of Ljubljana for extraordinary pedagogical, research and artistic achievements.
2019	Silver sign from the Slovenian Society INFORMATIKA for contributions in development and reputation of IT in Slovenia.
2018	Award for contributions in development and reputation from the Slovenian Society INFORMATIKA.
2013	First place in relationship extraction between proteins and genes from biomedical texts for gene regulatory networks construction at the BioNLP 2013 challenge.
2013	Seventh place in drugs and chemical compounds recognition from biomedical texts at the CHEMDNER challenge.
2010	Gold medal as a Cisco specialist at the EuroSkills 2010 competition.
2010	First place as an IT team at the SloSkills 2010 competition.
2010	Faculty award for best students.
2009	Bronze medal at the Imagine Cup Competition.
2007	Faculty award for best students.
2002–2010	Zois fellowship.

Invited teaching

2015–2015	In July 2015 prof. dr. Marko Bajec and I led the Summer School course of Database design and modeling at the Kyungpook National University in Daegu, South Korea.
2009–2016	CCNA courses instructor at the Cisco Networking Academy. Lecturing, conducting exercises and marking for CCNA 1–4 classes.
2010–2016	Teaching assistant for computer networking classes at Vocational College Kranj. Conducting exercises and helping students at practical work on Cisco network appliances.
2008–2010	Teaching assistant for programming classes at the Faculty of Computer and Information Science (Java, C and Python). Helping and mentoring undergraduate students at programming exercises.

Associations

2020–present	Professional Council of National and University Library of Slovenia Member of the council The association is dealing with the strategy and development of the Slovenian national library.
2019–present	Slovenian Society INFORMATIKA Member of the supervisory board The association is dealing with the development of the IT in Slovenia. I cooperate at the organization of the regular conferences DSI and IJU and as a technical editor take care for the Uporabna informatika journal.
2016–2021	Electronic Communications Council of the Republic of Slovenia Chairman of the Council The Council is a body to advise on the development of electronic communications and the protection of consumers' interests in the field of electronic communications in the Republic of Slovenia. The council prepares opinions, recommendations, proposals according to the Slovene legislation act ZEKom-1, articles 229 to 231. □
2016–2018	Agency for communication networks and services of the Republic of Slovenia Member of the supervisory board As a member of the supervisory board I am responsible for the supervision of operations at the Agency as defined by the Slovene legislation act ZEKom-1, articles 174 to 180. □
2012–2015	Doctoral consortium of the University of Ljubljana As a representative of the doctoral students at the University of Ljubljana I cooperated in decision making regarding the doctoral curriculum changes. Additionally, the main role of the consortium is to approve PhD thesis proposals. Because of that I learned a lot about the definition of a scientific excellence for specific research areas.
2012–2015	Association of Doctoral Candidates in Slovenia As a board member of the association I co-organized a two-day conference regarding the employment possibilities of PhD students in industry after they graduate. I was a chairman of comitee for selection of the best PhD supervisor of year 2013 in Slovenia. I also reorganized the IT infrastructure of the association.
2008–2010	Electrical engineering students' European Association

As a member and oversight committee I co-organized [JobFair](#) 2009 and 2010 events. This is a yearly fair at the Faculty of Computer and Information Science in Ljubljana, intended to connect students and companies. I also reorganized internal IT structure of the association (mailing lists, documents, permissions, transition to Google Apps).

Service to the profession

Reviewing	Computational Linguistics. Computers & Education: Artificial Intelligence. Computer Science and Information Systems: ComSis. Journal of Ambient Intelligence and Humanized Computing. PLoS One. Journal of Cheminformatics. Journal of Web Semantics. Journal of Computer Science and Technology. International Journal of Information System Modeling and Design. IEEE RCIS – International Conference on Research Challenges in Information Science. CAiSE – International Conference on Advanced Information Systems Engineering. ADBIS – East-European Conference on Advances in Databases and Information Systems. KMO – Knowledge Management Conference. BIR – International Conference on Perspectives in Business Informatics Research
Conferences	DSI 2015, <i>Co-organizer and student section chair.</i> □ DSI 2016, <i>Co-organizer, welcome roundtable chair, IoT section chair, student section chair.</i> □ CAiSE 2016, <i>Co-organizer.</i> □ DSI 2017, <i>Co-organizer and student section chair.</i> □ IJU 2017, <i>Organizer of IoT round table.</i> RCIS 2018, <i>Demo and poster session chair.</i> □ DSI 2018, <i>Co-organizer and student section chair.</i> □ DSI 2019, <i>Co-organizer and student section chair.</i> □ IJU 2019, <i>Member of the organization committee.</i> □ DSI 2020-2021, <i>Chairman of the organization committee, member of the program committee.</i> □ LDK 2021, SALLD-1 workshop, <i>Co-organizer and program committee member.</i> □ EACL 2021, Hackashop workshop, <i>Program committee member.</i> □
Editor	Uporabna informatika, <i>Technical editor (2018–present).</i> □

Publications & Patents

Patents & Applications

Platakis, M.; Makris, C.; Tonnesen Lied, T.; Herstad, B.; Thomaidou, S.; Žitnik, S. (2016). Identifying experts and areas of expertise in an organization, in *United States Patent Application*, Microsoft, Redmond, US 14/695,822. [□](#)

Šubelj L., Žitnik S., Lavbič D., Kumer A., and Bajec M. (2013). Combining heterogeneous data sources of data processing devices, in *Slovenian Institute of Intellectual Property*, Ljubljana, Slovenia, P-201300263, pp. 1-7. [□](#)

Books

Žitnik S. and Bajec M. (2013). Text Mining in Medicine, Chapter in *Computational Medicine in Data Mining and Modeling*, New York: Springer, pp. 105-134. [□](#)

Peer Reviewed Journals

Žitnik S., Blagus N. and Bajec M. (2022). Target-level sentiment analysis for news articles. Knowledge-based systems, in *Knowledge-Based Systems*, vol. 249, num. 108939, pp. 1-46. [□](#)

Kljun M., Teršek M. and Žitnik S. (2022). Semantic analysis of offensive language categories from existing annotated corpora, in *Uporabna informatika*, vol. 30, num. 1, pp. 3-18. [□](#)

Klemen M., Žitnik S. (2021). Neural coreference resolution for Slovene language, in *Computer science and information sy-*

stems, pp. 1-27. □

Smith GG., Haworth R., Žitnik S. (2020). Computer science meets education: Natural Language Processing for automatic grading of open-ended questions in eBooks, in *Journal of educational computing research*, pp. 1-29. □

Jankovič M., Žitnik S. and Bajec M. (2019). Reconstructing de facto software development methods. Computer science and information systems, in *Computer science and information systems*, vol. 16, num. 1, pp. 75-104. □

Epure EV., Compagno D., Salinesi C., Deneckere R., Bajec M. and Žitnik S. (2018). Process models of interrelated speech intentions from online health-related conversations, in *Artificial intelligence in medicine*, July 2018. □

Žitnik S. and Bajec M. (2018). Odkrivanje koreferenčnosti v slovenskem jeziku na označenih besedilih iz coref149, in *Slovensščina 2.0: empirične, aplikativne in interdisciplinarne raziskave*, pp. 36-66. □

Žitnik S., Žitnik M., Zupan B. and Bajec M. (2015). Sieve-Based Relation Extraction of Gene Regulatory Networks from Biological Literature, in *BMC Bioinformatics*, pp. 1-16. □

Krallinger M., Žitnik S., Bajec M., et al. (2015). The CHEMDNER corpus of chemicals and drugs and its annotation principles, in *Journal of Cheminformatics*, 7, suppl. 1, pp. 1-17. □

Fiala, D.; Šubelj, L.; Žitnik, S.; Bajec, M. (2015). Do PageRank-based author rankings outperform simple citation counts?, in *Journal of informetrics*, vol. 9, num. 2, pp. 334-348. □

Žitnik S., Šubelj L. and Bajec M. (2014). SkipCor: Skip-Mention Coreference Resolution using Linear-Chain Conditional Random Fields, in *PLoS One*, 9(6), pp. 1-14. □

Šubelj L., Žitnik S., et al. (2013). Node mixing and group structure of complex software networks, in *Advances in Complex Systems*, 17, 1450022, pp. 1-26. □

Jankovič M., Žitnik S., et al. (2013). Improvement of software development method with minimal involvement of developers, in *Uporabna Informatika*, 22(1), pp. 1-12. □

Šubelj L., Žitnik S., et al. (2013). Large real-world networks, in *Uporabna Informatika*, vol. 21, num. 2, pp. 107-119. □

Žitnik S., et al. (2013). General Context-Aware Data Matching and Merging Framework, in *Informatika*, vol. 24, num. 1, pp. 119-152. □

Zrnec A., et al. (2012). Podatkovne baze NOSQL, in *Uporabna informatika*, vol. 20, num. 3, pp. 164-172. □

Peer Reviewed Conferences

Knez T., Bajec M., Žitnik S. (2022). ANGLEr : a next-generation natural language exploratory framework, in *Research challenges in information science : 16th International Conference, RCIS 2022*, Barcelona, Spain, May 17-20, pp. 761-768, □

Lewandowska-Tomaszczyk B., Žitnik S., Baczkowska A., Liebeskind C., Mitrović J. and Oleskeviciene GV. (2021). LOD-connected offensive language ontology and tagset enrichment, in *LDK 2021 : LDK Workshops and Tutorials 2021 : proceedings of the Workshops and Tutorials - Language Data and Knowledge 2021, co-located with the Language, Data and Knowledge Conference (LDK 2021)*, Zaragoza, Spain, September 1-4, pp. 1-16, □

Prelevikj M. and Žitnik S. (2021). Multilingual named entity recognition and matching using BERT and dedupe for Slavic languages, in *Proceedings of 8th BSNLP Workshop on Balto-Slavic Natural Language Processing*, Stroudsboung, April 20, 2021, pp. 80-85. □

Bajec M., Jankovič M., Žitnik S., and Lebar Bajec I. (2020). Punctuation restoration system for Slovene language, in *Proceedings of Research challenges in information science : 14th International Conference*, Cyprus, Limassol, September 23-25, 2020, pp. 509-514. □

Blagus N. and Žitnik S. (2018). Social media comparison and analysis : the best data source for research?, in *RCIS 2018: conference proceedings*, France, Nantes, 29.-31. May 2018, pp. 1-10.

Blagus N., Žitnik S. and Bajec M. (2018). Družbeni mediji kot vir podatkov za raziskave, in *Digitalizacija in mi: zbornik konference Dnevi slovenske informatike*, Slovenija, Portorož, 17.-18. April 2018, pp. 1-12.

Jankovič, M.; Žitnik, S.; Bajec, M. (2017). Arhitektura sistema za oddaljeno spremljanje pacientov, in *Workshop Electronic and mobile health and smart cities : proceedings of the 20th International Multiconference Information Society*, Slovenia, Ljubljana, 9.-13. October 2017, pp. 27-28. □

Žitnik, S.; Drašković, D.; Nikolić B.; Bajec M. (2017). nutIE - a modern open source natural language processing toolkit, in *Proceedings of papers, 25th Telecommunications Forum (TELFOR)*, Belgrade, Serbia, 21.-22. November 2017, pp. 1-4. □

Žitnik, S.; Bajec, M.; Lavbič, D. (2017) LogMap+ : relational data enrichment and linked data resources matching, in *RCIS 2017 : conference proceedings*, United Kingdom, Brighton, 10.-12. May 2017, pp. 267-275. □

Žitnik, S.; Jankovič, M.; Petrovič, K.; Bajec, M. (2016). Architecture of standard-based, interoperable and extensible IoT

platform, in *Proceedings of papers of 24th Telecommunications Forum (TELFOR)*, Belgrade, Serbia, 22.-23. November 2016, pp. 1-4. □

Drašković, D.; Gencel, V.; Žitnik, S.; Bajec, M.; Nikolić, B. (2016). A software agent for social networks using natural language processing techniques, in *Proceedings of papers of 24th Telecommunications Forum (TELFOR)*, Belgrade, Serbia, 22.-23. November 2016, pp. 1-4. □

Janković, M.; Žitnik, S.; Bajec, M. (2016). MyDataHub, in *Workshop Electronic and Mobile Health: Proceedings of the 19th International Multiconference Information Society*, Ljubljana, Slovenia, 10.-11. October 2016, pp. 50-51. □

Žitnik S., Bajec M. (2015). Iterative joint extraction of entities, relationships and coreferences from text sources, in *Proceedings of the RCIS 2015*, Athens, Greece, 12.-15. May 2015, pp. 1-11. □

Furlan B., Žitnik S., Nikolić B. and Bajec M. (2013). The Role of Semantic Similarity for Intelligent Question Routing, in *Proceedings of the International Scientific Conference on Informatics 2013*, Spišská Nová Ves, Slovakia, 5.-7. November 2013, pp. 1-6. □

Žitnik S. and Bajec M. (2013). Token- and constituent-based linear chain CRF with SVM for named entity recognition, in *Proceedings of the Fourth BioCreative Challenge Evaluation Workshop*, Washington DC, USA, 8 October 2013, vol. 2, pp. 144-151. □ **Achieved rank 7 among 27 participating teams in extracting chemical compounds from text.**

Pavković M., et al. (2013). Intelligent techniques for searching Internet forums, in *Zbornik dvaindvajsete mednarodne Elektrotehniške in računalniške konference ERK 2013*, Portorož, Slovenija, 16.-18. september 2013, pp. 85-88. □

Žitnik S., et al. (2013). Iterative end-to-end information extraction based on linear models, in *Zbornik dvaindvajsete mednarodne Elektrotehniške in računalniške konference ERK 2013*, Portorož, Slovenija, 16.-18. september 2013, pp. 47-50. □

Žitnik S., Žitnik M., Zupan B. and Bajec M. (2013). Extracting gene regulation networks using linear-chain conditional random fields and rules, in *BioNLP shared task 2013: ACL 2013 - Proceedings of the workshop*, Sofia, Bulgaria, 4.-9. August 2013, pp. 178-187. □ **First place in extracting relationships between genes and proteins.**

Lavbič D., Žitnik S., et al. (2012). Traversal and relations discovery among business entities and people using semantic web technologies and trust management, in *Databases and information systems VII : selected papers from the Tenth International Baltic Conference*, DB&IS 2012, pp. 164-174. □

Žitnik S., et al. (2012). Collective information extraction using first-order probabilistic models, in *BCI 2012 : proceedings*, Novi Sad, Serbia, 16.-19. September 2012, pp. 279-282. □

Žitnik S. (2012). Collective ontology-based information extraction using probabilistic graphical models, in *CAISE-DC-2012 : proceedings of CAiSE 2012 Doctoral Consortium*, Gdansk, Poland, June 26, 2012. □

Šubelj L., Žitnik S., et al. (2012). Lastnosti velikih realnih omrežij in primeri uporabe, in *Ustvarimo nove rešitve!: zbornik prispevkov*, 19. konferenca Dnevi slovenske informatike, Portorož, 16.-18. april 2012, pp. 1-10. □

Zrnec A., et al. (2012). Podatkovne baze NOSQL, in *Ustvarimo nove rešitve!: zbornik prispevkov*, 19. konferenca Dnevi slovenske informatike, Portorož, 16.-18. april 2012, pp. 1-10. □

Bajec M., et al. (2011). Supervizor+, in *Informatika v luči reforme javne uprave: zbornik konference Informatika v javni upravi*, 28.-29. november 2011, Kongresni center Brdo pri Kranju, pp. 1-6. □

Žitnik S., et al. (2010). Trendi v inteligentnem pridobivanju podatkov iz heterogenih virov, in *Zbornik Konference Informatika v Javni Upravi 2010*. □

Šubelj L., et al. (2010). Analiza kompleksnih omrežij: Osnovni pojmi in primeri uporabe v praksi, in *Zbornik Konference Informatika v Javni Upravi 2010*. □

Corpora

Žitnik S. (2019). Slovene corpus for aspect-based sentiment analysis - SentiCoref 1.0, available at *CLARIN.SI data & tools*. □

Žitnik S. (2018). Slovene coreference resolution corpus coref149, available at *CLARIN.SI data & tools*. □