COMP 6961 Graduate Seminar Abstracts

Summarized by: Rahul Patil (40166394)

December 16, 2021

Seminar 1: Reducing the length of field-replay based load testing

Original presenter: Yuanjie Xia

Abstract by Rahul Patil

In today's modern culture, software quality has become a crucial part of any software process in leading sectors. Designing load test cases has become an important aspect of the software design process in order to test a realistic workload of a particular software, which is monitored by a broader term of SQA (Software Quality Assurance). Replacing load testing is a time-consuming operation that requires a lot of effort. As a result, it is not practicable in today's software development process, which necessitates time-saving optimization. Load testing ensures that the software system can deliver reliable service under particular conditions. As a result, designing realistic workloads that can replicate the actual burden in the field is one of the most typical issues of load testing. A proposal outlines an automatic method for lowering the cost of field-replay based load testing by classifying workloads, combining time periods with similar workloads, and analysing stability. As a result, performance vector sets resemble clusters of workloads with similar time periods, reflecting percentages of CPU use. The distribution of these sets and the stability of clusters can be determined using statistical insights gleaned from this data. The results from the analysis can significantly reduce workloads times. The load testing results obtained by replaying the workloads after reduction show a high degree of correlation that is similar to the original set of workloads. Research work done here can be used by practitioners to undertake realistic field-replay based load testing while saving resources and time.

Seminar 2: Implementing a Library to Calculate Surrogate Safety

Original presenter: Mohammad Hossein Nazemi

Abstract by Rahul Patil

There are numerous things in life of which safety is of paramount importance to any living being. The Seminar focuses on concepts of surrogate safety analysis, approaches, shortcomings, and results which duly create an impact on today's swift world. The data sourced from Traffic Safety Analysis showed an increase in injuries, casualties, and number of crashes by more than 20% in the last decade. The analysis based on crash data may not include applicable data, the before and after state of the crash, behavior of road users and numerous other attributes which could become actual cause of the happenings. Thus, PET (Post Encroachment Time) and Speed are used as parameters for analysis which can result in crossing times of two vehicles closing by on an encroachment zone, and speed module can be used to monitor the severity of PET. As a result, Automated PET Detection and calculation module can be used for precise continuous PET values. The library is usable with different classification algorithms in different solutions. Primitive precision particularly uses bounding boxes to detect conflicts and not the tracks which are especially useful for long vehicles and do not miss any value. Speed module uses moving average technique to remove noise from solution, considering last 4 frames. Real time analysis from blue city data with the annotation software resulted best after applying the filters reducing to actual PET's. The research conducted can help uncover patterns about traffic safety and indulge in action plans accordingly.

Seminar 3: A Recommender System for Scientific Datasets and Analysis pipelines

Original presenter: Mandana Mazaheri

Abstract by Rahul Patil

Open Science is all about making research transparent and accessible for public use and is achievable with FAIR principles in effect. As users generally would like to get the best matching datasets/pipelines for analysis in relevant projects. The research work done here examines the viability of a collaborative filtering system like provenance-based recommender system for scientific pipelines and analysis evaluated on datasets from Canadian Open Neuroscience Platform (CONP). Recommender systems in general can have input data with explicit as well as implicit feedback and approaches for the workflow based on either content based or collaborative filtering. As CONP was feasible with experiments using Boutiques framework, thus the framework was used to install, validate, publish, and execute pipelines. Provenance-based system duly captures low level technical features of a pipeline dataset, like the degree of pre-processing involved which can be easily overlooked by domain experts. Thesis concludes that provenance-based pipeline and dataset recommenders are possible and advantageous for open-science resource sharing and utilisation. The collecting of more detailed provenance traces, as well as the deployment of the system in production, will be the focus of future effort. Pipelines can be utilised in a variety of ways depending on how they are parametrized. In the provenance records, several parametrizations might be found and proposed for individual datasets. Furthermore, datasets frequently contain many sub-parts or formats relating to variable subjects or data. A recommender system might be built to provide analyses for such attributes, resulting in more precise recommendations.