

Laboratory of Artificial Intelligence, Institut Jožef Stefan seeking motivated students for work on European projects

Jožef Stefan Institute (www.ijs.si) is the leading research institution for natural sciences in Slovenia with over 900 researchers within 25 departments working in the areas of computer science, physics, and chemistry and biology.

The Artificial Intelligence Laboratory (<http://ailab.ijs.si/>), with approximately 40 researchers, is one of the largest European research groups working in the areas of machine learning, data mining, language technologies, semantic technologies and sensor networks. Our key research direction is combining modern statistical data analytic techniques with more semantic/logic based knowledge representations and reasoning techniques with the purpose to make progress in solving complex problems such as text understanding, large scale probabilistic reasoning, building broad coverage knowledge bases, and dealing with scale.

Our research team is dynamic, young and international and we are currently looking to expand by offering student positions to bright, motivated and driven individuals who are interested in working on innovative European projects.

If you are interested in joining our team, please send your CV along with a brief letter of motivation in either English or Slovene regarding your contribution to one or more of the projects described below to:

Polona Škraba Stanič - polona.skraba@ijs.si (Subject: Student work/šudentsko delo)

1. Project: X5gon (www.x5gon.org)

Description: Within this project, our team is developing a platform that will connect different Open Education Resource (OER) repositories across the world. The connection is established through different channels:

- Indexing - accessible resources are indexed, OERs are preprocessed and mapped into common semantic space, metadata is stored in the database,
- User-activity tracking - users are tracked when accessing a partner OER repository, and
- Recommendations - using indexed OERs and tracking data, we aim to provide personalized recommendations that are embedded into partner OER repository webpage.

We are looking for a student who would contribute to building Recommendations, which includes developing a recommender engine, designing evaluation strategy, and performing the evaluation.

2. Project: DataBench (www.databench.eu)

Description: The DataBench project is focused on the assessment of Big Data benchmarking tools and projects:

- In particular, our team is working on **modelling relationships between metrics, data and project methodologies**. Metrics and datasets, used within the benchmarking should be associated with the tasks and problems to be solved with the data. These will provide a set of indicators to measure the hardness of how the data is being used in terms of selection of project infrastructure, methods, tools, user interfacing and legal challenges. The output should be an approximate analytic mapping between the data characteristics and methodology being used in the observed projects.
- Another task is the **development of a specific database and REST API** (in order to search and retrieve information from the database) for project purposes.
- Developing **front-end** for user interaction.

In the course of the project, students will become acquainted with techniques, methods and specific libraries for modelling and analytics and get experience in back-end and front-end development.

3. Project: enviroLENS

Description: We are developing text-mining/reasoning services on top of corpora of environmental law documents and existing ontologies. The tasks to be done within the project are:

- Retrieval of existing law data and ontologies and basic analysis of the data.
- Existing ontology enrichment (based on InfoMiner analysis or something else).
- Relating ontology to Wikipedia.
- Text segmentation task (detecting relevant paragraphs from the text using active learning); to be implemented within InfoMiner, a system developed by our team for semi-automatic document analysis
- Rule Extraction based on Semantic Grammar Construction (<http://www2015.wwwconference.org/documents/proceedings/companion/p673.pdf>)
- Provision of Remote Sensing based ontology (describing capabilities of remote sensing).
- Development of “cross-ontology relations” (between law related ontology and remote sensing based ontology).

A student would be working on the above tasks and gain experience in language and semantic technologies.

4. Project: PerceptiveSentinel (<http://www.perceptivesentinel.eu/>)

Description: The aim of the project is to provide Big Data analysis of crop classification and yield predictions for 3 use cases. We are to develop:

- Stream mining methods for classification and regression, integrate them in Python (scikit-learn interface) and QMiner (<http://qminer.ijs.si/>, a data analytics platform developed by our team for processing large-scale real-time streams containing structured and unstructured data); show that we can achieve similar results with much faster algorithms than are currently available.
- Weather impact on classification and yield information; data fusion.
- Implement several state-of-the-art methodologies and implement them into a common Python library called EO-learn:
 - Feature extraction (on a stream, on time-series)
 - Feature fusion
 - Feature selection
- Contribute significantly to the EO-learn platform.

The student will have the opportunity to contribute in a wide variety of possibilities such as Deep learning architectures for learning from satellite imagery and tackling SAR (radar) data possibilities.

5. Project: EW-Shopp (www.ew-shopp.eu)

Description: The core task of the project is to support eCommerce businesses with analytics based on linking their data with external data sources - primarily pertaining to events and weather. Since the project is in its final year, we already have a somewhat mature toolkit codebase (Python and JavaScript) and most of the work focuses on implementing machine learning pipelines for individual business partners. This includes:

- Consuming business data and preprocessing it into a form suitable for analytics. This phase also includes ensuring it is correctly linked to appropriate weather data (e.g. based on store location) and events (e.g. relevant news articles based on topic).
- Formulating a prediction task, building a model using machine learning methodology and evaluating it. For example, setting up the model for predicting the increase in how many people in Berlin visit football-related websites before a big match to support online advertising or in what weather conditions the sales of air-conditioning units increases.
- Helping business partners incorporate the developed pipeline into their platforms and business processes. The partners include eCommerce businesses (online stores, price comparison engines, advertising agencies...) from both Slovenia as well as abroad.

A student would work on different aspects of this process and would gain experience with real-world analytics as well as deployment of such systems in business environments.