Zero-Knowledge Systematization of Knowledge:

Getting Blockchain Ready for Quantum Computers

dr. Cecilia Boschini, Prof. Stefan Wolf

Abstract. Zero-knowledge proofs are protocols that allow a prover to convince a verifier that she has some secret information that satisfies some public requirement without revealing the secret itself (e.g., to prove that she is of age without revealing how old she is). Research in this area has been thriving in recent years, due to its applications in privacy-preserving technologies and in blockchain. However, the incumbent threat posed by quantum computers requires to switch gear: as many of the current protocols are not secure against large universal quantum computers, applications need to be ready to switch to newer protocols, so-called post-quantum, that are secure even in case of a quantum attack. This project aims to survey the current usage of zero-knowledge proofs in blockchains, with a particular focus on Zcash (https://z.cash/), and to assess vulnerability to quantum attacks. The final goal would be to identify requirements for post-quantum alternatives in terms of length of produced tokens and efficiency.

Supervision. The student will work alongside dr. Cecilia Boschini, a Postdoc researcher in the Cryptography and Quantum Information group of Prof. Stefan Wolf. She has been doing research on post-quantum cryptography both in the academic and in the company environment (previously to USI, she was employed at IBM Research - Zurich). Her knowledge of zero-knowledge proofs and of post-quantum cryptography will help the student getting quickly comfortable with the topic.

Requirements. This project would be a good introduction to a research topic in cryptography that has gained a lot of attention lately, due to its wide-spread applications. As such, familiarity with cryptography, security and blockchain at the time of starting the project is required.

Quantum-safe Messaging: Security Assessment of the Signal Messaging App

Dr. Cecilia Boschini, Prof. Stefan Wolf

Abstract. Secure messaging relies on cryptography to guarantee both the integrity and the confidentiality of messages. The security of such building blocks comes from so called computational assumptions: they are secure if some mathematical problems are hard to solve. So far, this has been true: even the largest supercomputers would take decades before breaking such problems. However, the introduction of quantum computers has changed the paradigm: with universal, large enough quantum computer it will be possible to solve such problems in a short amount of time, thus breaking the security of the schemes. Quantum-safe alternatives (so-called *post-quantum*) are already available, and basic post-quantum cryptographic schemes are already in the process of being standardized. What is needed to use them in practice are security assessments: which part of the architecture of secure messaging apps is vulnerable, and what would it be necessary to fix it? The goal of this project is to analyze the security of a particular messaging app, Signal (`https://www.signal.org/`). Signal was chosen because of its security (arguably the best model among existing messaging apps), openness (the code and the architecture are open-source and peer-reviewed), and its commitment to an ethical and sustainable approach to security.

Supervision. The student will work alongside dr. Cecilia Boschini, a Postdoc researcher in the Cryptography and Quantum Information group of Prof. Stefan Wolf. She has been doing research on post-quantum cryptography both in the academic and in the company environment (previously to USI, she was employed at IBM Research - Zurich). Her knowledge of the Signal protocol and of post-quantum cryptography will help the student getting quickly comfortable with the topic.

Requirements. This project would be a good introduction to understanding the security of complex security architectures and to security assessments. As such, familiarity with cryptography/security (at least at a high-level) would be necessary at the time of starting the project.

Code transformation engine for the high-performance computing library *utopia.*

Abstract

Utopia is an **open source *C++* library** (bitbucket.org/zulianp/utopia) developed at ICS-USI in collaboration with CSCS (Swiss National Supercomputing Centre). The goal of this library is to provide hardware portable (CPU/GPU) codes for parallel algebra which can be employed for **scientific simulations**, **optimization problems**, and **machine learning algorithms**. Utopia is currently being used in several PASC (pasc-ch.org) projects such as FASTER, AV-Flow, and StagBL.

The main goal of this project is to implement a system for performing code transformations from small python code snippets to C/C++ code. This generated code is compiled just in time, and will run on acceleration hardware such as multicore CPUs and GPUs. This will allow utopia-python users to extend functionalities without the need of diving into complex C++ code nor sacrificing performance.

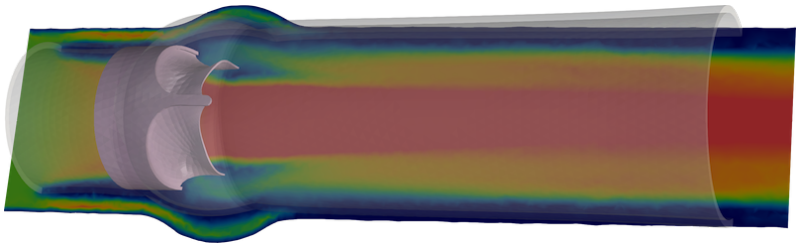
The project focus will be on ***Python* to C/C++ automated translation, code transformation, and just in time compilation.**

During the course of this project the student will:
- work independently on specific development goals;
- research the state of the art in scientific software;
- participate in developing open-source community codes;
- interact with developers and users;
- learn and use basic concepts of *C++;*
- use the GIT version control system, CMake/Makefile build systems, and Docker containers;
- implement a simple simulation tool using the developed tools.

Prerequisites

Basic knowledge of *C/C++* and *Python*.

Advisors: Prof. Dr. Rolf Krause, Dr. Patrick Zulian, Alena Kopanicakova



Simulation of a heart valve (AV-Flow)

GANs for solving PDEs with random data

Contact: Prof. Michael Multerer (multem@usi.ch) and Wei Huang (huangw@usi.ch)

In the last decade, the interest in the relationship between deep neural networks and partial differential equations (PDEs) has arisen among the communities of numerical analysis and machine learning. For example, a discrete residual neural network can be seen as a time discrete approximation to an ordinary differential equation in the continuous setting, see [1]. Vice versa, it is reasonable to consider the function space of neural networks as an ansatz space for approximating solutions to PDEs. Many PDEs have recently been solved with the help of neural networks, such as the Poisson equation, the Burgers' equation, and the Navier-Stokes equation, see [2] and [3]. As numerical guys active in the domain of uncertainty qualification, we are interested in solving PDE's with uncertain input parameters, like diffusivities or computational shapes which are subject to measurement imprecisions and missing data. The recently very popular generative adversarial networks (GANs) show the potential to solve PDEs with uncertain inputs with impressive performance due to its inherent stochastic property, see [3] and [4]. During this summer project, you may investigate all kinds of GANs to learn probability distributions from data which is governed by physical laws in the form of a PDE with random data and finally solve the forward and inverse stochastic problems.
We are open-minded when it comes to the concrete design of the project and any idea is welcome. Access to the ICS cluster's GPU resources for your experiments will be provided. You may extend this project to your master thesis and possibly to a scientific publication at a workshop or conference of machine learning. If you want to employ machine learning to crack mathematical problems in the physical world, please do not hesitate to contact us by e-mail for more details.

Prerequisites:

- Knowledge of programming, e.g., Python, Tensorflow or PyTorch.
- Knowledge of fundamental deep learning technique, GANs are preferred.
- Knowledge of PDEs is not necessary but highly appreciated.

[1] Chen, Rubanova, Bettencourt, Duvenaud "Neural Ordinary Differential Equations", Dec 2018.

[2] Zongyi Li, et. all "Fourier Neural Operator for Parametric Partial Differential Equations", Oct 2020.

[3] Raissi, Perdikaris, and Karniadakis, "Physics Informed Deep Learning (Part II): Data-driven Discovery of Nonlinear Partial Differential Equations", Nov 2017.

[4] Yang, Zhang, and Karniadakis, "Physics-Informed Generative Adversarial Networks for Stochastic Differential Equations", Nov 2018.

[5] Yang, et. all "Highly-scalable, physics-informed GANs for learning solutions of stochastic PDEs", Oct 2019.

Robust Sleep Detection Using Mobile and Wearable Devices

UROP project proposal at the Università della Svizzera Italiana (USI), Lugano. Switzerland

Background

A growing number of pervasive systems integrate sleep detection capabilities with the aim of improving their user's health and overall well-being. Sleep detection systems collect data from sensors embedded in personal devices such as, e.g., smartphones, wristbands, and other. The collected data are then analysed by applying signal processing and machine learning techniques to recognize users' sleep and wake stages. A system able to recognize such stages could enable several applications for self-assessment and self-management of sleep patterns on a day-to-day basis. It could for instance be used to provide feedback to their user regarding sleep (i.e., the total number of sleep hours), which in turn could help to diagnose sleep disorders. A system that infers sleep could further adapt its behaviour by, for instance, turning notifications of personal devices off when the user is sleeping or deciding to send a notification when it is the best time to wake the user up.

Sleep is traditionally studied in clinical environments using dedicated medical devices such as, e.g., polysomnography (PSG) and actigraphy. While these techniques are highly accurate (e.g., PSG) and appropriate to be used in a laboratory setting, they are very cumbersome to be used in naturalistic settings. Recent technological advancements have allowed researchers to propose several approaches for unobtrusive recognition of sleep outside clinical environments, in real-world settings. Existing home-based approaches can be grouped into *wearable* – such as, e.g., wristbands or chest bands – and *non-wearable* devices, such as, e.g., camera, radar signals or radio frequency. Non-wearable methods raise privacy concerns and are often immobile, which makes it difficult to monitor participants beyond their home-location. Wearable approaches, on the other hand, provide a mobile solution with limited privacy concerns and are able to distinguish between multiple people in one bed. Although several ready-to-use wearable-based applications are available in the market, their accuracy and validity towards gold-standard approaches are still open challenges. In this project we plan to address this challenge by developing a robust sleep detection method.

In particular the student will be asked to (1) apply standard signal processing and sensor-specific methods to prepare the data set (e.g., filtering, segmentation, interpolation and imputation of missing data), (2) apply supervised machine learning algorithms (e.g., CNN, LSTM) to identify users' sleep stages and (3) report the performance of the models using appropriate metrics and visualization techniques.

Required skills and knowledge

- Good knowledge of data analysis and machine learning techniques (e.g., CNN, LSTM, transfer learning, data augmentation).
- Good knowledge of Python programming (i.e., using Jupyter notebook)

Expected outcomes

- Implement a machine learning pipeline to automatically recognize sleep and wake stages.
- Pre-train a model on existing public data sets and fine-tune it on a new data set.

Supervisors and contact information: Shkurta Gashi: shkurta.gashi@usi.ch, Elena Di Lascio: elena.di.lascio@usi.ch, Prof. Dr. Silvia Santini, silvia.santini@usi.ch.

Personal Informatics System Using Earbuds: Feedback Component

UROP project proposal at the Università della Svizzera Italiana (USI), Lugano. Switzerland

Background

Mobile and wearable devices such as, e.g., smartphones and smartwatches, enable the continuous and unobtrusive collection of sensor data like, e.g., physical activity or heart rate. Adequate application of signal processing and machine learning techniques on this data allows deriving insights regarding user's behaviour in everyday life. Knowledge of peoples' behaviour allows advancing the personalization and effectiveness of applications that aim to promote users' health and well-being. The class of systems that help people collect and reflect on personal information are known as *personal informatics systems*.

One of the core features of personal informatics systems is the *feedback component*, which refers to the exploration of feedback mechanisms to be provided to the user with the goal of enhancing their health, performance and well-being. Some of the open challenges of developing the feedback component are *when* and *how* to provide the feedback to the user as well as *what* kind of feedback (i.e., whether a visualization metaphor can be applied). Building upon existing research, the goal of this project is to implement the feedback component of an existing personal informatics system that automatically detects users' facial expressions and head gestures using unobtrusive, ear-mounted devices.

Facial expressions and head gestures that people make reveal insightful information about their affect, empathy, engagement or boredom during human-to-human interactions. Systems able to recognize such activities have thus been investigated as means to support and improve both in-person and computer-mediated interactions. In remote teaching settings, for instance, the amount of nodding and yawning activities could hint at a general attentiveness or boredom of the students and thus trigger a corresponding feedback to the teacher.

The aim of this project is to extend an existing mobile application by (1) embedding an already implemented model that recognizes users' facial expressions and head gestures using data from earbuds – such as accelerometer and gyroscope – and (2) adding visualization mechanisms to provide feedback to the user (i.e., plots on mobile application, notifications on continuous detection of negative behaviour).

Required skills and knowledge

- Good knowledge of mobile application programming (using Android or a cross-platform programming framework like, e.g., Flutter)
- Good knowledge of technologies for data visualization and feedback.
- Basic knowledge of data analysis and machine learning techniques.

Expected outcomes

- Extend an existing personal informatics system by designing and implementing an Android/Flutter application that collects data from eSense earbuds and provides feedback to users in real-time.
- Embed an existing model into the system to automatically recognize user's head gestures and facial expressions.

Supervisors and contact information: Shkurta Gashi: shkurta.gashi@usi.ch, Elena Di Lascio: elena.di.lascio@usi.ch, Prof. Dr. Silvia Santini, silvia.santini@usi.ch.

Development of an Ethereum blockchain replayer

Supervisors:

- Dr. Pierre-Louis Roman <romanp@usi.ch>
- Prof. Patrick Eugster <eugstp@usi.ch>

Live blockchain systems like Bitcoin or Ethereum rely on a test network (testnet) to experiment with new features before they are rolled out in the main network (mainnet). In testnets, blocks are generated randomly and at a much higher pace than on the real network to help with fast feature prototyping. Testnets however do not enable prototypes to leverage the historical data stored in the blockchain since they generate random synthetic blocks. Features which results vary depending on the content of blocks therefore cannot treat the blockchain as an exploitable dataset.

One possible alternative to a testnet is a blockchain replayer. A replayer simply reenacts the creation of past blocks, possibly at a different pace and with modified content, and sends the replayed block to a blockchain node that includes the feature under test. Thanks to a replayer, the feature developer can observe the behavior of the modified node as if it was running live instead of in the face of synthetic blocks.

The goal of this internship is to develop a replayer for the Ethereum blockchain. Ethereum is host of many (if not too many) proposals for improvements [1] that could benefit from being evaluated in a replayed environment. The selected applicant will develop (likely in Go, up for discussion) an Ethereum replayer based on the official Ethereum node geth [2] and, if time permits, use it to evaluate select features proposed by Ethereum developers. The selected applicant is expected to have good programming skills while knowledge of the inner mechanisms of blockchains is a plus but is not a requirement.

This internship will provide the student with hands-on experience on software development on one of the most prominent blockchains and could see the replayer being adopted as a useful development tool by others. Software engineers with a background on blockchains is one of the top profiles researched on the job market.

[1] Ethereum improvement proposals: https://eips.ethereum.org
[2] Geth node: https://github.com/ethereum/go-ethereum

Routing algorithms and attacks on blockchain payment networks

Supervisors:

- Dr. Pierre-Louis Roman <romanp@usi.ch>
- Prof. Patrick Eugster <eugstp@usi.ch>

Existing blockchains have important and well-known limitations restricting their use. For one, transaction latency ranges from minutes to more than an hour for it to be safely recorded in the chain. Additionally, block size is usually capped (to hinder spamming) and is rapidly reached under high load. This saturation induces a competition between users to have their transaction included in a block which ultimately leads to an increase in transaction fees.

As a promising avenue to reduce contention on blockchains, developers and researchers are proposing off-chain solutions [1] (a.k.a. second layer solutions) that are designed to bypass the underlying blockchain (the first layer) as much as possible and use it only when needed. Included in these efforts are payment networks [2-4] that are composed of many one-to-one payment channels between individuals to form a global network capable of routing payments across channels. The payment routing algorithms used by these networks may only rely on local knowledge of the network/graph which may be vulnerable to attackers (by e.g., partitioning the network or creating sinks in the graph).

The goal of this internship is to study routing algorithms in blockchain payment systems, identify attack vectors in these algorithms and, if time permits, propose novel routing algorithms as potential solutions. The selected applicant is expected to have a good background in networks and graphs or a strong motivation to learn about these areas and about blockchains in general.

This internship will provide the student with exploratory research experience on a hot topic (second layer solutions) within a hot topic (blockchains). Software engineers with a background on blockchains is one of the top profiles researched on the job market.

[1] Gudgeon et al., "SoK: Layer-Two Blockchain Protocols", Financial Cryptography 2020: https://eprint.iacr.org/2019/360.pdf
[2] Sivaraman et al., "Routing Cryptocurrency with the Spider Network", HotNets 2018. Extended version: https://arxiv.org/pdf/1809.05088
[3] Bitcoin's lightning network: https://lightning.network
[4] Ethereum's raiden network: https://raiden.network

Programming Language for Reactive Distributed Monitoring

Supervisors:
- Dr Pavel Chuprikov <chuprp@usi.ch>
- Prof. Patrick Eugster <eugstp@usi.ch>

Existing network monitoring solutions primarily focus on querying information from the network rather than on the responses that would ensue these queries. Such approaches are limited both in presenting an incomplete abstraction, and in not taking advantage of the many optimization opportunities a holistic view on monitoring and management cycle could bring.

To overcome the limitations of the existing systems we focus on the design that combines monitoring and *management* in a single system. Our system aims at exploiting programmability of network devices to perform as many of the management actions as possible exactly at the point of data collection, i.e., at the switches, thus improving the management reaction time and saving the network capacity otherwise spent on collected data. As not all management decisions can be made locally, we introduce a distributed abstraction, where each monitoring and management task is represented by a set of interacting agents spread among several network devices. We call these agents *seeds*. When actually programming these seeds to perform the desired task, several performance factors must be taken into account: 1) switch resources are limited and some are already used by network control plane; 2) polling the data from the switch is a major bottleneck and must be under the system's control; 3) as many monitoring and management tasks can be active in the network simultaneously, the system have a good understanding of seed resource constraints so to utilize resources in the most efficient manner. To relieve the programmer (or a higher-level abstraction) from a burden of low-level resource management, we have designed a domain-specific programming language called *Almanac* and optimization algorithm for resource allocation and seed placement. Almanac features trigger variables for polling and timing, state-machine and messaging abstractions to simplify reasoning of a distributed system, and primitives for expressing placement requirements. At the same time the optimization algorithm takes into account the many resource and placement constraints from the seeds and from the switches to optimize overall utility of the monitoring and management tasks.

The selected student will work on automated translation from *Almanac* to the machine code accessing the existing API of our system gaining experience in programming language design and implementation.The translation would also need to include simple static analysis to produce the set of optimization constraints for the seed placement algorithm. The exact implementation framework is up to a discussion, but [1], [2], and [3] provide some examples.

[1] https://arzg.github.io/lang/
[2] https://www.stephendiehl.com/llvm/
[3] https://docs.racket-lang.org/guide/languages.html

Cost-based Mechanism Selection for Secure Cloud Computing

Supervisors:
- Dr Pavel Chuprikov <chuprp@usi.ch>
- Prof. Patrick Eugster <eugstp@usi.ch>

The value of what can be derived from customer data is being increasingly recognized by many industries, information is becoming the new currency. At the same, the amount of data has been growing exponentially, and many organizations have turned to the cloud in their search for cost-effective information processing. On the account of that, there is a huge demand for processing of sensitive data using third-party untrusted computational resources. While there are both software (homomorphic encryption) and hardware (secure enclaves) techniques with the potential to perform such processing without leaking any information, they have their own constraints and overheads, so that there is no single universal solution. For the best performance different techniques must be combined, but it quickly becomes hard to reason about end-to-end security for non-experts, which data analysts writing queries usually are not.

We have designed a system, called Hydra, that supports a multitude of security mechanisms nicely decoupling privacy policies from the business logic of the queries. To guarantee compliance with a chosen privacy policy, Hydra introduces a lambda-calculus-based domain specific language (DSL), whose type system tracks the privacy levels of program variables making sure that as the security mechanisms change no information leaks occur as different security mechanisms are being substituted. Hydra system is integrated with the Apache Spark streaming processor to provide users with the familiar query abstraction. The current limitation of Hydra is that the choice of the security mechanism is hardcoded, and while DSL is able to check compliance of any such choice, such flexibility is not fully exploited.

In the course of internship, the selected student will be working on the query optimization part of the Hydra (which uses standard Spark SQL extension points). The goal will be to use empirical performance measurements for different security mechanisms and Spark execution metrics in order to guide the mechanism choice at the query transformation phase in the fully automated fashion.

Evaluating the Use of Native Code in JVM Languages and Frameworks

The Java Virtual Machine (JVM) promotes the development of portable software, since applications are represented as platform-independent bytecode. However, the JVM also supports the integration of platform-specific native code, which does not have a corresponding bytecode representation. For example, many functions of the Java Development Kit (JDK) are implemented in native code, often to get access to otherwise unavailable lower-level functionality.

Unfortunately, a significant use of native code may result in suboptimal application performance. For example, the Java Native Interface (JNI), required to implement native code, may impose additional overhead when transitioning from bytecode to native code. Moreover, native code cannot benefit from the speculative optimizations done by the just-in-time (JIT) compiler based on profiling information collected at runtime. Furthermore, native code cannot be inlined in the caller by the JIT compiler, resulting in less effective optimizations.

This project aims at 1) assessing the contribution of native code during the execution of different applications running on the JVM and 2) detecting whether native code execution can significantly slow down application performance. Our focus is not limited to Java, but spans any language that can be run on the JVM, including Scala, Kotlin, and Clojure. Moreover, the projects consider frameworks built on top of the JVM, such as Spark, Flink, and Giraph. To achieve this goal, the student will be involved in several activities:

1. Design of a new methodology to characterize the execution of native code in applications running on the JVM.
2. Development of an efficient and accurate profiler that measures native-code execution.
3. Identification (or creation) of language-specific and framework-specific reference workloads to be used as benchmarks.
4. Analysis of the contribution of native code in the considered applications.
5. Detection of cases where native-code execution becomes a bottleneck that significantly impairs runtime performance.

The project is a unique opportunity for the student to deepen her/his knowledge in the domains of dynamic program analysis, managed runtimes (the JVM in particular) and empirical evaluation, all being important skills for a software engineer. The student will work side-by-side with the members of the Dynamic Analysis Group at USI, and will receive support in learning advanced topics that will strengthen her/his abilities as a computer scientist. Applicants interested in this project should be enrolled in the MSc program, have a good knowledge of JVM languages, C/C++ and UNIX-based operating systems, excellent programming skills, and deep interest in the field of dynamic analysis.

Advisors: Prof. Walter Binder, Dr. Andrea Rosà
Assistants: Matteo Basso, Filippo Schiavio

FACILE: Analyzing and enhancing text comprehension in the Italian language
UROP project proposal at the Università della Svizzera italiana (USI), Lugano, Switzerland

Reading and understanding text is a complex process that includes seeing the character (accessibility via visibility), being able to identify each character from the other distinctly (recognition via legibility) and finally making sense of the characters as a group (interpretation via readability).

This project seeks to develop a technological solution that improves accessibility providing systematic lexical simplification of text in the Italian language. It aims to help access content for people with intellectual disabilities, the elderly, nationals and immigrants with Italian as a second language, and all people in general. An example of a similar work can be find at the Figure below.

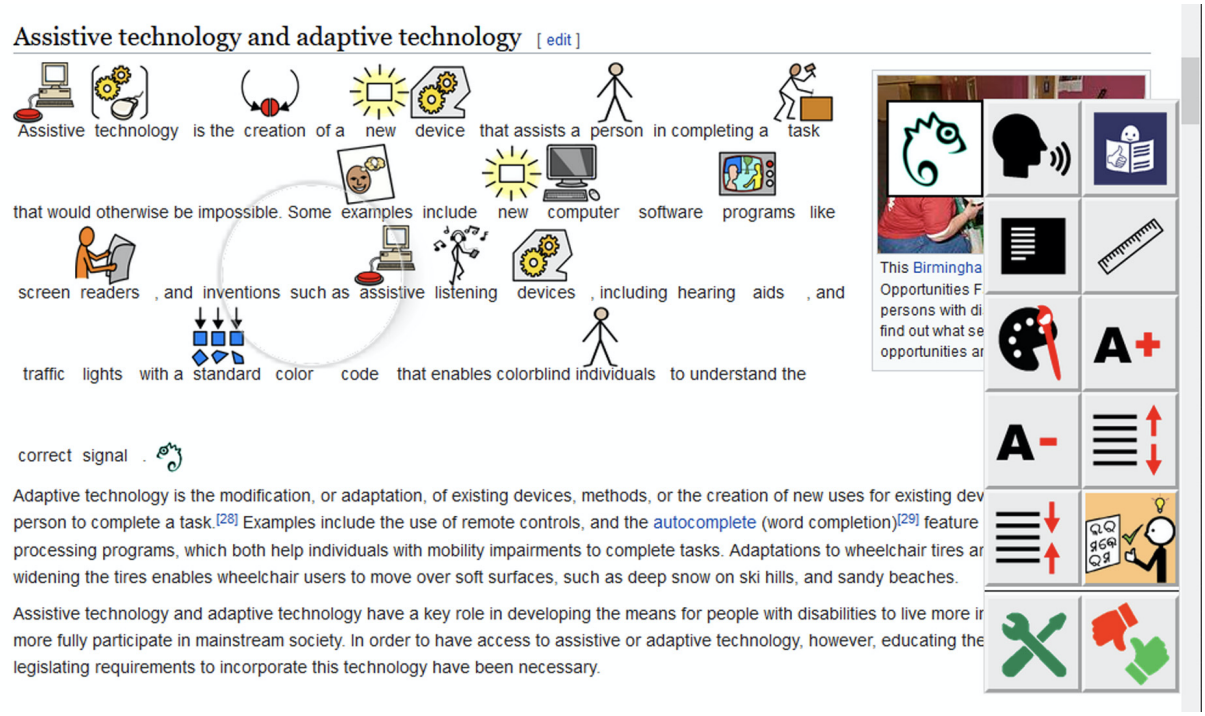


Figure 1: The Easy Reading graphical user interface (GUI) overlaid on a website, by Murillo-Morales et. al.

You will develop a web/mobile system with an accessible interface that increases cognitive access designed optimizing cognitive accessibility features following language guidelines, easy-to-read techniques, as well as WCAG 2.1. The system will also include symbols and the definitions of main words to help the users understand their meaning and context. Different services can be offered, such as synonyms.

You will be under the supervision of Dr. Monica Landoni and Leandro Guedes, a Doctoral assistant at USI. Collaborations can involve international researchers and Pro infirmis - Servizio di Lingua facile, located in Ticino.

+ACCESS: Designing accessible solutions for people with disabilities
UROP project proposal at the Università della Svizzera italiana (USI), Lugano, Switzerland

Digital accessibility and assistive tools have been used to enhance the experience of people with different abilities. Designing accessible solutions with and for people with disabilities makes a huge contribution to their life quality.

In this project, you will follow design steps, develop prototypes, and a final version of an application to be used before, during, and after a museum visit. The application will allow our users to listen and record audio files, see pictures and videos, and write important annotations.

You will personalize the user interface to fulfill essential accessibility guidelines and augmentative and alternative communication (AAC) techniques. You can see an example of a child interacting with an AAC application in the image below.

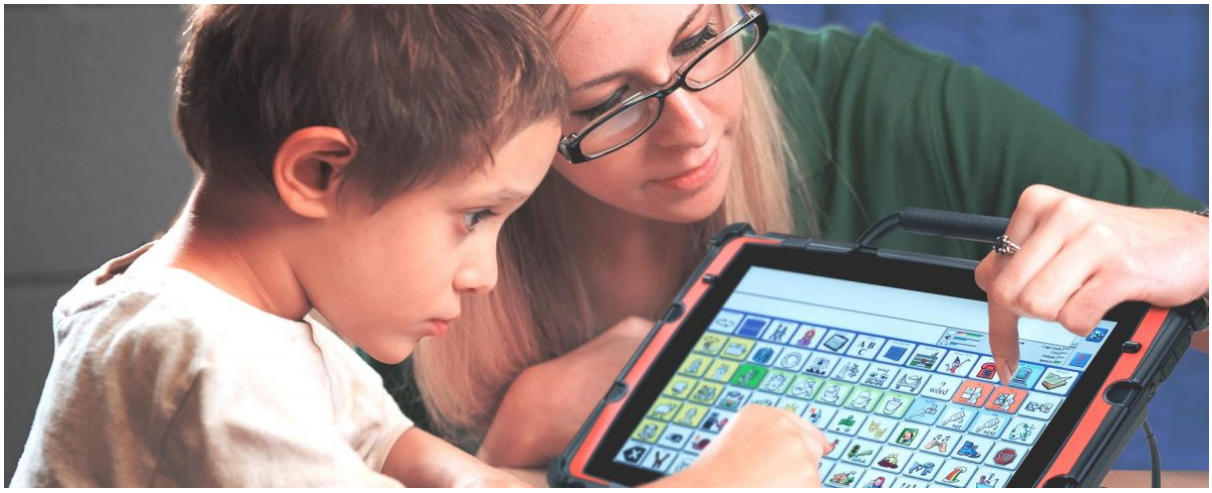


Image courtesy: http://teamworktherapies.com/?p=1970

In the future, your work will be assessed by members from Atgabbes, a local association of people with disabilities, focusing on users with mild cognitive impairments. Real local users will use your contribution.

You will be under the supervision of Dr. Monica Landoni and Leandro Guedes, a Doctoral assistant at USI. The work will be assessed at the local museum LAC, Lugano Arte e Cultura. Collaborations can involve international researchers and companies.

Are you interested in this project? Contact us if you have any questions, we will be happy to answer your questions.

Multisensory experiences at museums: mixing visual experiences with audio and tangible user interfaces

UROP project proposal at the Università della Svizzera italiana (USI), Lugano, Switzerland

Most of our everyday experiences are multisensory in nature. What we can see, hear, feel, taste, smell, and much more are part of our daily lives. Almost any experience you can imagine, such as walking at the park or watching content on your smartphone, involves a magnificent sensory world.

Multisensory experiences, where the senses meet technology, are available in multiple scenarios, including museums and exhibitions, where we can support multisensory experiences and enhance the user experience.

Image courtesy: https://alchetron.com/Tangible-user-interface

This summer project aims to use tangible user interfaces with visual and audio feedback to build a multimodal experience for users visiting an exhibition at a museum. It will help visitors make sense of the art and receive feedback that can enhance their experience.

You will be under the supervision of Dr. Monica Landoni and Leandro Guedes, a Doctoral assistant at USI. The work will be assessed at the local museum LAC, Lugano Arte e Cultura. Collaborations can involve international researchers and companies.

Are you interested in this project? Contact us if you have any questions, we will be happy to answer your questions.

DIES - Design and implementation of emoji enriched interfaces for search engines to help children searching in the classroom

UROP project proposal at the Università della Svizzera italiana (USI), Lugano, Switzerland

We have run an exploratory study to better understand children's ability to recognise relevant results when searching in the classroom. Teachers in two European schools sharing the same language assigned their students (ages 10 and 11) an online information discovery exercise about a history topic covered in class.
For this, children used a classic search interface and two novel ones enriched with emojis associated to relevant vs. irrelevant results. You can see the 3 mock-ups interfaces enriched with emojis as cues for relevant and non -relevant results in the Figure below.

(a) Light Bulb **(b) Tick/Cross** **(c) No Emoji**

At the end of the exercise, children filled out a post-task questionnaire meant to elicit their perception on usability of the interfaces.
We learned various lessons from our examination of children's search behaviour that will guide the design of future interfaces, including the fact that emoji-enriched interfaces result in significant improvement in terms of children identifying relevant resources.

Now we need your help to implement these emoji enriched interfaces and make them available online for future studies. Thus, if you are interested in Interactive Information Retrieval (IIR), while being an expert in Web development, and want to work on the future of Search Engines for children, this is the perfect project for your summer!

You will be part of an international team including IIR experts in Europe and US and an expert in education to provide us with valuable insights into how children use SE in the classroom.

Choose Your Own Adventure: design and implementation of a Web and/or Mobile Gamebook app for preschool children

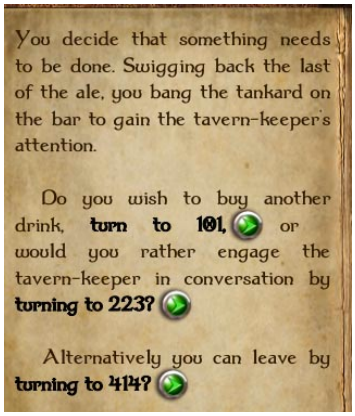
UROP project proposal at the Università della Svizzera italiana (USI), Lugano, Switzerland

Storytelling and narrative skills are among the building blocks that allow young children to learn how to read and write in their formative years. Together with an inter-disciplinary team composed of designers, preschool teachers and child psychologists, we are looking to design technology meant to help children aged 3-6 years old gain an interest in reading and telling stories, allowing them to develop their skills at home or in the classroom.

One of the avenues that we are researching is the design of an **Interactive Digital Gamebook Platform**: a game where children can participate in the stories by making choices that will influence the narrative and eventually lead to different possible endings.

Gamebooks were popular in the 1980s and 1990s, after which they were slowly replaced by other forms of interactive media such as videogames. However, they are now experiencing a resurgence as many digital gamebooks are being developed for mobile and eBook platforms.

Italian traditional gamebook for children (Amazon.it)


You decide that something needs to be done. Swigging back the last of the ale, you bang the tankard on the bar to gain the tavern-keeper's attention.

Do you wish to buy another drink, **turn to 101,** or would you rather engage the tavern-keeper in conversation by **turning to 223?**

Alternatively you can leave by **turning to 414?**

Gamebook on an Android platform

However, gamebooks are usually designed and developed for adults, while children are a special category of users who have specific needs and characteristics, for which designers and developers need to account.

Moreover, we want to go one step further and design a platform, a "container" in which we will be able to upload different books in different languages, so that we will be able to gather data from children and caregivers from all around the world.

If you are interested in:

- Learning how to gather requirements and tailor your software for special groups of users.
- Developing a web app and/or an Android app
- Learning how to conduct usability testing
- Gathering and analysing data from log files

This is the perfect project for you! You will work both independently and together with the team, develop your coding and designing skills, and participate in a real-world research project with real users and stakeholders!

Reading Buddy: design and implementation of a voice-activated ChatBot to support caregiver-child dialogic reading

UROP project proposal at the Università della Svizzera italiana (USI), Lugano, Switzerland

Storytelling and narrative skills are among the building blocks that allow young children to learn how to read and write in their formative years. Together with an inter-disciplinary team composed of designers, preschool teachers and child psychologists, we are looking to design technology meant to help children aged 3-6 years old gain an interest in reading and telling stories, allowing them to develop their skills at home or in the classroom.

One of the directions in which we are taking our research is the design of a **conversational agent to support caregiver-child dialogic reading**: a voice activated chatbot – for example, supported by Alexa or Google Assistant – to help parents and caregivers read more effectively with their children, allowing them to develop pre-reading and narrative skills.

Echo Show 8 (Amazon.it)

Dialogic reading is a proven and effective method to read **with** young children, instead of reading **to** them: the act of reading becomes a conversation instead of a monologue, engaging children and encouraging them to expand their use of language, improving their language and vocabulary skills.

There are many different types of prompts and sequences that can be used in dialogic reading – for example the **PEER** sequence consists in **P**rompting the child to answer a question, **E**valuating the response, **E**xpanding it and then encouraging the child to **R**epeat it.

However, many parents do not know these techniques, or they do not have the time or the language skills to practice dialogic reading with their children – for example, immigrant and refugees families in which the parents do not speak the language of the country in which they are living. Our aim is to level the playing field by providing each and every family with a way to help their children the best possible start in life, as developing their literacy and storytelling skills early will help them become better students in the future.

During this project, you will work together with a team of researchers, preschool teachers and education experts to design and implement a voice-activated chatbot that will interact with caregivers and children during a shared reading experience.

This is the ideal project for you if you:

- Have an interest in conversational agents and NLP (Natural Language Processing)
- Are a proficient Italian speaker
- Want to explore usability and engagement for a special group of users – children!

Defects4DL: A Database of Real Faults in DL Systems

Paolo Tonella, Gunel Jahangirova, Nargiz Humbatova
Software Institute@USI
paolo.tonella@usi.ch,gunel.jahangirova@usi.ch,nargiz.humbatova@usi.ch

Faults in Deep Learning Systems

Deep Learning (DL) models are becoming an integral component of systems performing complex, human-competitive tasks, such as automated driving, speech recognition and natural language processing. As such systems involve life, business or ethics critical activities, the types of faults that can occur in them are a crucial topic. However, the notion of a fault in DL systems is more complex than in traditional software, because the behaviour of these systems depends on how they are trained (i.e., the training data, the training process, the adopted model architecture, etc.) and is not programmed explicitly in the code. As a consequence, faults are not necessarily the logical faults that affect traditional software: they can be faults that affect any phase of the training process.

The existence of DL specific faults has been recognised by researchers and multiple studies performing classification and analysis of such faults have been performed. However, no database of real faults for DL systems has been created as extracting and reproducing such bugs is an expensive and challenging task.

The Goal of the Project

The overall goal of this project is to create a framework that provides an easy access to a set of reproducible and isolated real bugs collected from version control histories of various DL systems. A bug that will be a part of Defects4DL needs to satisfy the following conditions: 1) it is not in a DL framework (e.g. Keras), but in the DL system itself 2) it is not a common programming error 3) its root cause can be traced. Once these conditions are met, we need to obtain, for each bug, a faulty (V_{bug}) and a fixed (V_{fix}) versions that differ only by the bug fix. Moreover, as shown in Figure 1, we need to isolate the bug, which means that the bug fix should not include unrelated changes such as features or refactorings[1].

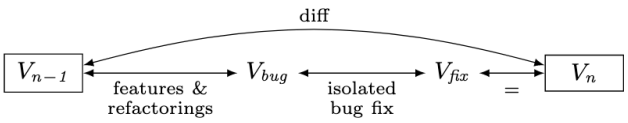


Figure 1: Versions V_{bug} and V_{fix} differ by only a bug fix. V_{n-1} and V_n represent the versions of two consecutive revisions in a project's version control history.

For each bug, the reproducibility needs to be ensured by including a test suite that contains at least one test case that exposes the bug – that is, the test case passes on the fixed but fails on the faulty version of the DL system. Figure 2 shows the information that Defects4DL will provide for each bug. Along with already mentioned V_{fix}, V_{bug} and a test case discriminating between them, a script that sets up all required configurations and dependencies will be also included.

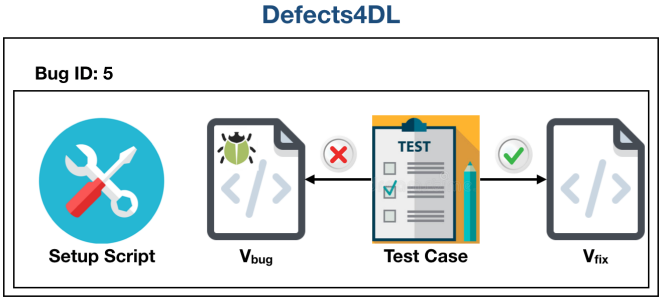


Figure 2: Information provided for each Bug by Defects4DL.

As the workload associated with setting up one single bug is not known, there is no fixed number of bugs that the project aims to collect. The subject DL systems the bugs of which will be mined are not decided yet and can be selected based on your interests.

Prerequisites

The only technical requirement to participate in this project is an average knowledge of the Python programming language and some familiarity with the Keras DL framework. Some experience with Git and test cases writing would also be helpful.

Why Choose This Project?

This project has a huge potential: if a good number of bugs will be collected, this dataset will change the course of DL testing research from using artificially seeded/adversarial faults to using real faults, thus making the field more representative of real-world practice. Hence, the natural follow up of this work, e.g., during the MSc thesis, could be a comparative assessment of DL test generators on the collected real faults. The results could be described in a scientific paper for possible presentation at major scientific venues.

In the frame of this project, you will get an invaluable experience, specifically, in debugging DL systems and repository mining, but more generally, in software development and testing. The results of this work will be made publicly available, i.e. you will contribute to an open source project. Moreover, you will be a part of the Precrime European project team:

https://www.pre-crime.eu

For any further questions, do not hesitate to contact us by email.

[1]Figure is from Just René, Darioush Jalali, and Michael D. Ernst. "Defects4J: A database of existing faults to enable controlled testing studies for Java programs", ISSTA 2014

Developing Self-Driving Car Autopilots using Reinforcement Learning Algorithms

Summer Project Proposal

Paolo Tonella, Matteo Biagiola, Andrea Stocco
Software Institute@USI
{paolo.tonella|matteo.biagiola|andrea.stocco}@usi.ch

Self-driving cars are nowadays a reality. Most major manufacturers including Tesla, GM, Ford, Mercedes, BMW, and Waymo/Google are building and actively testing different types of autonomous vehicles. Recent results show that autonomous cars have become very efficient in practice and have already driven millions of miles without any human intervention.

The key component of an autonomous vehicle is the perception module controlled by Deep Neural Networks (DNNs). DNNs take input from different sensors like camera, light detection, ranging sensor, and infrared sensor, and outputs the steering angle, braking and other commands necessary to operate the car safely.

Project Proposal

The goal of this project is to develop a self-driving car model that uses Reinforcement Learning (RL) algorithms, and compare its effectiveness with respect to more traditional Supervised Learning (SL) models (i.e., models based on DNNs).

To understand the differences between these two learning models we need to distinguish predictions from decision making. SL models typically generate predictions which can be used to make decisions. Moreover, SL models do not take into account that each decision influences future events, which in turn influence future decisions. RL, on the other hand, allows learning a *policy*, thereby creating models able to make their own decisions, take actions, react and adapt based on the feedback they receive.

Since the task of driving is sequential in nature, our hypothesis is that autonomous driving can be better addressed when formalized as a RL problem rather than a SL problem. No rigorous comparisons between SL and RL models for self-driving cars exist in the literature, and you will develop an open-source tool that will allow such comparison in a simulation environment.

Project Platform

The simulation platform that will be used for the project is the Donkey Car simulator (Figure 1). The simulator supports the creation, training, and testing of self-driving cars based on both technologies of interest for this project (i.e., RL and SL).

Tasks

For the realization of this project, the candidate is expected to perform a series of tasks, among which:

- familiarize with the Donkey Car simulator, and the self-driving car SL models;
- familiarize with RL-related concepts (i.e., reward, action, state) and algorithms (e.g., PPO, SAC, DQN);

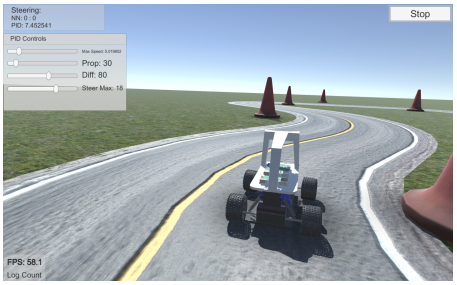


Figure 1: Donkey Car simulated environment.

- design an experiment to compare RL models trained with different RL algorithms with more traditional SL models in the Donkey Car simulator, and report the results.

Prerequisites

We are looking for a student which is passionate about the domain of self-driving cars and who is motivated to contribute to the project. No prior knowledge of the simulation platform or the Donkey Car is required, nor any prior knowledge of SL or RL, but willingness to learn these technologies and adapt them depending on the project's needs. During the project, the candidate can build the project upon an existing software infrastructure that works with RL/SL algorithms, and we will work closely with you and provide assistance when needed.

Why You Should Choose This Project

This project is unique in its multi-disciplinary nature: you will learn and use concepts in machine learning (ML), in particular SL and RL, software engineering and the ability to set up an experimental setting to rigorously compare different methods.

Upon successful completion of the project, you will have contributed to an open-source project that will allow developers to compare different self-driving car models based both on SL and RL. On your CV side, you will be able to show practical experience with training and above all evaluating ML models that have different characteristics.

Further Information

The proposed work is part of the Precrime (Self-assessment Oracles for Anticipatory Testing) ERC project. The interested student can find more information about Precrime on the project's website: https://www.pre-crime.eu/. Interested in this project? Any still-unanswered questions? Drop us an email now!

LüCE
Lugano Computing Education
Research Lab

UROP project proposal at the Software Institute of the Faculty of Informatics
Università della Svizzera Italiana (USI), Lugano, Switzerland

Learning Algebra with Programming Language Concepts

Advisor: Matthias Hauswirth and the Luce team

Motivation

Algebra is one of the harder subjects students learn in school. At Luce we are working on pedagogical tools that aim to make algebra a less daunting subject. We do this by reinterpreting middle school algebra as a pure functional programming language.

Goal

In this project you will use ideas from programming language theory to develop an interactive component for learning to work with algebraic expressions, adding the notational features known from middle school algebra, and you will integrate that component into our web-based educational platform.

Prerequisites

Excellent analytical and algorithmic thinking skills and ability to quickly familiarize yourself with new formalisms (such as formal syntax and semantics of programming languages).

Platform / Languages / Frameworks

In this project you will develop an extension to our existing platform. Prior experience with the following technologies is beneficial: JavaScript, React, React Konva, Material UI.

More Information

If you are interested in this project, please contact Matthias.Hauswirth@usi.ch to discuss the details.

LüCE
Lugano Computing Education
Research Lab

UROP project proposal at the Software Institute of the Faculty of Informatics
Università della Svizzera Italiana (USI), Lugano, Switzerland

Expression Misconceptions in Multiple Languages

Advisor: Matthias Hauswirth and the Luce team

Motivation

In most programming languages, a sizable fraction of source code consists of expressions. While some expressions seem easy to understand, in our work we have identified a significant number of misconceptions students hold about expressions.

Goal

In this project you will analyze the grammars of three programming languages, Java, JavaScript and Python, and you will define precisely which syntax rules represent parts of expressions. You will then annotate those rules and use a parser generation framework to convert the annotated grammars to parsers. With your parsers, you will analyze a corpus of student-written code to accurately quantify the fraction of code representing expressions. Finally, you will develop rules that help to detect symptoms of expression-related misconceptions.

Prerequisites

Excellent analytical and algorithmic thinking skills and ability to quickly familiarize yourself with new formalisms (such as formal syntax and semantics of programming languages).

Platform / Languages / Frameworks

In this project you will develop extensions to our existing platform. Prior experience with the following technologies is beneficial: Java, ANTLR.

More Information

If you are interested in this project, please contact Matthias.Hauswirth@usi.ch to discuss the details.

LüCE
Lugano Computing Education
Research Lab

UROP project proposal at the Software Institute of the Faculty of Informatics
Università della Svizzera Italiana (USI), Lugano, Switzerland

Type Checking and Inference

Advisor: Matthias Hauswirth and the Luce team

Motivation

Types are an essential feature of programming languages. Types help to detect errors, they help in abstraction, they serve as documentation, and they provide language safety. Understanding how a type system works is helpful when learning new programming languages.

Goals

In this project you will extend an existing platform with interactive educational activities for learning about typing and type checking. You will develop interactive visual components as well as the underlying model representing formal type system rules, and you will combine the two to provide a learning experience guided by the foundations of programming languages.

Prerequisites

Excellent analytical and algorithmic thinking skills and ability to quickly familiarize yourself with new formalisms (such as formal syntax and semantics of programming languages).

Platform / Languages / Frameworks

In this project you will develop an extension to our existing platform. Prior experience with the following technologies is beneficial: JavaScript, React, React Konva, Material UI.

More Information

If you are interested in this project, please contact Matthias.Hauswirth@usi.ch to discuss the details.

LüCE
Lugano Computing Education
Research Lab

UROP project proposal at the Software Institute of the Faculty of Informatics
Università della Svizzera Italiana (USI), Lugano, Switzerland

Semi-automatic Videos for Programming Language Misconceptions

Advisor: Matthias Hauswirth and the Luce team

Motivation

Knowing which programming language misconceptions students might develop while learning to program is an important element for instructors. Showing real-world examples of students who exhibit symptoms of those misconceptions can motivate and inspire teaching practices.

Goals

In this project you will extend our current manual approach to extract video segments from video recorded mastery checks with students to develop a tool that semi-automatically produces videos to be published in a platform that collects programming language misconceptions.

Prerequisites

Excellent analytical and algorithmic thinking skills and ability to quickly familiarize yourself with new formalisms (such as formal syntax and semantics of programming languages).

Platform / Languages / Frameworks

In this project you will develop extensions to our existing platform. Prior experience with the following technologies is beneficial: Python, FFmpeg, video-editing software.

More Information

If you are interested in this project, please contact Matthias.Hauswirth@usi.ch to discuss the details.

LüCE
Lugano Computing Education
Research Lab

UROP project proposal at the Software Institute of the Faculty of Informatics
Università della Svizzera Italiana (USI), Lugano, Switzerland

Expression Tree Data Analysis and Visualization

Advisor: Matthias Hauswirth and the Luce team

Motivation

Grading students' exercises is a time-consuming process. The instructors start creating rubrics, which then are used to grade the solutions. However, once they identify new errors, they may decide to update the rubrics and by consequence go over each solution once more. This process can be repeated several times, progressively delaying the grading.
We can partially automatize this process, and help instructors recognize erroneous patterns in students' solutions by clustering them together. The benefit is twofold: we speed up the grading process, and we can relate pattern of errors to programming language misconceptions.

Goals

In this project you will extend an existing platform that allows the collection of student's submissions. You will develop interactive visual components for enhancing the analysis of expression tree diagrams, and automatically label submissions with suggested related misconceptions.

Prerequisites

Excellent analytical and algorithmic thinking skills and ability to quickly familiarize yourself with new formalisms (such as formal syntax and semantics of programming languages).

Platform / Languages / Frameworks

In this project you will develop an extension to our existing platform. Prior experience with the following technologies is beneficial: JavaScript, React, React Konva, Material UI.

More Information

If you are interested in this project, please contact Matthias.Hauswirth@usi.ch to discuss the details.

Unsupervised Machine Learning Methods for Aspect Extraction in Information Retrieval

Information retrieval (IR) concerns itself with satisfying the user need by retrieving a set of documents relevant to the user's query. These queries can often be ambiguous and have multiple possible meanings. For example, the query "kiwi" might refer to the fruit, the bird, or the airline company, among others.

In recent years IR systems have become interactive and can now ask clarifying questions to the user when the query is ambiguous, to better understand what the user wants. Such systems might even start a conversation with the user lasting a good number of turns to better understand the user search interest. The challenge for conversational IR ystems is in knowing what kind of question to ask to the user.

In this project we aim at automatically extract potential aspects/facets of the user's query from a set of potentially relevant documents retrieved by the system with the original query, making it possible to form a clarifying question such as for example: "Would you like to know about kiwi fruit, kiwi bird or the kiwi airline?"

The student will implement several aspect extraction methods using algorithms described in recent scientific literature. These methods are largely based on unsupervised machine learning methods, like clustering, topic models, as well as graphical models.

The ideal candidate is a highly motivated student with willingness to learn and we encourage the scientific publication of the work at an appropriate conference in the IR field.

For further information please contact Prof. Fabio Crestani and/or Mr Ivan Sekulic.

Deep Neural Models for Natural Language Understanding in Conversational Search

Conversational assistants, like Alexa and Siri, are becoming increasingly popular in today's world. While their abilities are clearly improving, there are still significant limitations to their understanding of natural language, especially as the conversation progresses.

Conversational Search a subfield of Information Retrieval (IR), aimed at building Conversational IR Systems that will answer users' queries in a conversational manner. One of the biggest challenges of these systems is tracking the conversation and understanding the changes in user information need as the conversation progresses. In fact, after several conversation turns, the user might refer to a previous turn without explicitly mentioning it, or even change the topic completely. This a difficult challenge for these systems.

In this project, this challenge would be tackled with deep neural models that showed great capabilities in reading comprehension in various NLP/IR tasks, such as BERT and GPT-2. The student will train several such models on well-established conversational IR datasets with the goal of improving the fully understanding of the user need at any given turn.

The ideal candidate is a highly motivated student with willingness to learn this new challenging topic and with good knowledge of Python. We encourage the scientific publication of the work at an appropriate conference in the IR field.

For further information please contact Prof. Fabio Crestani and/or Mr Ivan Sekulic.

PROJECT PROPOSAL:
IMPROVING COMPUTATIONAL EFFICIENCY IN INFERENCE OF EVOLUTIONARY PROCESSES

Ernst C. Wit (Professor of Data Science)

Francisco Richter (Postdoc in Statistical Network Science)

DESCIPTION

Dynamical networks and trees are commonly used to describe real evolutionary processes. Those networks are typically incomplete and performing statistical inference on available trees is computationally challenging.

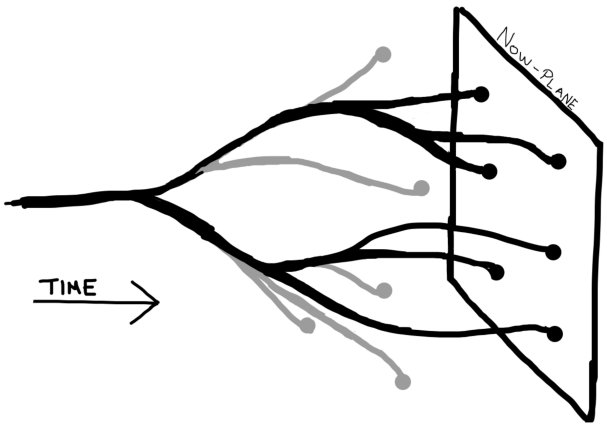


Figure 1: Evolutionary process are usually described by phylogenetic trees. A phylogenetic tree has two components, the topology or shape of the tree, and the time at its nodes.

Well-designed data augmentation algorithms play a key role when dealing with missing-data scenarios, as advanced statistical methodologies relies heavily on them. Some of these methodologies includes EM-type of algorithms, Stochastic Gradient Descent or Bayesian approaches. At the moment, little is known about the computational efficiency and comparison of those approaches.

In this project, we will quantify the computational efficiency of different tree inference algorithms. For this you will develop your own code. The results from the study will be written up in a manuscript, evaluating various statistical methodologies to perform inference on point processes involving trees and networks.

At the completion of the project the students will have (1) become familiar with general and advanced statistical methodologies, (2) gotten in touch with real applications in bioinformatics and systematic biology, (3) explored the field of statistical network science and (4) developed expertise in algorithmic design and data augmentation algorithms for trees and networks.

DATA & TECHNICAL REQUIREMENTS

We will work with real phylogenies. But as part of this project, millions of alternative phylogenetic trees need to be simulated efficiently. Therefore, R or C(++) programming skills are required. Understanding of optimization and numerical algorithms is required. Affinity with probability is essential to enjoy the project and a mathematical background in stochastic processes is an advantage, but can be learned on the job.

REFERENCES

Francisco Richter, Bart Haegeman, Rampal S Etienne, and Ernst C Wit. Introducing a general class of species diversification models for phylogenetic trees. *Statistica Neerlandica*, 74(3):261–274, 2020.