

Kratek opis usposabljanja mladega raziskovalca (*Short description of the Young Researcher's training*)

1. Raziskovalna organizacija (*Research organisation*):

0510-0381 Univerza v Ljubljani, Medicinska fakulteta
University of Ljubljana, Faculty of Medicine

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

Andrej Kastrin, andrej.kastrin@mf.uni-lj.si

3. Šifra in naziv raziskovalnega področja (*Research field*):

3.08 Javno zdravstvo (varstvo pri delu)
Public health (occupational safety)

4. Kratek opis usposabljanja mladega raziskovalca (*Short description of the Young Researcher's training*):

slo:

Pomembno področje delovanja raziskovalne skupine je razvoj metodologije za odkrivanje zakonitosti iz nestrukturiranih (biomedicinskih) podatkovij. Ukvarjali smo se predvsem s samodejnim luščenjem informacij iz bibliografskih zapisov in njihovo sistematično transformacijo v strukturirane oblike znanja. Pred kratkim pa smo v poglobljeni analizi pokazali, da bo za prebojne raziskovalne dosežke na širšem področju znanosti o življenju nujno potrebna integrativna analiza heterogenih podatkovij (npr. genomskih, proteomskih, metabolomskih, bibliomskih).

Usposabljanje mladega raziskovalca (m/ž) (v nadaljevanju: MR) po potekalo na interdisciplinarnem področju podatkovnih znanosti in tehnologij znanja, s poudarkom na analitiki klasičnih omških podatkovij in rudarjenju besedil. MR se bo vključil v delo programske skupine *Metodologija za analizo podatkov v medicini*. Usposabljanje bo MR začel s poglobljenim študijem izbranih tem z domenskega področja, vključno z analizo kompleksnih omrežij, reprezentacijskim učenjem in statistično podprtim odkrivanjem zakonitosti iz (ne)strukturiranih podatkov. Hkrati bo potekala identifikacija ožjega področja dela doktorske naloge. Sledil bo poglobljen pregled literature, razvoj metodologije in analiza podatkov. Validacija razvite metodologije bo najverjetneje potekala na področju odkrivanja in pojasnjevanja novega znanja na področju raziskovanja mikrobioma. Predvidevamo, da bo MR s področja obravnavanega problema objavil dva izvirna znanstvena prispevka. V okviru usposabljanja bo MR spoznal delo sorodnih raziskovalnih skupin, s katerimi tudi sicer intenzivno sodelujemo.

Od kandidata pričakujemo visoko stopnjo motiviranosti in delovne vneme. Kandidat naj izpolnjuje pogoje za vpis na doktorski študijski program *Računalništvo in informatika* (UL FRI), *Matematika in fizika* (UL FMF) ali *Interdisciplinarni doktorski študijski program Statistika* (UL). Po dogovoru z mentorjem je možen tudi vpis na drug, a vsebinsko podoben študijski program. Zaželeno je solidno znanje angleščine.

eng:

An important area of activity of the research group is the development of a methodology for knowledge discovery from unstructured (biomedical) data. We have mainly focused on automatic information extraction from literature data and its systematic transformation into structured forms of knowledge. Recently, we have shown that breakthrough scientific achievements in the life sciences require integrative analysis of heterogeneous data (e.g. genomic, proteomic, metabolomic, bibliographic).

The PhD candidate (m/f) will be trained in the interdisciplinary field of data science and knowledge technologies, with a focus on the analysis of classical omics data and text mining. The PhD candidate (m/f) will be involved in the research activities of the research program *Methodology for Data Analysis in Medical Sciences*. The training will start with an in-depth study of state-of-the-art topics in the field, including analysis of complex networks, representation learning, and statistically based knowledge discovery from (un)structured data. Next, we will define specific goals of the dissertation. Computational validation of the developed methodology will most likely be performed in the field of microbiome research. We expect the PhD candidate to prepare and publish two original scientific papers. As part of the training, the PhD candidate will work closely with our recognized research partners.

We expect the candidate to be highly motivated and with high level of grit. The candidate should meet the requirements for admission to the PhD program in *Computer and Information Science* (UL FRI), *Mathematics and Physics* (UL FMF), or the *Interdisciplinary Doctoral Study Programme in Statistics* (UL). In consultation with the supervisor, enrollment in another program with similar content is also possible. Solid knowledge of the English language is desirable.