

Opis delovnega mesta mladega raziskovalca/ke (*Description of the Young Researcher's position*)

1. Članica UL (*UL member*):

Fakulteta za računalništvo in informatiko (*Faculty of Computer and Information Science*)

2. Ime, priimek in elektronski naslov mentorja/ice (*Mentor's name, surname and email*):

Tomaž Curk, tomaz.curk@fri.uni-lj.si

3. Raziskovalno področje (*Research field*):

Bioinformatika, Umetna inteligenca

Klasifikacija ARRS:

2.07 Računalništvo in informatika

2.07.07 Inteligentni sistemi – programska oprema

4. Opis delovnega mesta mladega raziskovalca/ke (*Description of the Young Researcher's position*):

Vključuje morebitne dodatne pogoje, ki jih mora izpolnjevati kandidat/ka za mladega raziskovalca/ko, ki niso navedeni v razpisu za mlade raziskovalce.

GLOBOKO UČENJE ZA ZLIVANJE HETEROGENIH, NEOZNAČENIH IN NESTRUKTURIRANIH BIOLOŠKIH PODATKOV

Mladi raziskovalec bo raziskoval na področju strojnega učenja in podatkovnega rudarjenja velikih količin heterogenih, neoznačenih in nestrukturiranih podatkov. Preučeval in razvijal bo nove metode globokega učenja za modeliranje in zlivanje heterogenih, neoznačenih in nestrukturiranih bioloških podatkov. Pristopi zlivanja podatkov omogočajo vključitev dodatnih heterogenih podatkov, ki so le posredno povezani s ciljno relacijo, kar navadno izboljša natančnost napovednega modela. Glavni izziv pristopov je pretvorba nestrukturiranih podatkov (slika, video, zaporedje, graf znanja, graf relacij med objekti) v predstavitev nizke dimenzionalnosti, ki jo je možno uporabiti za zlivanje. Globoke nevronske mreže so izredno uspešen pristop, saj med učenjem odkrivajo diskriminativne, nizko dimenzionalne latentne vzorce v podatkih. Mladi raziskovalec bo razvil metode za hkratno učenje vložitve bioloških podatkov in njihovega medsebojnega zlivanja, z namenom izboljšanja napovedne natančnosti naučenega modela in vpogleda v mehanizme delovanja bioloških sistemov.

Mladi raziskovalec bo vključen v Laboratorij za bioinformatiko UL FRI (<http://www.biolab.si>). Kandidat bo razvite metode ovrednotil v sodelovanju z našimi raziskovalnimi partnerji na bioinformatičnih domenah, predvsem na podatkih o interakcijah med proteini in RNA, o interakcijah med bakterijami in bakteriofagi ter na drugih domenah. Med usposabljanjem bo tako sodeloval z domačimi raziskovalnimi partnerji (NIB, IJS, UL BF) in imel možnost daljšega gostovanja v tujini: EMBL, Heidelberg, Nemčija; The Francis Crick Institute, London, UK; Baylor College of Medicine, Houston, ZDA in drugi.

Od kandidata pričakujemo odlično pisno in ustno sporočanje v angleškem jeziku, znanje programiranja in metod strojnega učenja, dobre temelje računalništva in biologije ter veselje do združevanja obeh področij.

DEEP LEARNING FOR DATA FUSION OF HETEROGENEOUS, UNLABELED, AND UNSTRUCTURED BIOLOGICAL DATA

The young researcher will be trained in machine learning and data mining of large heterogeneous, unlabeled, and unstructured data. The candidate will develop deep learning methods for modeling and fusion of heterogeneous and unstructured biological data. Data fusion methods allow the inclusion of heterogeneous side information, which improves the prediction of a target relation. The main challenge of these approaches is the conversion of unstructured data (images, videos, sequences, knowledge graphs, relations among objects) into a low-dimensional representation that can be used for data fusion. Deep neural networks are an extremely successful approach, as they discover discriminative, lower-dimensional latent patterns in data during learning. The young researcher will develop methods for concurrent data embedding and data fusion with the aim of improving the predictive accuracy of models and providing better insight into the mechanisms of biological systems.

The research will be conducted in the Bioinformatics Laboratory at UL FRI (<http://www.biolab.si>). In collaboration with our research partners, the methods will be evaluated on various bioinformatics domains, including protein-RNA interaction data, bacteria-bacteriophages interaction data, and other domains. During the training, the candidate will thus collaborate with our domestic research partners (NIB, IJS, UL BF) and will have the opportunity for extended research visits abroad: EMBL, Heidelberg, Germany; The Francis Crick Institute, London, UK; Baylor College of Medicine, Houston, USA, and others.

We expect the candidate to have excellent written and oral communication skills in English, programming and data mining experience, good foundations in computer science and biology, and enthusiasm for combining both fields.