Propostas de Dissertação de Mestrado para 1º semestre 2024/25
1-jul-2024

July 1, 2024

Contents

1 Descoberta de classes de cetáceos

Proponente(s): Joaquim Silva e Sofia Cavaco

Contacto: jfs@fct.unl.pt ; scavaco@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Nesta tese pretende-se descobrir possíveis classes de cetáceos a partir de amostras de vocalizações não etiquetadas, utilizando técnicas de Machine Learning, em particular relacionadas com classificação não supervisionada. Os alunos candidatos deverão ter frequentado a cadeira de Aprendizagem Automática com aproveitamento. O aluno selecionado será apoiado pelos co-orientadores em reuniões numa base semanal.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

2 Privacy-preserving analysis of misinformation data

Proponente(s): Alex Davidson

Contacto: a.davidson@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

Statistical analysis of data seeks to identify trends and patterns of behaviour in datasets, in order to draw meaningful conclusions. Such analysis is fundamental to all empirical scientific studies of any social or physical phenomena, and is the central concept behind the creation of machine learning tools that are able to "learn" probability distributions, and then "infer" classifications of so-far unobserved phenomena with respect to said distribution.

A severe impediment to performing such analysis is that effective tools require access to raw data points, which may contain sensitive or private information that should not be shared (or ideally, even seen). As a concrete example, valuable scientific research uses data obtained from healthcare settings to identify relationships (either based on correlation or causation) between illnesses and different types of indicators obtained from alternative datasets (e.g. relating to other illnesses/demographies/lifestyle choices). Obviously, such data is highly sensitive, and typically contains information about individuals that should never be released to the wider world. As such, it is almost impossible to find a trusted entity to take in data from such different sources, and run existing statistical analysis tools. Therefore, this begs the question: how can we perform efficient and highly accurate data analysis over highly sensitive data?

In this project, we will experiment with cryptographic tools for performing statistical analyses over sensitive (and fine-grained) survey and social media data, obtained within the FARE project (`https://cordis.europa.eu/project/id/853566`). The FARE project seeks to analyse what personal factors may expose individuals to sharing fake news / misinformation. Concretely, a requirement of the project is that no individual ever views both the raw survey and social media data, apart from specific summarised aggregations run over the combination of the two. Therefore, this project will require the specification and design of cryptographic mechanisms for performing these summary aggregations, while ensuring this privacy requirement. The final result will include the programming and implementation of the solutions, alongside proof-of-concept applications that demonstrate the working system.

The ideal student profile would include those with a strong background in mathematical and statistical reasoning, alongside the prior knowledge and desire to implement such tools in software. The student will be expected to partake in semi-regular meetings with the FARE project research team (led by Prof. Joana Gonçalves de Sá), for determining progress and future steps.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

3 A new cryptographic toolkit for performing private set operations

Proponente(s): Alex Davidson

Contacto: a.davidson@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

Private set operation protocols allow two entities holding two disjoint datasets to perform collaborative computation of intersections, unions, and more complex combinations, without ever revealing their entire sensitive dataset. Such protocols can be used to encourage information sharing between competing businesses, and to perform comparisons of secret data between nation states.

Previous work on such protocols has established individual efficient mechanisms for many operations, but it would be advantageous to have toolkits that are able to compute multiple set operations, for enhanced flexibility in applications. While such toolkits exists, they tend to be lacking in efficiency, and implementations are scarce. Without such toolkits, implementers are forced to combine/implement the most efficient protocol for each individual operation, which is both a complex task, and leads to highly convoluted codebases.

This project will seek to develop a new cryptographic toolkit for private set operations, which is both more efficient and simple to implement. The project will use innovations related to the area of private information retrieval (e.g. such as those used in `https://eprint.iacr.org/2024/092.pdf`) to enhance the performance of existing mechanisms for building such protocols. The project will also aim to produce a comprehensive implementation, that demonstrates the performance of flexibility of the toolkit, with respect to previous work.

The ideal student profile would include those with a familiarity in formal computer science and statistical reasoning. The aim of the work is to develop academic publications, and students with an interest in scientific research will be prioritised.

Related reading:
- `https://eprint.iacr.org/2024/092`
- `https://eprint.iacr.org/2016/108`

Existe pré-acordo com algum aluno: Não

Nome do aluno:

4 Implementing anamorphic cryptography

Proponente(s): Alex Davidson

Contacto: a.davidson@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

Anamorphic cryptography allows innocent individuals to send covert messages without giving themselves away, whilst under the observation of an adversarial "dictator". More specifically, such individuals send messages that have a non-covert meaning that is safe to reveal to the dictator, and a covert meaning that is only readable by select individuals. The challenge for anamorphic cryptography is build this capability into existing encryption schemes, without the dictator seeing (rather than designing new schemes).

Existing anamorphic schemes are theoretical by design, but involve using standard public-key encryption schemes (such as El Gamal: `https://en.wikipedia.org/wiki/ElGamal_encryption`). However, implementations of such schemes are scarce, which means that analysing their real-world practicality is difficult.

In this project, the aim is to implement existing and experimental anamorphic encryption schemes and analyse their performance on believable messages. Ultimately, the student will produce a codebase for experimenting with anamorphic cryptography, and for providing a basis for future implementations.

The ideal student will have an interest in mathematics, and its application to cryptography. The implementations will ideally be written in Go or Rust, but other languages may be considered. The ultimate goal of the project is to have academic impact, in the form of writing publications. Students with an interest in scientific research will be prioritised.

Related reading:

- `https://eprint.iacr.org/2022/639`
- `https://eprint.iacr.org/2023/434`

Existe pré-acordo com algum aluno: Não

Nome do aluno:

5 Broadcasting encryption over FM radio

Proponente(s): Alex Davidson

Contacto: a.davidson@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
A common problem of secure Internet communication is that it is possible for powerful adversaries (e.g. nation states) to "turn off" the Internet, and force users to communicate via different mechanisms, without the security provided by Internet protocols such as TLS. While people can continue to communicate using alternative communication mechanisms (e.g. Bluetooth, FM radio) such mechanisms can be listened to by malicious entities who can act based on what they hear.
This project will investigate the capacity for building efficient protocols for broadcasting encrypted data over FM radio, which is a much more resilient technology for data transmission. The core challenges of this work will be identifying and building specific encryption methods that are suitable for producing decryptable outputs, even after taking into account noise added by FM radio transmission. The main challenge of the work will be in specifying encryption mechanisms that can be decrypted by intended recipients efficiently, without revealing the intended recipient. Ultimately, the project aims to lead to proof-of-concept implementations and experiments that can be run over real FM communication channels.
This project intends to produce novel academic results, that could serve as the basis for concrete research outputs/publications. For the mathematical sides of this work, familiarity with basic probability and linear algebra will be considered a positive. Note that the project has fairly broad and open goals, and the student will be expected to take part in, and be motivated by, the process of defining the concrete research path.
This project will be co-advised by Dr. Fernando Virdia, a cryptography researcher in the DI.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

6 Extracting explicit and implicit keywords from documents in big corpora

Proponente(s): Joaquim Francisco Ferreira da Silva

Contacto: jfs@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Nesta tese pretende-se extrair automaticamente as keywords de cada documento como uma forma de obter um sumário dos tópicos nele versados. As keywords podem ser explícitas porque estão escritas nos documentos ou implícitas por não lá estarem mas, por serem semanticamente relacionadas com os tópicos explícitos do documento. Para além de palavras isoladas, as keywords podem ser n-grams de tamanho superior a 1, isto é, multi-palavras. Pretende-se explorar abordagens alternativas para o cálculo da proximidade semântica entre termos textuais, a partir de grandes corpora.
Será dado acompanhamento ao aluno com reuniões numa base semanal.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

7 Model Checking of RESTful APIs with Graph Queries

Proponente(s): Carla Ferreira & Stefania Dumbrava (ENSIIE & Télécom SudParis)

Contacto: carla.ferreira@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

Microservice architectures are an emerging technology that builds business logic from small services. These systems are built upon independently deployed software deemed to be reliable. Nowadays, industries are adopting microservice architectures without an effective and automated methodology for testing the non-proprietary software they're using. In this architectural setting, where source code is unavailable, we propose leveraging each service API as a test artifact. Current API specification languages only offer basic types of information, i.e., the types and structure of the shared data. Therefore, we extend OpenAPI Specifications (OAS) with semantic contracts based on first-order logic to define data integrity invariants and operations' pre- and post-conditions. The M.Sc. thesis work builds upon the Magma framework that generates inputs, performs requests, verifies the results' contract compliance, and computes the test suites' coverage, both in an isolated and concurrent scenario, fully automating the microservice testing process. In this approach, we extract a TLA+ specification from the OAS that is then used to explore all the possible states of the microservice, thus serving as an oracle. The focus of the internship will be on leveraging graph database technologies, in particular, the Neo4j system and its Cypher query language, to efficiently analyze state space graphs.
Internship Objective: The goal of this internship is to efficiently compute test harnesses for RESTful API based on a graph representation of their state space. To this end, we will first import the TLA+ state space graphs provided by the TLC model checker into the Neo4j graph database. Next, we will transform these into a simplified representation that stores minimal relevant state metadata. We will encode interesting test case suits as graph query workloads and leverage Neo4j's efficient path traversal mechanism to extract the needed operation sequences.
Milestones:
Gaining familiarity with the Neo4j system [1], its Cypher query language [2], and its Python driver.
Reviewing related works on leveraging graph queries for model checking [3]
Designing an appropriate graph data model for representing the state space in Neo4j.
Implementing an efficient import script in Python.
Encoding test suites using Cypher graph queries.
Assessing the efficiency and effectiveness of the generated test suites on real-world RESTful APIs, such as Gitlab
Investigating extensions that account for dynamic analyses, integrating the recent PG-Trigger mechanisms [4].
Writing a report that provides a pedagogical account of the methodology, implementation and potential opportunities for improving the developed tool.
Technologies: Neo4j graph database, Python.
Project team: Ana Ribeiro, Carla Ferreira, Stefania Dumbrava
Bibliography:

[1] Neo4j (`https://neo4j.com/`)
[2] Cypher query language (`https://neo4j.com/developer/cypher/`)
[3] Stefano Ceri, Anna Bernasconi, Alessia Gagliardi, Davide Martinenghi, Luigi Bellomarini, Davide Magnanimi: PG-Triggers: Triggers for Property Graphs. SIGMOD Conference Companion 2024: 373-385
[4] Hojat Khosrowjerdi, Hamed Nemati, Karl Meinke: Spatio-Temporal Model-Checking of Cyber-Physical Systems Using Graph Queries. TAP@STAF 2020: 59-79

Existe pré-acordo com algum aluno: Não

Nome do aluno:

8 Measuring the climate impact of HTTPS

Proponente(s): Alex Davidson

Contacto: a.davidson@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

The rapidly increasing amount of computation brought about by the global rise of cryptocurrencies (specifically Bitcoin), and the training of hugely sophisticated artificial intelligence algorithms, has given a platform to the discussion of what impact this will have on the world's climate. In particular, since most of this computation is driven by companies and entities in the global north, we appear to be at the dawn of a new "industrial" era, that could have no less impact on the quickly evolving change observable in today's natural ecosystems.

This project will instead analyse to what extent our desire for privacy and encryption across the entire Internet, driven by the wide-scale adoption of HTTPS, could also have (and have had) an impact on the climate. In particular, the project will measure energy usage of widely-used encryption ciphersuites employed in TLS (e.g. AES-GCM), to ascertain whether such security mechanisms are potentially driving real-world change in natural environments. While such computations are much cheaper than proof-of-work, or huge machine learning training procedures, their ubiquity means that there could still be noticeable impact. Finally, the project will consider whether scaling back the guarantees of TLS (the underlying security mechanism in HTTPS) by using cheaper, less well-known ciphers — or by only targeting authentication, instead of confidentiality — may reduce the energy requirements of the protocol.

The project will require benchmarking energy usage of cryptographic implementations of well-known TLS ciphersuites, that are made available in the SUPERCOP benchmarking suite (`https://bench.cr.yp.to/supercop.html`). Energy benchmarks will be derived using Running Average Power Limit (RAPL) interfaces on suitable computing clusters. The ideal student will be familiar with C/C++ code. The aim of this project is to develop academic publications showcasing the findings of the research.

This project will be co-advised by Dr. Fernando Virdia, a cryptography researcher in the DI.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

9 Unacceptable behaviour: checking something bad does not happen in your code (with an application to Smart Contracts)

Proponente(s): António Ravara

Contacto: aravara@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

Context: some sequences of operations should not happen in program execution, because errors will arise, like trying to read a file without previously opening it, or freeing a reference already freed. When the expected behaviour of a stateful software system is specified as a finite state machine (i.e., an automaton), any valid sequence of operations on that system is an accepted word in the language of the automaton. So, sequences that are words not in the language, are necessarily invalid (i.e., unsafe, as leading to errors). If one generates automatically (some of) such sequences, one can then use an automatic verifier to check if the code implementing the desired behaviour is not allowing any such unsafe sequences.

Given the immutability and irrevocability of Smart Contracts, and the multiple real-life examples of their vulnerabilities, a mechanism to formally declare what should not happen and to automatically check that the code won't ever go wrong, would be useful. Assisted verification systems are quite demanding and lightweight automatic methods not requiring the programmer's intervention are key to supporting the sound development of critical applications.

Problems: how to automatically generate bad sequences of operations from an automata-like specification of the good behaviour of a stateful software system? How to automatically verify that a program won't ever run into such sequences?

Objectives: devise an algorithm that takes a deterministic object automaton (DOA) [1], possibly enriched with quantitative information (in the form of assertions), and produces a set of invalid sequences of operations to be verified as unsafe by the model checker Cubicle [2], providing an implementation of the automaton in the Cubicle input language.

Work plan:

1) Understand how to use DOAs combined with assertions to specify stateful software systems

2) Study and use/refine implementations of word generators for deterministic finite automata (DFA) and regular expressions (RE)

3) Study and use/refine implementations to construct a complement automaton of a DFA and to negatively test words w.r.t the language of a RE [3]

4) Develop and implement algorithms to complement a DOA and get a significant set of words to that automaton

5) Understand how to use Cubicle to specify and verify stateful software systems, getting inspiration from recent works like [4] and recent MSc theses [5,6]

6) Develop examples based on Smart Contracts to check in Cubicle, designing the DOAs of the systems, getting from the developed tool the sets of unsafe sequences, and using Cubicle to check that an implementation of the system does not run into unsafe sequences.

Research environment: the student will be part of the team of an ongoing research project, supported by NOVA LINCS and the FCT. The project has collaborations with Italian and Argentinean Universities. The work can be supported by a grant, should the student be interested.

Requisites: motivation letter (explaining also why are you fitted to this project) and cv (with course grades).

Resources:

[1] `https://arxiv.org/abs/2009.08769v1`
[2] `https://link.springer.com/chapter/10.1007/978-3-642-31424-7_55`
[3] `https://dl.acm.org/doi/10.1145/3278122.3278133`
[4] `https://link.springer.com/chapter/10.1007/978-3-030-54994-7_23`
[5] `http://hdl.handle.net/10362/161077`
[6] `http://hdl.handle.net/10362/167654`

Existe pré-acordo com algum aluno: Não

Nome do aluno:

10 Gesture classification for assembly tasks with event cameras

Proponente(s): Filipa Valdeira; Nuno Mendes; Cláudia Soares

Contacto: f.valdeira@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
In the manufacturing industry, collaboration between robots and humans is highly beneficial. A key component of this process is for robots to detect and understand the actions of human operators in a timely manner. Event cameras, able to detect changes brightness, are a recent alternative well-suited for this task, as they are sensitive to movement and ignore static information. Although promising, it is still an understudied field. The main goal of this thesis is to research, implement and evaluate state-of-the-art methods for classification of different assembly tasks in a manufacturing context, on an event-based dataset.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

11 Interpretability of recommender systems for specialist doctor referral

Proponente(s): Filipa Valdeira; Cláudia Soares

Contacto: f.valdeira@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

The choice of specialist doctors during the referral process by primary care physicians is a crucial aspect of patient care. Recommender systems are a helpful tool in this context, as they can suggest suitable specialists based on past patient history and the characteristics of both doctors and patients. A fundamental point is the interpretability of the model, as this affects the trust that patients and doctors place in it. Using a dataset of past referrals, the main goal of this thesis is to research and implement different approaches for the recommender system and evaluate their interpretability.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

12 Go deadlock fixer

Proponente(s): António Ravara and João Lourenço

Contacto: aravara@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
(Co-supervised by Prof. João Lourenço)
Context: Go is now, according to the TIOBE index (`https://www.tiobe.com/tiobe-index/`), the 12th most used programming language. The success comes also from the approach to deal with concurrency: communication-centred, instead of shared memory-centred. Nonetheless, concurrency faults like deadlocks still pose challenges to developers. Although Go has a dynamic deadlock detector, finding bugs before shipping the code
is far more effective. Some tools are already available, but their ability to deal with real-world situations needs to be evaluated. Moreover, signaling problems is not enough - finding solutions could be as challenging.
Problems: current approaches to automatically finding and fixing deadlocks in Go are restrictive - either deal with particular patterns or propose ad-hoc solutions that change significantly the behaviour of the program.
Objectives: to deal with a significant collection of bugs, one needs to include in the analysis channel closing and passing. Building on recent algorithms and tools, the aim is to devise an automatic approach taking as input Go code and detecting most of the typical cases of deadlocks, covering a larger spectrum than the one now treated. Moreover, in most cases, the tool should explain how to solve the problem(s).
Work plan:
1) examine the state-of-the-art in deadlock detection and fixing for the Go language;
2) devise a detection algorithm tackling situations due to erroneous channel closing or unsound channel passing;
3) devise a correction algorithm tackling the situations covered;
4) validate the approach with a rig and representative suite of use-cases.
Requisites: motivation letter (explaining also why are you fitted to this project) and cv (with course grades). Grades in courses of functional and of concurrent programming, as well as operating systems and theory of computation, must be at least 14. Enrolling in the course PCLT is strongly encouraged.
Resources:
[1] `https://dl.acm.org/doi/10.1145/3180155.3180157`
[2] `https://arxiv.org/abs/2004.01323v1`
[3] `http://hdl.handle.net/10362/155578`
[4] `https://dl.acm.org/doi/abs/10.1145/3551349.3561154`
[5] `https://ieeexplore.ieee.org/abstract/document/9668278`
[6] `https://www.semanticscholar.org/paper/Automated-Verification-of-Go-Programs-via-Bounded-Dilley-Lange`
`28946b5c41c396716a2c936ef3c2dcc1b77401b3`
[7] `https://www.semanticscholar.org/paper/GoDetector%3A-Detecting-Concurrent-Bug-in-Go-Zhang-Qi/`
`f1aadbf406217036965f9065eecb45b41e21eeb6`
[8] `https://dl.acm.org/doi/10.1145/3180155.3180157`

Existe pré-acordo com algum aluno: Não

Nome do aluno:

13 Automatically Verified, Coordination-Free Graph Distribution

Proponente(s): Carla Ferreira & Stefania Dumbrava (ENSIIE & Telecom SudParis)

Contacto: carla.ferreira@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

The proliferation of highly interconnected data in a wide variety of areas ranging from telecom, social, and genomic networks, to transportation graphs and linked open scientific datasets, has made it crucial to efficiently and scalably be able to process these. As such, various distributed graph data stores have been developed, such as InfiniteGraph, HyperGraphDB, ArangoDB, Neo4j Fabric, etc. These systems provide weaker consistency models than the ACID one supported in the relational setting, precisely in order to allow for better performance and seamless data evolution.

These models range from eventual consistency level (EC), which is the weakest one, to the strong consistency level, which provides the most guarantees. Each particular distributed graph system supports a different set of consistency models, which can also be subject to custom tuning. Under the EC model, replicated data can accept updates without remote synchronization, which, while advantageous performance-wise, can be error-prone. Thus, recently, a novel data type that satisfies sufficient convergence conditions has been developed. This is called a Conflict Free Replicated Data Type (CRDT) and has been implemented for graphs, for example, as part of the GunDB and graph-crdt libraries. However, no functional implementation of a Graph CRDT exists and, in particular, no adaptation of this data structure to the expressive property graph model (allowing attached properties to both nodes and edges) exists. As such, the overarching goal is to support future verification efforts aimed at improving the reliability of distributed graph systems that can capture rich data models, such as the property graph one. The internship milestones consist of:

Designing a property graph CRDT library containing custom data structures for supporting state-based and operation-based replication and proving their strong eventual consistency.

Extending the aforementioned data structures and proofs with invariants targeting the maintenance of topological properties (label-constraint reachability, transitive closure, etc.)

Enriching the property graph CRDTs with operations allowing for the extraction of graph patterns and, potentially, combining them into transactions with provable isolation properties.

The work will be carried out in the VeriFx verification framework that leverages the Scala programming language and the Z3 solver. Previous knowledge of distributed systems or databases is not required, but appreciated.

Bibliography:

[1] Marc Shapiro, Nuno M. Preguiça, Carlos Baquero, Marek Zawirski: Conflict-Free Replicated Data Types. SSS 2011: 386-400

[2] Marc Shapiro, Nuno Preguiça, Carlos Baquero, Marek Zawirski: Conflict-free Replicated Data Types, INRIA Report, 2011

[3] Jean-Christophe Filliâtre, Andrei Paskevich: Why3 - Where Programs Meet Provers. ESOP 2013: 125-128

[4] Mário Pereira, António Ravara: Cameleer: A Deductive Verification Tool for OCaml. CAV (2) 2021: 677-689

[5] Kevin De Porre, Carla Ferreira, Elisa Gonzalez Boix: VeriFx: Correct Replicated Data Types for the Masses. ECOOP 2023: 9:1-9:45

Existe pré-acordo com algum aluno: Não

Nome do aluno:

14 Using AI to Inject Noise in Java Programs

Proponente(s): João Lourenço

Contacto: joao.lourenco@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
CO-ADVISER: Prof. Jeremy Bradbury, Ontario Tech University, Canada
PROBLEM: Concurrent software is inherently non-deterministic, with huge state spaces. Many programming errors are hidden in uncommon states that depend on specific improbable interleavings (thread schedules). Testing these programs and finding the right interleaving to reveal bugs is often an overwhelming task in terms of both time and computational resources.
PROPOSAL: Noise injection is a technique inserts "white noise" (e.g., minimal program sleeps) into a running concurrent program with a goal of disturbing the thread interleaving and causing uncommon interleavings to occur, thus revealing any latent bugs hidden in these interleavings (`https://www.cs.purdue.edu/homes/xyzhang/spring07/Papers/test-thread.pdf`). This work aims to explore the benefits of using AI and specifically Large Language Models (LLMs) to identify the location and type noise to insert into a concurrent Java program.
REQUIREMENTS: The candidate must have had approval in the courses of "Concorrência e Paralelismo" and "Inteligência Artificial" with a grade >= 15 points.
APPLICATION: Please email me (joao.lourenco@fct.unl.pt) your CV and two/three paragraphs explaining your interest in this topic and which of your skills you believe are the most relevant for this project (and of course, "why?").

Existe pré-acordo com algum aluno: Não

Nome do aluno:

15 Replay of Concurrent Computations under Partial Control

Proponente(s): João Lourenço

Contacto: joao.lourenco@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
CO-ADVISER: Eitan Farchi (IBM Haifa Research Labs, Israel)
PROBLEM: Much work has been done in the past with respect to recording and replaying message-passing oriented distributed computations, from using simple logical clocks to the preservation of global orderings with vector clocks. These approaches usually assume that the record-replay infrastructure has full control over the system, intercepting sends and/or receives as needed. However, that is not always the case, as sometimes the record-replay infrastructure has only partial control over the system. Consider the case of multiple clients accessing a third-party web server: the developers of the web server may intercept the message receive and send events in the server, but have no control of those operations in the clients; reciprocally, the clients' developers may intercept the message receive and send events in the clients, but have no control of those operations in the server..
PROPOSAL: There are cases where some of the parties in a distributed computation are "closed systems", whose code cannot be instrumented, while other parties are "open systems", that can be instrumented freely. We propose to address the reproduction of message-passing oriented computations in such systems, where the record-replay infrastructure has only partial control over the system. In this project, we propose to: Study different strategies to enforce specific event orderings in a distributed computation; Implement and evaluate some of those strategies.
REQUIREMENTS: Candidates should have completed "Concorrência e Paralelismo" with a grade >= 15 points.
APPLICATION: Please email us (joao.lourenco@fct.unl.pt, eitanfarchi@gmail.com) your CV and two/three paragraphs explaining what triggered your interest in this topic, which of your skills are most relevant for this project (and of course, "why?").

Existe pré-acordo com algum aluno: Não

Nome do aluno:

16 Extending FLeec: a Non-Blocking Applicational Cache

Proponente(s): João Lourenço

Contacto: joao.lourenco@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
PROBLEM: FLeeC is an application-level cache system based on Memcached, which leverages re-designed data structures and non-blocking (or lock-free) concurrency to improve performance by allowing an unlimited number of concurrent writes and reads to its main data structures, even in high-contention scenarios. Although operational and with very good performance speedups, FleeC can be further improved.
PROPOSAL: We propose to extend FLeeC by redesigning one of the main data structures, the Slab allocator, that is used to store the cached data. The redesign must take into account the subsequent adaptation from lock-based to non-blocking concurrency control mechanisms.
REQUIREMENTS: The candidate must have had approval in the course of "Concorrência e Paralelismo" with a grade >= 15 points..
APPLICATION: Please email me (joao.lourenco@fct.unl.pt) your CV and two/three paragraphs explaining what triggered your interest in this topic and which of your skills you believe are the most relevant for this project (and of course, "why?").

Existe pré-acordo com algum aluno: Não

Nome do aluno:

17 Lazy State Determination for SQL Databases

Proponente(s): João Lourenço

Contacto: joao.lourenco@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
PROBLEM: In the past we proposed LSD: Lazy State Determination, a system that enables the use of futures in database systems. The use of futures in this context reduces the "conflicting window" of concurrent operations, which may increase considerably the commit rate and the system's throughput. Currently, there are prototype implementations of LSD for a simplified KV-store and a JDBC driver.
PROPOSAL: In this work we propose to continue the development of the current prototype of LSD for JDBC. The new implementation will do an extensive profiling of the existing prototype, followed by designing and implementing strategies to improve its performance.
REQUIREMENTS: The candidate must have concluded (approval) in the course of "Concorrência e Paralelismo" with a grade >= 15 points.
APPLICATION: Please email me (joao.lourenco@fct.unl.pt) your CV and two/three paragraphs explaining what triggered your interest in this topic and which of your skills you believe are relevant for this project.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

18 Probabilistic Loss Functions for Generative AI Models

Proponente(s): Claudia Soares

Contacto: claudia.soares@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Modelos generativos são todos os modelos de aprendizagem automática que são capazes de produzir amostras a partir de uma distribuição aprendida. Estes modelos podem ser muito simples, como misturas de distribuições normais, ou complexos como modelos de difusão usados para gerar imagens, ou ainda Transformers generativos como o conhecido GPT.
Todos estes modelos aprendem uma representação do espaço probabilístico através da minimização do erro entre a probabilidade real e a probabilidade gerada.
Atualmente, as medidas de erro usadas nestes modelos, ou são puramente não-probabilísticas, ou sofrem diversas aproximações e simplificações para se tornarem tratáveis e diferenciáveis. Nesta tese vai-se investigar "Probabilistic Loss Functions for Generative Models", sendo que se vai estudar, e desenvolver funções de custo/erro que melhorem a performance de modelos generativos do estado da arte.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

19 AutoML for Reinforcement Learning

Proponente(s): Claudia Soares

Contacto: claudia.soares@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Reinforcement Learning (RL), in conjunction with deep learning, has achieved numerous significant accomplishments, leading many to believe that RL could pave the way to generally capable agents and Artificial General Intelligence (AGI). However, RL agents' effectiveness often hinges on the design choices during the training process, which can entail tedious and error-prone manual adjustments. This makes RL application to new problems challenging and restricts its full potential. AutoML, with its proven ability to automate such design choices in several machine learning areas, has started showing promise when applied to RL. Nevertheless, Automated Reinforcement Learning (AutoRL) not only encapsulates standard AutoML applications but also introduces unique RL challenges, leading to a different set of methods. AutoRL is thus emerging as a crucial RL research area, showing potential in various applications, from RNA design to game playing to orchestration in swarms of computing nodes.

This thesis will be conducted in the context of the TaRDIS project: Trustworthy and Resilient Decentralised Intelligence for Edge Systems. TaRDIS's main aim is to markedly simplify the process and reduce the effort required to build accurate and efficient heterogeneous swarms. TaRDIS is dedicated to aiding the correct and efficient development of applications for swarms and decentralized distributed systems, combining a novel programming paradigm with a toolbox to support the development and execution of applications.

Existe pré-acordo com algum aluno: Sim

Nome do aluno: Ricardo Gonçalo

20 Attribute grammars in OFLAT

Proponente(s): Artur Miguel Dias

Contacto: amd@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
<hr><hr><h1>Attribute grammars in OFLAT</h1>
<h4>Artur Miguel Dias</h4>
<h2>Context</h2>
This thesis will expand on work already developed around the following two software tools: (1) OCamlFLAT, an OCaml library supporting many concepts of Formal Languages and Automata Theory (FLAT), and (2) OFLAT, an interactive pedagogical Web application for FLAT concepts, developed in OCaml using the previous library, the Js_of_ocaml compiler, and the Cytoscape.js Javascript library.
<h2>Objectives</h2>
The MSc student is expected to extend the OCamlFLAT library and the OFLAT Web application with support to attribute grammars. The concepts should be supported in the library and in the interactive pedagogical graphical interface. Considering the already existing support for LL and LR parsing in OFLAT, L-attributed grammars should also be investigated. Some of the expected functionalities: specification, word recognition; animation of the recognition process with display of derivation trees decorated with attribute values; circularity test and evaluation of circular attributes; develop a comprehensive collection of pedagogical exercises in the format supported by OCamlFLAT.
<p>Supervisor: Artur Miguel Dias.
<p>Nota: Nos quatro últimos anos, oito alunos do MIEI já realizaram com sucesso as suas dissertações no contexto do sistema OCamlFLAT/OFLAT.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

21 Contributions to the OFLAT platform

Proponente(s): Artur Miguel Dias

Contacto: amd@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
<h1>Contributions to the OFLAT platform</h1>
<h4>Artur Miguel Dias</h4>
<h2>Context</h2>
This thesis will expand on work already developed around the following two software tools: (1) OCamlFLAT, an OCaml library supporting many concepts of Formal Languages and Automata Theory (FLAT), and (2) OFLAT, an interactive pedagogical Web application for FLAT concepts, developed in OCaml using the previous library, the Js_of_ocaml compiler, and the Cytoscape.js Javascript library.
<h2>Objectives</h2>
In recent years, the OFLAT platform has grown a lot, accumulating numerous features. However, the pedagogical quality of parts of the user interface is questionable.
<p>Here are some of the objectives:
<ul>
<li> Use critical analysis to identify inadequacies in the user interface, from a pedagogical point of view
<li> Research a better layout for the presentations of the FLAT models.
<li> Introduce contextual help.
<li> Improve the interface for dealing with the repository of models.
<li> Expand the existing "settings" page.
<li> If necessary, add or improve in the existing models a limited number of functionalities with impact on the user interface. For example, a good simplifier of regular expressions is missing, but is not hard to write.
</ul>
<p>Supervisor: Artur Miguel Dias.
<p>Nota: Nos quatro últimos anos, oito alunos do MIEI já realizaram com sucesso as suas dissertações no contexto do sistema OCamlFLAT/OFLAT.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

22 Evaluating the quality of requirements specifications using Generative AI techniques

Proponente(s): João Araújo

Contacto: joao.araujo@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Co-supervisor: Ana Moreira
The objective of this project is to explore the use of Generative AI (GAI) techniques to improve the quality of a requirements specification. Requirements engineering is a critical phase that involves identifying, understanding, and describing the needs and constraints of stakeholders. Specifically, the project focuses on evaluating the quality of specification concerning a set of quality attributes such as Understandability, ambiguity, completeness.
Inputs: existing requirements data sets;
Outputs: diagnosis of requirements datasets and automated suggestions for improvements.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

23 Creating a Threat Modeling Protocol for Non-hierarchical Organizations

Proponente(s): Kevin Gallagher

Contacto: k.gallagher@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

Modern cybersecurity solutions often assume hierarchy. However, many organizations, including worker cooperatives, trade unions, activist organizations, open source software projects, and more attempt to avoid hierarchical structures. In this project we study how to develop threat modeling protocols that consider horizontality itself as an asset, specifically aiming for the above mentioned organizations. We will then evaluate this new protocol with members of groups of differing levels of horizontality. This work will feed into a developing body of scholarship seeking to understand how we can secure non-hierarchical organizations, and will serve as a foundation on which we can build security technology for organizations with non-traditional governance structures.

This project may involve collaboration with individuals from North America, thus requiring meetings at non-standard times, as well as the ability to communicate in spoken and written English. This project may also involve collaboration with activists groups, requiring special attention to the security of communications.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

24 Determining how people use Tor

Proponente(s): Kevin Gallagher

Contacto: k.gallagher@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Anonymity and privacy are human rights. However, the modern Internet was not designed with anonymity nor privacy in mind, forcing users who want these properties to use external tools such as the overlay network Tor. Though Tor does grant anonymity for its users assuming a specific threat model, it fails to address traffic correlation attacks. The scientific community has sought to address this issue, but experimental evaluations always contain the same problem – the scientific community does not actually know what Tor traffic looks like. In order to address this problem, we seek to study how people use Tor; that is, we seek to determine what programs people use with Tor, how frequently they use Tor, for how long, and in how many concurrent sessions. We will use this information to create a traffic model that could be used by researchers who wish to experimentally test different traffic analysis attacks and defenses.
This theme may include collaboration with the Tor Project, as well as potential publication of a scientific article.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

25 Creating a k-anonymous reporting tool for Tor using STAR

Proponente(s): Kevin Gallagher and Alex Davidson

Contacto: k.gallagher@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Anonymity and privacy are human rights. However, the modern Internet was not designed with anonymity nor privacy in mind, forcing users who want these properties to use external tools such as the overlay network Tor. Though Tor does grant anonymity for its users, it faces large user experience problems including differential treatment and broken functionality. To address these issues, the Tor Project would need to know where these issues occur, but the Tor Browser does not collect these metrics due to potential privacy problems. This dissertation would address this issue by developing a k-anonymous reporting feature for Tor using a modification of STAR. It will be co-advised by Kevin Gallagher and Alex Davidson, and may involve interaction with the Tor Project. The student working on this theme is expected to attempt to publish their findings in a relevant academic conference.

Existe pré-acordo com algum aluno: Sim

Nome do aluno: Bruno Melo

26 Measuring the UX and Usability of Operating Tor Relays

Proponente(s): Kevin Gallagher

Contacto: k.gallagher@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Anonymity and privacy are human rights. However, the modern Internet was not designed with anonymity nor privacy in mind, forcing users who want these properties to use external tools such as the overlay network Tor. Though Tor does grant anonymity for its users, it relies on nodes by volunteers who run nodes in the network. The security and performance of Tor is directly related to the number of volunteer run nodes. Despite the importance of these volunteer-run relays, nobody has yet performed a usability or user experience study to determine how easy they are to set up and administer. In this work we will address this gap by performing the first user experience study for Tor relay operation. We will attempt to identify issues that new relay operators may have, and how they attempt to reason about and resolve these issues. Finally, we will analyze what these problems may imply for the growth of the Tor network.
The student who chooses this theme will be expected to attempt to publish this work in a relevant academic conference. Findings will be sent to the Tor Project for their potential use.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

27 Studying the UX of the Element application

Proponente(s): Kevin Gallagher

Contacto: k.gallagher@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Secure Messaging (SM) has become intertwined with our everyday lives. In principle, the guarantees are simple: when you send a message, anyone that is not yourself or the recipient should not be able to read it. However, in practice, applications suffer from usability and user experience problems that could potentially endanger the security of communications, such as difficult key verification processes. In this work we will study the usability and user experience of one SM application in particular: Element, a multi-platform implementation of the federated Matrix protocol. We will seek to understand how usable users find the application to be, what issues users have when using the application, and what security issues these UX problems could lead to.
The results of this work are intended to be submitted as a potential publication to a top-tier conference in the field, such as PETS or SOUPS.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

28 A Textbook of Verified OCaml Programs

Proponente(s): Mário Pereira

Contacto: mjp.pereira@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
If we think carefully on how the World currently depends on computational infrastructures, we should really care about building bullet-proof, bug-free software. Testing, although being the most used technique at an industry level, can only provide limit guarantees about program correctness. To achieve the highest assurance guarantees, one must turn into Formal Methods, a field of study within Computer Science that whose goal is to employ mathematical techniques to rigorously analyze the behavior of a program. In particular, Deductive Software Verification, a sub-field of Formal Methods, proposes an ambitious path: turn the correctness of a piece of code into a mathematical statement and then prove it using a computer.
Over the past decades, deductive verification tools have evolved from purely academic prototypes into robust frameworks, able to analyze industrial-scale software. It is, thus, of utmost importance to properly train the next generation of verification engineers, able to apply such tools to realistic case studies. An important contribution to this endeavor is the publication of mechanized textbooks, which comprehensively introduce verification tools ranging from simple examples to research-scale problems.
In this thesis, we intend to build a textbook of verified classic algorithms and data structures. In particular, we aim to verify an extensive set of programs written in the OCaml language, a functional-first language. Using OCaml, one can very naturally write elegant, efficient, and correct code, as the language provides a very clean syntax (when compared, for instance, with Java), a state-of-the-art type system, and a flexible module system. To conduct the verification effort, we will use the Cameleer tool, a mostly-automated deductive verification tool for annotated-OCaml code [1].
Knowledge of software verification, OCaml programming, and foundations of programming languages is required. ICL, LAP, and CVS courses are mandatory pre-requisites.
[1] "Cameleer: a Deductive Verification Tool for OCaml", Pereira and Ravara, CAV 2021.

Existe pré-acordo com algum aluno: Sim

Nome do aluno: Pedro Gasparinho

29 Smart tests for smart contract languages

Proponente(s): António Ravara

Contacto: aravara@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Context: Specialised programming languages for developing smart contracts lack automatic tools to help getting the code right. Many of the existing ones require a high degree of expertise with thus high costs that many SMEs cannot support. However, the widespread adoption of blockchain technology leads to a growing need for smart contracts to manage applications on data.
Problems: with smart contracts, only at development time can one fix bugs - once deployed, contracts are immutable and any vulnerability will be exploited. Rigorous, yet automatic, approaches are crucial to support sound code development.
Objectives: building on modelling tools available for advanced smart contract languages, using model-based testing approaches, the aim is to build algorithms to automatically extract from the models tests for the code.
Work plan:
1) get acquainted with the most advance programming languages for smart contracts;
2) examine the state-of-the-art proposals in automatic testing and verification of smart contracts;
3) devise test extraction algorithms from modelling tools for the main languages;
4) validate the algorithms with a rich and representative suite of use-cases.
Requisites: motivation letter (explaining also why are you fitted to this project) and cv (with course grades). Grades in courses of functional and of concurrent programming, as well as Software Engineering and theory of computation, must be at least 14.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

30 Integrating COMA in the Deductive Verification of OCaml programs

Proponente(s): Mário Pereira

Contacto: mjp.pereira@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

If we think carefully on how the World currently depends on computational infrastructures, we should really care about building bullet-proof, bug-free software. Testing, although being the most used technique at an industry level, can only provide limit guarantees about program correctness. To achieve the highest assurance guarantees, one must turn into Formal Methods, a field of study within Computer Science that whose goal is to employ mathematical techniques to rigorously analyze the behavior of a program. In particular, Deductive Software Verification, a sub-field of Formal Methods, proposes an ambitious path: turn the correctness of a piece of code into a mathematical statement and then prove it using a computer. In practice, one must build together an implementation and its intended mathematical model, normally referred to as the logical specification. An external tool is then responsible to generate the said mathematical statement that relates the specification with the code.

Over the past decades, deductive verification tools have evolved from purely academic prototypes into robust frameworks, able to analyze industrial-scale software. However, such tools still require a high degree of expertise from users, which prevent their wide adoption by regular programmers who are not necessarily verification experts. One should aim to build more user-friendly frameworks, in particular one should invest on the construction of specification languages that can easily capture the logical behavior of some implementation.

In thesis, we aim to integrate COMA [1], a recently proposed specification technique, into the Cameleer tool [2]. Cameleer is a mostly-automated deductive verification for OCaml code. It takes as input an OCaml file, whose behavioral specification is described using the GOSPEL specification language, and translates such annotated program into an equivalent program in the Why3 verification framework. The integration of COMA in the Cameleer pipeline will allow the user to write more natural and easy to maintain specification, while maintaining the automation provide by the use of Why3.

Knowledge of software verification, OCaml programming, and foundations of programming languages is required. ICL, LAP, and CVS courses are mandatory pre-requisites.

[1] "Flexible Verification Conditions with Continuations and Barriers", Paskevich, 2023 (`https://inria.hal.science/hal-04115885/document`)

[2] "Cameleer: a Deductive Verification Tool for OCaml", Pereira and Ravara, CAV 2021.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

31 Information extraction from biomedical documents in low-resource scenarios

Proponente(s): André Lamúrias

Contacto: a.lamurias@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

Advisor: André Lamúrias

Co-advisor: Diana F. Sousa, European Commission Joint Research Centre (JRC)

The increasing volume of biomedical literature in the form of scientific papers, clinical trials and health news articles, presents a challenge for researchers and healthcare professionals to stay updated with the latest developments. Information extraction (IE) from biomedical documents is crucial for organizing and utilizing this vast amount of data effectively. However, for many research purposes there is a severe lack of annotated corpora which is required to train and evaluate deep learning models. For example, considering languages other than English, there are very few annotated datasets for biomedical information extraction. The creation of silver standard corpora can alleviate these issues [1] but more resources are necessary to improve performance for some tasks. Generative AI approaches have also been proposed, in various biomedical NLP tasks such as entity linking [2] and named entity recognition [3], however this type of approach has been underexplored in multilingual datasets.

The project aims to explore data augmentation strategies using text generation models and prompt engineering to enhance the performance of information extraction. This project will study generative models like BART on biomedical corpora and generate synthetic biomedical texts to augment training datasets. These methods should improve model performance even in languages where resources are scarce. The developed approaches will be evaluated on open-source datasets and competitions.

[1] Sousa, Diana, Andre Lamurias, and Francisco M. Couto. "A Silver Standard Corpus of Human Phenotype-Gene Relations." In Proceedings of NAACL-HLT, pp. 1487-1492. 2019.

[2] Yuan, Hongyi, Zheng Yuan, and Sheng Yu. "Generative Biomedical Entity Linking via Knowledge Base-Guided Pre-training and Synonyms-Aware Fine-tuning." Proceedings of the 2022 Conference of NAACL-HLT. 2022.

[3] Yan, Hang, et al. "A Unified Generative Framework for Various NER Subtasks." Proceedings of ACL-IJCNLP (Volume 1: Long Papers). 2021.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

32 Using graph neural networks for ligand-protein affinity prediction

Proponente(s): André Lamúrias

Contacto: a.lamurias@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

Advisor: André Lamúrias

Co-advisor: Arménio Barbosa, Biomolecular Engineering Lab.: UCIBIO – i4HB, NOVA School of Science and Technology, UNL

Molecular docking is used to estimate the affinity of a molecule to a protein, which can be used to identify candidate molecules for the purification of biopharmaceuticals for therapeutics and diagnostics [1]. However, classical docking methods do not scale well when we take into account that millions of ligands can be considered just for one study, since these methods use the full 3D structure of the atoms to calculate the affinity score. Deep docking methods [2] have been recently proposed where each molecule is represented by a set of descriptors, which can be used to train a model that predicts if that molecule is a candidate hit. This type of method can result in a 100-fold reduction of the number of molecules to be considered. On the other hand, Graph Neural Networks are a type of deep neural network that can incorporate connectivity information into a model alongside node-specific features, through message passing, that has been applied successfully to several problems [3].

There have been a number of deep docking models proposed using Graph Neural Networks [4]. In this project the aim is to develop a deep docking method based on GNN where the ligands are represented as nodes and connected according to their similarities. Additional features related to the specific application should also be incorporated into this model. Scores from existing docking simulations will be used to train and evaluate the model.

[1] Giancristofaro, A., Barbosa, A. J., Ammazzalorso, A., Amoia, P., De Filippis, B., Fantacuzzi, M., ... & Amoroso, R. (2018). Discovery of new FXR agonists based on 6-ECDCA binding properties by virtual screening and molecular docking. MedChemComm, 9(10), 1630-1638.

[2] Gentile, F., Agrawal, V., Hsing, M., Ton, A. T., Ban, F., Norinder, U., ... & Cherkasov, A. (2020). Deep docking: a deep learning platform for augmentation of structure based drug discovery. ACS central science, 6(6), 939-949.

[3] Lamurias, A., Sereika, M., Albertsen, M., Hose, K., & Nielsen, T. D. (2022). Metagenomic binning with assembly graph embeddings. Bioinformatics, 38(19), 4481-4487.

[4] Masters, M. R., Mahmoud, A. H., Wei, Y., & Lill, M. A. (2023). Deep learning model for efficient protein–ligand docking with implicit side-chain flexibility. Journal of Chemical Information and Modeling, 63(6), 1695-1707.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

33 AI-powered Requirements Engineering

Proponente(s): Ana Moreira

Contacto: amm@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
The objective of this project is to explore the use of Generative AI (GAI) techniques to enhance the requirements engineering process in software development. Requirements engineering is a critical phase that involves identifying, understanding, and describing the needs and constraints of stakeholders. By leveraging the capabilities of GAI, this project aims to automate and improve the accuracy of these tasks, thereby reducing ambiguities and facilitating better communication among stakeholders. Specifically, the project will focus on creating a RE AI-power method that can assist in identifying key requirements from initial inputs, enhance understanding by contextualizing and clarifying these requirements, and aid in their precise and comprehensive description. This innovative approach has the potential to significantly streamline the requirements engineering process, leading to a more efficient project outcomes and higher-quality software products.
Inputs: transcripts of interviews; survey responses detailing stakeholder needs and expectations; existing documents; user feedback (from previous versions of the product); technical regulatory guidelines and/or industry standards & best practices
Outputs: initial list of functional and non-functional requirements; categorization of requirements priorities based on stakeholder needs & project goals; traceability matrices (e.g., between requirements and stakeholders' inputs); initial set of user stories and use cases
Final note: Ideally, applicants have AI knowledge AI, and are doing or have done Machine Learning.
Superviors: Ana Moreira, João Araújo

Existe pré-acordo com algum aluno: Não

Nome do aluno:

34 Creating a Plugin For IDES to Save Gas in Smart Contracts Using Siphon

Proponente(s): Kevin Gallagher

Contacto: k.gallagher@fct.unl.pt

Tipo de dissertação: Projeto de engenharia

Descrição

Ethereum introduced a new era of software development within the blockchain realm. Smart contracts, which are software programs, are deployed, operated, and executed in a decentralized fashion by harnessing the capabilities of the Ethereum Virtual Machine. Smart contracts possess the capacity for endless complexity, resulting in computations for associated transactions that are equally intricate. The expense linked to performing these computations is referred to as Gas and must be provided by users to miners each time they wish to execute a transaction.

Gas limits a miner's ability to complete a transaction, meaning that the execution of inefficient Smart Contracts might only be partially carried out. When this happens, the user suffers monetary losses. Recent work has proposed Siphon, a tool that enables users to submit their Solidity-based Smart Contract and receive an optimized version that minimizes gas consumption. However, Siphon is still a difficult tool to use, and lacks many of the finishing touches that production level software would need. More, Siphon as a stand-alone tool has limited impact - integrating it into several IDEs would increase its potential impact.

In this project the student will finish implementation of Siphon to a production-ready level, and will create plug-ins for several IDEs that allow Siphon to be used to optimize Solidity smart contracts during development.

Existe pré-acordo com algum aluno: Sim

Nome do aluno: Tomé Dias

35 On-demand Task Creation for Concurrent Execution

Proponente(s): Hervé Paulino

Contacto: herve.paulino@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
The goal of this thesis is to enhance the traditional stack-based sequential execution model by incorporating the capability to generate tasks for concurrent execution, when there are available workers searching for your to do.
This project will be implemented in C++ and will integrate with the Low-Cost Work Stealing scheduler for multi-threaded computations (`https://dl.acm.org/doi/pdf/10.1145/3558481.3591099`)

Existe pré-acordo com algum aluno: Sim

Nome do aluno: Samuel Costa Fernão Pires

36 Metagenomic language models

Proponente(s): André Lamúrias

Contacto: a.lamurias@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Foundation models have shown impressive results on many natural language tasks. At the same time, there has been an explosion of genomic data thanks to the continuous improvement of sequencing technologies, both in quality and cost per genome. Although language models have been adapted to learn the language of DNA sequences, these adaptations have focused mostly on the human genome and a few selected species [1,2]. In metagenomics, the DNA of many different species needs to be classified accordingly, in order to study the microbial community of a given sample. This challenging task requires advanced methods that could be unlocked using pre-trained language models.

The main objective of this project is to expand the existing DNA language models to the metagenomic scenario, where genomic diversity is crucial for the model to understand the DNA language. As such, the first objective of this project is to train a language model on a diversity-focused training dataset. This initial model will then be finetuned for specific environments to classify sequences according to their species. The results will be compared with existing models and evaluated on standard metagenomic datasets.

[1] Y. Ji, Z. Zhou, H. Liu, and R. V. Davuluri, "DNABERT: pre-trained Bidirectional Encoder Representations from Transformers model for DNA-language in genome," Bioinformatics, vol. 37, no. 15, pp. 2112–2120, 2021, doi: 10.1093/bioinformatics/btab083.

[2] H. Dalla-Torre et al., "The Nucleotide Transformer: Building and Evaluating Robust Foundation Models for Human Genomics." bioRxiv, p. 2023.01.11.523679, Sep. 19, 2023. doi: 10.1101/2023.01.11.523679.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

37 Leveraging cell embeddings to classify neural cell types in neurodegenerative and neurodevelopmental disorders

Proponente(s): André Lamúrias

Contacto: a.lamurias@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

Advisor: André Lamúrias

Co-Advisor: Muhammad Asif, Neuroscience Department, University of Cambridge

Complex diseases such as neurodegenerative and neurodevelopmental disorders pose significant challenges in contemporary healthcare, affecting millions globally. Previously, conventional machine learning has been employed to establish phenotypic and genotypic associations [Asif 2020], disease marker identification [Asif 2018], and their associations with diseases [Vilela 2022 & 2023]. However, such studies lack information at single cell level. Elucidating the intricate cellular mechanisms underlying these pathologies is essential for the development of efficient treatments.

Recently developed single-cell omics technologies capture cellular expression at the single-cell level, providing a snapshot of thousands of features (e.g., genes) across numerous cells at multiple time points during disease development. The generated data is characterized by sparsity, high dimensionality, heterogeneity, technical noise across time points, technical discrepancies across different datasets, and challenges in data integration. Conventional statistical and machine learning methods have shown limitations in addressing the data analysis challenges posed by single-cell data. Effective and comprehensive analysis of single-cell omics datasets may elucidate the cause of neurodegenerative and neurodevelopmental diseases.

Cell embeddings [Rosen 2023] and graph learning [Veličković 2018] provide a promising approaches to unravel the complexity of these diseases, as single-cell experimental data can be used to generate graphs, which can then be used for cell type/state predictions. This project aims to combine these techniques and apply them to experimental data from patients and research projects [Szebényi 2021].

Asif M, Martiniano HF, Marques AR, Santos JX, Vilela J, Rasga C, Oliveira G, Couto FM, Vicente AM. Identification of biological mechanisms underlying a multidimensional ASD phenotype using machine learning. Translational psychiatry. 2020 Jan 28;10(1):43.

Asif M, Martiniano HF, Vicente AM, Couto FM. Identifying disease genes using machine learning and gene functional similarities, assessed through Gene Ontology. PloS one. 2018 Dec 10;13(12):e0208626.

Vilela J, Martiniano H, Marques AR, Santos JX, Asif M, Rasga C, Oliveira G, Vicente AM. Identification of Neurotransmission and Synaptic Biological Processes Disrupted in Autism Spectrum Disorder Using Interaction Networks and Community Detection Analysis. Biomedicines. 2023 Nov 4;11(11):2971.

Vilela J, Asif M, Marques AR, Santos JX, Rasga C, Vicente A, Martiniano H. Biomedical knowledge graph embeddings for personalized medicine: Predicting disease-gene associations. Expert Systems. 2023 Jun;40(5):e13181.

Rosen, Yanay, et al. "Universal Cell Embeddings: A Foundation Model for Cell Biology." bioRxiv (2023): 2023-11.

Veličković, Petar, et al. "Deep Graph Infomax." International Conference on Learning Representations. 2018.

Szebényi K, Wenger LM, Sun Y, Dunn AW, Limegrover CA, Gibbons GM, Conci E, Paulsen O, Mierau SB, Balmus G, Lakatos A. Human ALS/FTD brain organoid slice cultures display distinct early astrocyte and targetable neuronal pathology. Nature neuroscience. 2021 Nov;24(11):1542-54.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

38 Sistema de entrada/saída para o ambiente de programação NextBlocks

Proponente(s): Fernanda Barbosa & Carmen Morgado

Contacto: fb@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
NextBlocks é um plugin do moodle, que permite criar e executar actividades de programação por blocos. Neste ambiente de programação foi criado um sistema de entrada de dados (input) simples, onde foram incorporados 3 tipos de blocos de leitura de dados, que são usados como ”parâmetros” de entrada no programa.
Nesta tese pretende-se criar um novo sistema de entrada/saída no ambiente de programação, que seja semelhante a uma consola. O objetivo deste novo sistema é poder desenvolver na plataforma programas mais complexos e interativos com o utilizador. Este novo sistema de entrada/saída levará também à reformulação do sistema de testes automáticos.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

39 Extending support for the Heritage Digital Twin in Conserva

Proponente(s): Armanda Rodrigues, Matthias Knorr, Marcia VIlarigues

Contacto: a.rodrigues@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

The preservation and protection of heritage artifacts is generally based on conservational physical techniques led by trained professionals, after detailed studies of the artifacts. Current developments into digital tools, focused on supporting these processes, include digital restauration tools and, where relevant, the use of 3D models to represent the artifacts as they are available today and originally. 3D Printing can today support this work and the dissemination of its results. Moreover, the data collected, used and generated in these contexts needs to be structured and curated so that it supports future analyses and reflects the evolution of the artifacts, while supporting the conclusions of the researchers. This need can be answered by the application of the Digital Twin concept, which has been used in diverse areas as a virtual (as complete as possible) representation of an object, spanning its lifecycle, and constantly updated from the evolution of the real artifact.

A preliminary prototype has been developed in a prior thesis, but it has a number of limitations, such as being designed for a single collection and thus causing problems for managing user's access rights over different collections. This hinders possibilities of integration with other existing similar repositories, which would, if possible, allow integrated search and comparative studies. In this dissertation, we aim to design a comprehensive structure for the Heritage Digital Twin in Conservation studies, in collaboration with researchers from the Conservation - Restoration Department of NOVA FCT, and to improve the existing prototype for viewing, interacting with and analyzing associated data. The prototype will be tested with data from an ongoing project on the magic lantern (`https://www.magica-project.com/`).

Existe pré-acordo com algum aluno: Não

Nome do aluno:

40 Facilitating the usage of ASP

Proponente(s): Matthias Knorr, Ricardo Gonçalves

Contacto: mkn@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Answer Set Programming (ASP - see `https://potassco.org/` for documentation) allows for the solution of difficult combinatorial search problems, representing them in a declarative logic-based way and applying efficient solvers to obtain the solutions. Despite its declarativity, newcomers often struggle with the conceptual ideas that are fundamentally different from imperative programming languages hindering the adoption of ASP. Recently, a tool has been developed based on the methodology EZASP for facilitating the creation of ASP encodings (`https://arxiv.org/pdf/2111.06366v1.pdf`) that builds on the idea of restricting to a limited language (that can be viewed as a normal form of ASP programs) and formally add structure to programs in a way that is commonly used in encodings anyway. The tool, available as a VS Code extension, allows for detecting a number of syntax errors and provides warnings and error messages if the structure of the program is not according to the methodology. Building on this work, in this thesis, the objective is to develop a tool that allows to provide these ideas but in a dynamic fashion, that is, based on the observed problems/mistakes make suggestions and actually reorganize the developed code automatically. Also, the usage of LLMs within VSCode may be explored for suggesting improvements on syntax errors.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

41 Digital tools for studying luxury glassmaking in Roman Portugal

Proponente(s): Armanda Rodrigues, Inês Coutinho (DCR/VICARTE), Nuno Correia

Contacto: a.rodrigues@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
The primary aim of this project is to investigate Tróia's luxury glassmaking in Roman Portugal, uncovering its economic and cultural importance. The available data includes archaeological findings and material analyses, and these data will be used to address and examine production, trade, and consumption of these objects during Roman times. We aim to develop several digital tools that will support dissemination of the data and results of the project, preserving its heritage and sharing its findings. As this is a project that is currently starting, the focus of the dissertation may be on part of the digital tools to be developed, such as:
1. An interactive map of the Tróia Peninsula, involving several layers of information, to position the glass found throughout the various excavation campaigns on the ground. These layers will involve not only data from excavations but also, existing knowledge on heritage shapes, decorations, and usage of glass in the area, which will be combined with the excavation reports for spatial-temporal context;
2. Repository featuring literature on Tróia excavations, notably those concerning glass finds, aiding future research on Roman glass in Lusitania;
3. An augmented reality (AR) experience focusing on augmenting a visit of an existing physical tomb to enable visitors to handle precious fragile artifacts, added to the environment as 3D models. It should use a mobile phone and its display as the main interaction device given its availability but may enable other emerging display methods such as the ones provided by augmented reality glasses.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

42 Assessment of MATLAB's OO features based on the GoF patterns

Proponente(s): Miguel Pessoa Monteiro

Contacto: mtpm@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

In the past, the popular "Gang-of-Four" (GoF) design patterns were used as a means for assessing a language's support for modularity, module (de)composition and "separation of concerns". Implementations of all GoF patterns exist in many languages and several of them are freely available. Some of these "repositories" were used as a basis for assessments of the languages concerned. However, no such assessment of MATLAB was made, particularly its OO features. The aim of this MSc thesis is to fill this gap.

This thesis was devised with a paper in view, which will document both the MATLAB and Octave languages. It makes sense to compare the two languages since Octave is often touted as being a clone of MATLAB. Even so, it is suspected that the support for the OO paradigm of the two languages has significant differences. This thesis aims to assess this in detail. The work on Octave was already carried out by Diogo Escaleira. The present thesis covers MATLAB and will provide an analysis covering both languages.

This thesis entails producing two complete collections of MATLAB implementations of the GoF patterns. The code examples should be those used by Diogo Escaleira (originally in Java), in order to form a basis for analyzing results, including comparisons between languages.

In addition to the usual Introduction and Workplan, the Preparation Report will comprise chapters on the following topics:

(1) MATLAB: a thorough review of how MATLAB deals with well-known OO features (class instantiation, constructors, self variable, (emulation of) abstract classes and interfaces, constructors and constructor chains across inheritance hierarchies, visibility control, polymorphism, overloading, overriding, object identity, garbage collection and more).

(2) state-of-art in past GoF-driven studies, including the analysis criteria used.

(3) a review of composition mechanisms that some patterns emulate (e.g., Iterator vs for(each), Decorator vs mixins, Visitor vs double dispatch, Command vs callable objects, Abstract Factory vs family polymorphism and virtual classes).

(3) Ilustrating MATLAB implementations of a few patterns (e.g., Observer, Singleton and Command).

Existe pré-acordo com algum aluno: Não

Nome do aluno:

43 Synthesizing Session-Typed Programs

Proponente(s): Bernardo Toninho

Contacto: btoninho@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

Session types provide a powerful formalism for specifying interactions between multiple components by modelling the interactions at the type level.

The goal of this work is to integrate ideas from type-based program synthesis in a session typed language.

Specifically, the work will explore how to leverage technology from linear logic proof search, through the propositions-as-types correspondence, to synthesize concurrent functional programs from type-based specifications.

The work will explore both the theoretical aspects of proof search as well as the pragmatics of implementing a program synthesis tool in a setting with advanced features such as recursion and polymorphism. The implementation will target an existing compiler for a session-typed functional language.

Candidates should have a strong interest in programming languages, type systems and logic. Knowledge of compilers and programming language implementation is mandatory (students must have completed ICL or an equivalent compilers course).

Existe pré-acordo com algum aluno: Não

Nome do aluno:

44 Mechanizing the Metatheory of Featherweight Go

Proponente(s): Bernardo Toninho

Contacto: btoninho@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

Featherweight Go is a core formal model of the Go language, codifying many of Go's idiomatic features such as duck typing, interfaces and type assertions. In previous work, this model has been shown to be type safe following the standard type preservation and progress formalization of type safety.

Since Featherweight Go is a small subset of Go, it lacks several features that the full language includes. Moreover, being a core model it naturally acts as a playground for exploring many language extensions — Polymorphism (as proposed in the paper that introduced Featherweight Go), which has subsequently been incorporated in Go; Compilation strategies; method overloading; etc.

However, Featherweight Go's correctness proofs are informal in the sense that they have never been mechanized. While this does not compromise the correctness of the argument since the system is small enough for a pen and paper proof to be manageable, extending Featherweight Go beyond its core feature set raises questions on the verifiability of a type soundness argument.

Thus, the goal of this work is to formalize the syntax, semantics, type system and type soundness of Featherweight Go in a proof assistant (e.g. Coq, Agda, Why3), providing additional trust on the correctness guarantees of Featherweight Go and paving the way for soundly growing the model in a sustainable way through mechanized reasoning.

Candidates to this topic should have a strong interesting in programming languages, type systems and logic. Interest in programming language modeling and verification is mandatory (students must have completed ICL and CVS or equivalent courses).

Existe pré-acordo com algum aluno: Não

Nome do aluno:

45 Lazy Load Docker

Proponente(s): Sérgio Duarte e Vitor Duarte

Contacto: smd@fct.unl.pt, vad@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

A tecnologia de contentores, tipo docker, é uma forma de virtualização leve que permite distribuir comodamente aplicações prontas a executar num ambiente alvo. Para tal, são preparadas imagens que procuram satisfazer todas as dependências dessa aplicação e apenas essas. Mesmo assim, é frequente, as imagens resultantes ocuparem muito espaço em disco, gasto com ficheiros que efetivamente não são necessários ou que poderiam ser descartados uma vez que a aplicação já se encontra em execução.

Este trabalho procurará conceber uma solução, tipo docker, onde a fase de carregamento dos ficheiros contidos na imagem será feito a pedido para uma cache, em tempo de execução. Será estudada a viabilidade da solução, avaliando o seu desempenho em função da dimensão da cache para um leque de aplicações.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

46 Machine learning-driven decision support tool for exam scheduling in radiology

Proponente(s): Cláudia Soares; Filipa Valdeira

Contacto: f.valdeira@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Efficiently scheduling medical exams for a large number of patients is a complex problem that must take into account not only scheduling constraints, but also patients' medical history, priority with respect to other patients, interactions with other exams and appointments, as well as established guidelines. The main goal of this thesis is to develop a reliable and explainable system that leverages machine learning techniques to learn from historical data and assists in the decision-making process for exam scheduling. The project will be conducted in cooperation with the radiology department of Instituto Português de Oncologia (IPO), resorting to their databases.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

47 Automated and Semi-automated Monitor Generation for Distributed Protocols

Proponente(s): Bernardo Toninho

Contacto: btoninho@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

In systems where the entire codebase's source is not available (e.g. distributed systems, proprietary software) or where static verification is prohibitive, ensuring proper interaction between communicating components can only be achieved via dynamic verification, by effectively monitoring the exchanged messages between the system components.

In such systems, components are meant to carry out intricate protocols and deviations from the protocol can give rise to vulnerabilities, errors and overall system misbehavior. Thus, it becomes critical that the monitoring of communication be protocol-aware, so that monitors can detect protocol compliance failures, identifying the misbehaving component and even potentially compensating for the error.

In this thesis, the candidate will explore the problem of protocol-aware dynamic monitoring of systems in a real-world setting. The student will choose a suitable protocol specification framework from which monitors can then be automatically or semi-automatically generated in a real-world language/framework. The student will implement the generation procedure and evaluate it with real-world systems, measuring performance impact and evaluating protocol compliance.

Existe pré-acordo com algum aluno: Sim

Nome do aluno: Vladyslav Mikytiv

48 Creation of Synthetic Patient Data for Health Care Research

Proponente(s): Matthias Knorr, João Moura Pires

Contacto: mkn@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

Research in Health Care often requires collecting sensitive data from patients to be able to realize a study. However, for reasons of ethics and privacy, the collected data is commonly only available for the sake and the duration of the specific study according to the corresponding signed legal agreement, and deleted by the end of the legal agreement. This considerably complicates complementary later studies as the data is no longer available.

In this thesis, the aim is to explore the creation of synthetic data using the data collected in a concrete study within the project DSAIPA/AI/0094/2020 - An intelligent system to improve patients' safety and remote surveillance in follow-up for cardiothoracic surgery (`https://cardiofollowai.vohcolab.org/`). Building on techniques used for example in Synthea (`https://synthea.mitre.org/`) or based on machine learning, e.g., autoregressive models as used in PARSynthesizer (`https://docs.sdv.dev/sdv/sequentialdata/modeling/parsynthesizer`), the objective is to create synthetic data that replicates the data of the original study according to criteria developed in co-operation with experts, without being able to re-establish the origin of the data.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

49 Informing neural network classifiers with ontologies and self-supervised learning

Proponente(s): Matthias Knorr, Ricardo Gonçalves, Ludwig Krippahl

Contacto: mkn@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

Deep neural networks (DNN) are powerful models, but too complex for humans to understand without assistance, which is particularly critical in situations where the DNN's recommendations may affect human lives. One possible aid to understanding how a DNN reaches some result is to have a correspondence from neuron activations to intelligible concepts. This can help provide explanations for how a deep neural networks is responding.

Ontologies are used to describe what are relevant concepts (and the relations between these) in a domain, and have been successfully applied, e.g., in biology and healthcare. Previous results in our research suggest that ontologies can indeed be used to inform the training of DNN classifiers by incorporating this explicit knowledge either in the network architecture or the loss function of the training process. However, one problem we detected is that the DNN will focus on correlations with the target labels (in supervised learning) and may ignore concepts that would actually be useful to explain the predictions.

This thesis aims at evaluating how self-supervised learning can help solve this problem by forcing the network to also preserve information about the input vector in addition to providing the correct classification. The hypothesis is that self-supervised learning can prevent the network from discarding information about the examples that is important for explaining the network's predictions.

For this thesis the candidate should have previous experience with deep learning libraries (e.g. Pytorch or Tensorflow).

Existe pré-acordo com algum aluno: Não

Nome do aluno:

50 Concept mapping with self-supervised deep learning

Proponente(s): Ricardo Gonçalves, Matthias Knorr, Ludwig Krippahl

Contacto: Ricardo Gonçalves

Tipo de dissertação: Dissertação científica

Descrição

Deep neural networks (DNN) are powerful models but too complex for humans to understand without assistance, which is particularly critical in situations where the DNN's recommendations may affect human lives. However, it is known that DNN can learn compact representations of data that capture their fundamental aspects. This is evident for example in text embeddings, image representation with autoencoders etc.

One open question is whether we can make these aspects match explicit domain knowledge by making parts of the network correspond directly to relevant concepts of an ontology. Ontologies are used to describe what are relevant concepts (and the relations between these) in a domain, and have been successfully applied, e.g., in biology and healthcare.

The main idea here is to include such an ontology in the training of the network, either by tailoring the loss function to account for the logical relations in the ontology (using e.g. Logical Tensor Networks) or by previously training a part of the network to distinguish between sets of attributes that are consistent or inconsistent with the ontology. The function of this addition is to force a larger DNN trained in a self-supervised manner to obtain representations of the data that also respect the given ontology under a predetermined mapping between neurons and concepts. The hypothesis to evaluate is whether this representation can match target concepts in the ontology even though it is learned from unlabelled data. Evaluating this hypothesis can help provide important insights into how to combine symbolic explicit knowledge with complex sub-symbolic models like DNN.

For this thesis the candidate should have previous experience with deep learning libraries (e.g. Pytorch or Tensorflow).

Existe pré-acordo com algum aluno: Não

Nome do aluno:

51 Using computer visions and machine learning to generate representations of cultural heritage

Proponente(s): Nuno Correia

Contacto: nmc@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
This proposal aims at using machine learning and computer visions to generate novel representations of cultural heritage from existing materials. It will create interactive experiences on historical artifacts and cultural narratives. This approach aids in the preservation of cultural heritage and makes it more accessible and engaging to a broader audience. The proposed project will utilize a combination of techniques including machine learning, and computer vision, to analyze and reinterpret existing cultural heritage materials. Algorithms will be trained on datasets of cultural artifacts (2D and 3D content) and texts. The results will be new representations of cultural heritage that can be utilized in contexts such as arts, education and tourism. These representations will include interactive 3D models, augmented and virtual reality experiences.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

52 Performances of K-Means Clustering

Proponente(s): Susana Nascimento

Contacto:

Tipo de dissertação: Dissertação científica

Descrição

K-means algorithm remains the most popular and straightforward clustering algorithm. Its wide application across many clustering domains is due to its ease of implementation and low computational cost.

In K-means initialization process, users must specify the number of clusters in the dataset beforehand, with the initial cluster centers being chosen at random. Moreover, the algorithm's performance is highly dependent on this initial cluster selection. For large datasets, identifying the optimal number of clusters to start with becomes complex and challenging.

The main objective of this dissertation is to develop an experimental schema to study performance of K-means clustering with respect to the former aspects and to compare with recent state of the art. Real world and synthetic data will be used.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

53 Generative videos for traffic encapsulation-based internet censorship evasion systems

Proponente(s): Henrique Domingos

Contacto: hj@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Online censorship and digital authoritarianism are on an upward trend worldwide, with
some sources estimating that more than 60 countries implement information access
control with many countties strongly involved in internet censorship activities. Censorship evasion systems and tools that provide anonymity (e.g., Tor) have
emerged to allow users in repressive regimes to access uncensored web resources. Some of other tools, such as TorKameleon or Oktopus (developed at DI-FCT-UNL and NOVA LINCS), evade known traffic identification and blocking
tactics used by censors by using real-time video streams to build covert channels encapsulating user traffic, by replacing the content of video frames with it. However, it is still unclear which types of video characteristics and features are best suited in terms of undetectability (i.e., the likelihood of the censor detecting that the user is using a censorship evasion system by analyzing (capturing, correlating or finding fingerprints) the video-generated traffic) and performance characteristics for traffic encapsulation, and thus, which should be used with protectio systems (ex., TorKameleon).
In this dissertation topic the idea is to study the use of novel automatic generative video (from novel AI based tools and APIs - ex :https://www.veed.io/) to dynamically create/generate video-encapsulated traffic morphed covert channels. The targeted solution can be addressed to extend privacy and anonymity mixnet-typ protection solution for censorship circumvention (avoidance) to protect users' privacy in internet communication. The targeted solution can be leveraged by existent research systems developed in DI-FCT-UNL and NOVA LINCS (TorKameleon or also Oktopus Solutions). For this dissertation, students will work in collaboration with other MSc and PhD students at DI/FCT/UNL. For initial references see: https://www.computer.org/csdl/proceedings-article/trustcom/2023/819900b490/1XldQwabdoQ, https://github.com/net4people/bbs/issues/331, https://arxiv.org/abs/2303.17544
The work will be also developed in open collaboration for other researchers at Nova Lincs
1) Review the state of the art.
2) Study the impact of different types of videos (e.g., different resolution, entropy,
subject) on detectability and performance for traffic encapsulation for privacy and anonymity preserved covert channels in internet communication
3) Develop a generative video component to automatically create videos with such
characteristics for video-morphing covert channels protecting internet communication traffic

Existe pré-acordo com algum aluno: Não

Nome do aluno:

54 Use of Differential Privacy for Unobservable Privacy-Preserved Communication

Proponente(s): Henrique Domingos

Contacto: hj@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

The main goal of this thesis is to create a dynamic communication environment capable of withstanding state-of-the-art correlation attacks to protect anonymity and privacy while maintaining good throughput and latency conditions for practical use in different application contexts. In order to achieve that goal, we will combine the capabilities of Oktopus - a solution previusly developed in DI-FCT-UNL, which uses K-Anonymization and WebRTC-based and TLS tunnels to establish covert communications, with a dynamic multipath routing. The dissertation will extend the Oktopus solution providing an alternative form of privacy-preserved transport chanels based on differential-privacy circuits.

Differential privacy (DP) is a mathematically rigorous framework for releasing statistical information about datasets while protecting the privacy of individual data subjects. In the context of the thesis, the technique will enable data flows and traffic patterns to be transmitted limiting the information that can be leaked by advanced censorship techniques breaking anonymity and privacy, even when the traffic is encrypted). To implement differentially-privacy circuits the idea is to inject carefully calibrated noise into statistical computations and generated communication patterns such that the utility of the statistic is preserved while provably limiting what can be inferred about traffic captured and analyzed by a censor.

This dissertation work will be developed in a strong collaboration with João Vilalonga, a PhD student at DI-FCT-UNL working on related aspects of the topic, being a co-author of a base platform (Oktopus) that will be used as a leveraging system to develop the targeted solution. The thesis will be also developed in collaboration with Prof. Kevin Gallagher at DI/FCT/UNL - NOVA LINCS

Existe pré-acordo com algum aluno: Não

Nome do aluno:

55 Use of Social Networks as Sources for Cybersecurity Management Systems

Proponente(s): Henrique Domingos

Contacto: hj@fct.unl.pt

Tipo de dissertação: Projeto de engenharia

Descrição
Cybersecurity is a growing concern as the number and gravity of cyber- attacks and incidents are continuously increasing. The knowledge of latest updates on vulnerabilities (just in time), patching, and effectiveness of these information bases are crucial to maintaining an IT infrastructure's high-security level and operation of Cybersecurity Engineers and Teams in Security Operation Centers of different companies and organizations. Typically, security auditing tools and platforms (such as SIEM monitoring platforms) use continuous feeds of cybersecurity information from relevant CVE and CERT Centers' databases, nationally and widely produced from very good reputation origins but also with very high usage licensing costs. Alternative to purchasing expensive cybersecurity news feeds is to learn and audit systems from collected OSINT (Open Source Intelligence) sources: a wealth of knowledge published daily by a huge number of users, security companies, researchers, cybersecurity auditors or hackers, among others. In particular, Twitter or Reddit ar examples of information hubs for obtaining cutting-edge information about many subjects, including cybersecurity.
In this thesis the goal is to analyze if social networks are today valid sources for collection and processing of cybersecurity-related news or posts that must/can be combined with valid input sources in combination or in comparison with other sources. For the expected analysis, it is necessary to conduct a systematic and scientific approach, addressing a qualitative and quantitative study about the security data found on specific social networks (with a possible focus on Tweeter/X or Reddit (and comparative data to databases that publish confirmed vulnerabilities or exploits. The final milestone is to answer the following question: is X or Reddit good or possibly valid or better sources for cybersecurity analysts ?
To address the proposal, different tasks must be developed:
1) Study of related wrk on using informal and social network solutions as sources of cybersecurity incidents and vulnerability assessment
2) Selection of a target to conduct a systematic approach (next steps). In this case, th starting point is X (Tweeter), in which we can explore some interesting advantages for the intended purpose
3) Design a methodology for assessment purposes
4) To design tool and analytics environmentto support a pipeline comprising data aggregation, text filtering, text feature extraction, text-to-binary classifier, clustering, and event's compromise generation.
5) To conduct a validation and experimental analysis for one (or both) of the following objectives:
a) Integration of the detected events (from the previous pipileline processing) with a SIEM platform (ex., based on Elastic Search / logstash analysis), ex: OSSIM
b) A comparative study of effectiveness of the provided pipeline processing with information from other cybersecurity event database sources

Existe pré-acordo com algum aluno: Não

Nome do aluno:

56 Fuzzy Spectral Clustering over Domain Taxonomy to Explore Research Tendencies in Data Science

Proponente(s): Susana Nascimento

Contacto:

Tipo de dissertação: Dissertação científica

Descrição

The need to interpret high-dimensional data sets has led to developments on dimensional reduction and kernel based clustering techniques. Crisp and fuzzy spectral clustering play a key role in this context.

Fuzzy Additive Spectral Clustering (FADDIS) [1,2], has proven to be effective when applied to different types similarity data such as affinity data, graph data, and real world applications like the analysis of research tendencies in Data Science.

The goal of this work is to extend FADDIS with additional similarity measures as well as other technicalities of the method.

The application domain is the analysis of research tendencies in Data Science from published papers in reference journals in the field.

[1] Boris Mirkin, and Susana Nascimento (2012). Additive Spectral Method for Fuzzy Cluster Analysis of Similarity Data Including Community Structure and Affinity Matrices, Information Sciences, 183(1), pp. 16-34, Elsevier, January 2012, (http://dx.doi.org/10.1016/j.ins.2011.09.009).

[2] R. Felizardo A study on parallel versus sequential relational fuzzy clustering methods (Master thesis) http://hdl.handle.net/10362/5663

Existe pré-acordo com algum aluno: Não

Nome do aluno:

57 [SI-MORENA] - LAND IT - Edição Assistida e Colaborativa do Reordenamento do Território

Proponente(s): João Moura Pires

Contacto: João Moura Pires

Tipo de dissertação: Dissertação científica

Descrição
O LAND IT (Land ANalisys and Design of the Integrated areas of the Territory) é um sistema de apoio à decisão, no formato de aplicação web, que pretende auxiliar na criação e gestão das áreas integradas da gestão da paisagem. O seu objetivo é otimizar todo o processo de criação, tornando-o mais rápido e evitando o aparecimento de erros. Este processo tem como base a paisagem atual do município de Mação, e, através de transformações ao território, pretende-se chegar a um novo desenho da paisagem, de modo a reduzir bastante os incêndios florestais de grande escala que já chegaram a atingir quase 70
Ao longo deste processo são criados vários cenários, ou seja, propostas de ordenamento, que podem ser partilhados entre utilizadores. Assim, pretende-se que o sistema inclua a
funcionalidade de edição simultânea de cenários, em que um utilizador cria um cenário e
partilha-o com um ou vários utilizadores e todos podem fazer alterações em tempo real. Sendo estas alterações maioritariamente manipulação de geometrias. Será necessário garantir um correto funcionamento da funcionalidade, e que os dados da aplicação sejam coerentes entre sessões. Além de que é essencial ter um histórico de versões que inclua as edições de todos os utilizadores envolvidos, bem como que seja sempre possível saber quem atualizou uma geometria. Deste modo, pretende-se facilitar a criação de cenários envolvendo múltiplos utilizadores, visto que atualmente um cenário só pode ser editado por um dos utilizadores que lhe tenha acesso à vez.
Outros aspetos também deverão ser tratados, como garantir que os utilizadores têm acesso a atualizações de outros utilizadores em tempo real, como eliminar ou criar um novo cenário. Para este tema será necessário utilizar linguagens de programação como Java e TypeScript.
O aluno que ficar com este tema fará parte da equipa do LAND IT e após a obtenção da
aprovação na preparação de dissertação poderá obter uma bolsa de investigação.
O LAND IT teve início nesta tese: `https://run.unl.pt/handle/10362/163563`

Existe pré-acordo com algum aluno: Não

Nome do aluno:

58 Decentralized, Oblivious and Privacy-Preserving Cloud-Enabled Single Sign-On

Proponente(s): Henrique Domingos

Contacto: hj@fct.unl.pt

Tipo de dissertação: Projeto de engenharia

Descrição
Currently, in single sign-on authentication schemes on the web, users are required to interact with identity providers securely (such as Google, Facebook, Amazon etc) to set up delegated authentication data during a registration phase and receive authentication tokens (sometimes as multifactor based authentication credentials) for future access to cloud services and applications. As an example (among others), OAuth2 is a well known protocol for Web authentication. This type of interaction can make authentication schemes challenging in terms of security and availability. However from a security perspective, a main threat is theft or abuse (or exfiltration/leakage dangers) of authentication reference data stored with or from those identity providers. An adversary also could easily abuse such data to mount an offline dictionary attacks for obtaining the underlying passwords, secrets or biometric elements (and these practices happened). From a privacy perspective, identity providers are able to track user activity, learn from inference perspetive about authentication procedures and control sensitive user data, with no control from the involved users. In terms of availability and security, users rely on the trusted third-party servers that need to be available during authentication.

In this dissertation the idea is to propose a novel decentralized privacy-preserving single sign-on scheme through a Decentralized Anonymous, Privacy-Preserving and Oblivious Multi-Factor Authentication solution (DAPO-MFA). Moreover, the authentication protocol must eliminate the dependence on an always-on central identity provider authority during user authentication, allowing service providers to authenticate users at any time by interacting with a decentralized and trustable authentication solution. In the solution multiple servers will cooperate by using novel privacy-preserving crypto primitives supporting the authentication protocol. Moreover, the solution will be designed for byzantine intrusion tolerance (even that one or more of the multiple servers are targeted by intruders) and each server cooperating in th process doesn't learn nothing from the authentication protocol or about users being authenticated.

As validation of the solution, we will implement a MultiCloud-based DAPO-MFA Oauth2 Service for Web Applications, showing the correctness of operation, comparative performance of authentication with a centralized SSO solution and demonstrating/discussing the achieved privacy-preservation guarantees.

The dissertation work will be developed in a collaborative context involving previous ideas and preliminary work from a PhD student and solution addressed at NOVA LINCS

Existe pré-acordo com algum aluno: Não

Nome do aluno:

59 Multimodal Gamified Fitness application companion with Conversational Agent and Camera detection

Proponente(s): Rui Nóbrega

Contacto: rui.nobrega@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

Colaboração: Pedro Valente

In this master thesis, you'll be developing a multimodal gamified fitness application that hosts a sophisticated conversational system powered by the Generative Pre-trained Transformer 4 (GPT-4). By integrating Intel RealSense camera technology for automatic skeleton detection and depth imaging, the application aims to create an engaging and immersive user experience. The project will leverage a client-server architecture with a user-friendly interface for participants and a control interface for researchers. This setup will facilitate Wizard of Oz studies, a research method where a human operator simulates the behavior of an intelligent system, allowing for real-time simulation of conversational agents.

The research will explore the impact of these advanced conversational agents on user engagement and task performance in a practical, task-oriented setting. By addressing key areas in Human-Computer Interaction (HCI), multimodal interfaces, and computer graphics, this thesis aims to contribute significantly to the field, providing valuable insights and a robust tool for future HCI research and practical applications.

Some example publications:

1. Cooking With Agents: Designing Context-aware Voice Interaction, CHI 2024, `https://doi.org/10.1145/3613904.3642183`
2. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents, CHI 2016, `https://doi.org/10.1145/2858036.2858288`
3. What Makes a Good Conversation?: Challenges in Designing Truly Conversational Agents, CHI 2019, `https://doi.org/10.1145/3290605.3300705`
4. Designing Conversational Agents: A Self-Determination Theory Approach, CHI 2021, `https://doi.org/10.1145/3411764.3445445`
5. Beyond Self-diagnosis: How a Chatbot-based Symptom Checker Should Respond, ACM Transactions on Computer-Human Interactions 2023, `https://doi.org/10.1145/3589959`
6. Designing a Motivational Agent for Behavior Change in Physical Activity, Short in CHI 2015, `https://doi.org/10.1145/2702613.2732924`

Existe pré-acordo com algum aluno: Não

Nome do aluno:

60 Geração Procedimental de Interiores em Realidade Virtual (VR) explorando Novos Paradigmas de Navegação.

Proponente(s): Rui Nóbrega

Contacto: rui.nobrega@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Colaboração: Ana Rita Rebelo
Um dos desafios em Realidade Virtual (VR) é a limitação do espaço físico disponível para os utilizadores se movimentarem. Frequentemente, a locomoção em VR depende do uso dos comandos de VR, usando técnicas como teletransporte ou joystick. Embora sejam soluções eficazes, estas técnicas não são naturais e podem limitar a sensação de imersão dos utilizadores.
O objetivo desta tese é utilizar Procedural Content Generation (PCG) para adaptar ambientes virtuais amplos, de forma que se ajustem ao espaço físico disponível ao redor do utilizador. O trabalho será desenvolvido essencialmente utilizando a plataforma Unity.
Serão ainda exploradas novos paradigmas de exploração de espaços virtuais baseados nos conceitos de expaços impossíveis e hiperbólicos.

Existe pré-acordo com algum aluno: Sim

Nome do aluno: Sofia Monteiro

61 Video see-through Augmented Reality for Collaboration and Exploration in Museum contexts

Proponente(s): Rui Nóbrega

Contacto: rui.nobrega@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

This thesis will explore the advantages and limitations of using video see-through mixed reality devices to enable interactive in-situ augmented reality in spaces such as museums, cultural venues and collaborative information sharing between art conservation technicians.

Taking advantage of the 3D space detection of modern Mixed Reality devices, the goal is to explore their capabilities to the fullest while integrating with traditional AR frameworks such as AR Core or Vuforia. Additionally, there should be integration of procedural content generation that can be superimposed into reality considering existing objects and current illumination.

As a use case, we will use the visualization of objects from Portuguese museums in AR with the ability of sharing them collaboratively online or simulating an archaeological site in the lab for further research context.

Keywords: Mixed-reality, Computer Graphics, HCI, Cultural Heritage

Existe pré-acordo com algum aluno: Sim

Nome do aluno: Guilherme Figueira

62 Secure and Trusted Cloud-Edge-IoT Pipelining Computing for ML Applications

Proponente(s): Henrique Domingos

Contacto: hj@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

New applications exploring computation pipelines managed in Cloud-Edge-IoT continuum environments are examples of novel applications of autonomous systems. Some of these systems need to address computations based on ML models. As an example of such systems, Autonomous stores providing checkout-free shopping systems for the retail industry are leveraged by more or less complex networks of sensors and take decisions from machine-learning models. As the retail industry rushes to adopt this technology, ensuring the privacy of sensor data and necessary guarantees for the proprietary software and their ML models with digital rights protection. Trusted Execution Environments (TEE) and solutions such as ARM Trust Zone, Intel SGX (`https://en.wikipedia.org/wiki/Trusted_execution_environment`) are attractive technololy to the problem, the impact on system performance and isolated processing resources limit their widespread adoption in scenarios where high performance and workloads imposed by ML-based procesisng and low latency for applications are crucial.

The goal of this dissertation project is to develop a secure computing environment that protects the confidentiality of sensitive information, such as sensor data and other SW intellectual property, running critical code in HW-enabled isolation, while minimizing the performance impact.

As a validation environment for the dissertation, we will address a system that aims to protect IP-sensitive video-streaming and input data as we will found in the next generation of autonomous stores running on untrusted platforms and computing critical code (far from away in the cloud). In a different approach, we want to provide an hybrid Cloud-Edge-IoT application, leveraging critical computations running in the Edge with Intel SGX TEE. To ensure optimal performance, the system will implement mechanisms for securely offloading havy weight computations to untrusted GPUs, while keeping critical processing functions isolated in the TEE.

Plan:

The work will target i) a design model and specification of the intended system and protocols; ii) a prototype implementation of the designed system in a Cloud-Edge-IoT distributed environment, iii) an experimental evaluation for the solution validation.

Sep-Nov: study of related work and background

Dec: first approach to write the preparation report (introduction and related work)

Jan: definition for elaboration guidelines and implementation options, based on a desitn specification of components for implementation

Feb: defense of the report preparation

Mar: Complete/Revise related work

Apr-Jun: Design and implementation

Jul: Validation and experimental evaluation of the produced prototype

May to Sep: Dissertation report and submission for defense.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

63 [SI-MORENA] - Análise do Impacto do tráfego urbano na qualidade do ar

Proponente(s): João Moura Pires, Francisco Ferreira (Dep. Ambiente)

Contacto: jmp@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Numa tese de mestrado que está a decorrer (termina em Setembro de 2024) estão a ser analisados dados da qualidade do ar em Portugal continental desde 2001 até 2022 com o objectivo de perceber as tendências globais e locais para os principais poluentes que estão a ser monitorados. A fonte destes dados é `https://qualar.apambiente.pt/`. Este estudo incorpora a análise do impacto da meteorologia nas concentrações dos poluentes.
Tendo como ponto de partida este trabalho, pretende-se nesta tese de mestrado focando-se nas cidades principais (Lisboa e Porto) estudar a relação entre as concentrações de diversos poluentes do ar e a intensidade do tráfego na cidade de Lisboa, recorrendo a dados das estações de monitorização da qualidade do ar e a dados de tráfego de Google ou equivalente. Naturalmente que terá que integrar a informação meteorológica.
Trata-se de uma tese de visualização analítica de dados podendo vir a desenvolver modelos de previsão concentrações de diversos poluentes com base nas previsões de tráfego e da meteorologia.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

64 [SI-MORENA] - LAND IT - Análise e Visualização de Cenários

Proponente(s): João Moura Pires

Contacto: jmp@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
O LAND IT (Land ANalisys and Design of the Integrated areas of the Territory) é um sistema de apoio à decisão, no formato de aplicação web, que pretende auxiliar na criação e gestão das áreas integradas da gestão da paisagem. O seu objetivo é otimizar todo o processo de criação, tornando-o mais rápido e evitando o aparecimento de erros. Este processo tem como base a paisagem atual do município de Mação, e, através de transformações ao território, pretende-se chegar a um novo desenho da paisagem, de modo a reduzir bastante os incêndios florestais de grande escala que já chegaram a atingir quase 70
Durante este processo serão criadas diversas propostas de ordenamento (cenários), e é
fundamental que os gestores as possam analisar, de modo a decidir quais os próximos
elementos a alterar, com o fim de atingir uma versão final. Nesta análise é importante retirar
conclusões sobre:

- as áreas ocupadas por cada proposta de uso (florestas, agricultura,...);
- quais os custos de cada transformação e quais as transformações mais dispendiosas;
- como são distribuídos os apoios nos próximos 20 anos à transformação;
- quais os lucros a obter com estas transformações.

Atualmente, o LAND IT inclui algumas visualizações para estes tópicos, mas estão
desatualizadas ou incompletas. Assim, o objetivo é reformular e criar novas visualizações que auxiliem os gestores nas suas decisões. Estas visualizações têm de ser incorporadas no LAND IT de modo a garantir que cada utilizador só tem acesso às visualizações que lhe compete, o que envolverá utilizar linguagens como o TypeScript e Java.
O aluno que ficar com este tema fará parte da equipa do LAND IT e após a obtenção da
aprovação na preparação de dissertação poderá obter uma bolsa de investigação.
O LAND IT teve início nesta tese: `https://run.unl.pt/handle/10362/163563`

Existe pré-acordo com algum aluno: Não

Nome do aluno:

65 [SI-MORENA] - Avaliação dos caudais dos rios de Portugal

Proponente(s): João Moura Pires e Francisco Ferreira (Dep. Ambiente)

Contacto: jmp@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
O Sistema Nacional de Informação de Recursos Hídricos (`https://snirh.apambiente.pt/`) é um repositório algo antiquado em termos de interface, mas extremamente valioso em termos de informação sobre qualidade e disponibilidade da água nas suas diversas vertentes (desde água nos rios e albufeiras, a águas balneares). Sistematizar, compreender e sistematizar os padrões de determinados caudais é um elemento fundamental na gestão de um recurso cada vez mais escasso e com situações de conflito como o facto dos principais rios em Portugal iniciarem o seu percurso em Espanha.

Neste tese pretende-se a avaliação dos caudais dos principais rios portugueses nos últimos dez anos em diversos troços e relação com dados da meteorologia e clima. Trata-se de uma tese de visualização analítica de dados podendo desenvolver instrumentos que permitam analisar os dados disponíveis para perceber tendências e padrões na disponibilidade da água nesses cursos e nas bacias e apoiar a procura de correlações importantes com a meteorologia.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

66 Interactive tool for practicing and evaluating logic exercises

Proponente(s): Ricardo Gonçalves e João Costa Seco

Contacto: Ricardo Gonçalves

Tipo de dissertação: Dissertação científica

Descrição
Logic is a basic topic for several disciplines, such as mathematics and computer science. Whereas some tools exist that allow an interactive environment for learning logic, these are usually either based on an installable program, or static, not easily allowing the extension with additional material.
Recent work has started to develop a system that could provide the basis for an online system that can allow the interactive resolution and assessment of logic related exercises, with the flexibility to incorporate new types of exercises and open to allow teachers to easily add new exercises of different types. Nevertheless, the set of specific types of logic exercises that have been implemented was just for testing purposes and is very limited, both in number and depth. The goal of this thesis is to study, design and implement different types of exercises typical of an introductory logic course, both in propositional and first-order logic. The proposed system should be interactive, not only to allow students to automatically verify their solutions, but, as important, to provide detailed and specific feedback in case of incorrect solutions (without revealing the solutions). Moreover, the system should allow teachers to easily add new exercises and provide the means to automatically grade the exercises done by the students. This would allow its usability for automatic evaluation purposes, namely when integrated with existing online e-learning platforms such as Moodle.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

67 Trustworthy Medical LLM to inform clinical decision support systems

Proponente(s): Claudia Soares, Filipa Valdeira

Contacto: claudia.soares@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

General Large Language Models (LLMs) like PaLM, the LLaMA family, GPT-series, and ChatGLM, have significantly improved various natural language processing tasks such as text generation, summarization, and question answering. These advancements have inspired the development of medical LLMs specifically adapted to the field of medicine.

Models like MedPaLM-2 and MedPrompt, based on PaLM and GPT-4, have achieved competitive accuracy rates compared to human experts in the United States Medical Licensing Examination (USMLE). Furthermore, a variety of medical LLMs, including ChatDoctor, MedAlpaca, PMC-LLaMA, BenTsao, and Clinical Camel, have been developed based on publicly available general LLMs like LLaMA. The use of medical LLMs to assist medical professionals in improving patient care is an area of growing interest.

However, despite promising results, there are key issues that need to be addressed in the development and application of medical LLMs. Many models focus mainly on medical dialogue and question-answering tasks, but their practical use in clinical practice is often overlooked. While some research has begun to explore the potential of medical LLMs in different clinical scenarios, there's a lack of practical guidelines for their development and insufficient evaluation datasets for assessing their performance in these scenarios.

Moreover, most medical LLMs report their performances primarily on answering medical questions, neglecting other biomedical domains like medical language understanding and generation.

Finally, the trustworthiness of these LLMs' outputs is not adequately addressed.

This thesis will propose an LLM pipeline that can parse Electronic Health Records text boxes and retrieve important information for patient healthcare, while guaranteeing trustworthy information. The thesis is to be done in collaboration with the IPO Lisboa, the largest national center for cancer treatment and research.

Existe pré-acordo com algum aluno: Sim

Nome do aluno: Ricardo Pereira

68 Database for the historical manufacture of artist's materials

Proponente(s): Matthias Knorr

Contacto: mkn@fct.unl.pt

Tipo de dissertação: Projeto de engenharia

Descrição

The Winsor & Newton 19th century archive database is a unique primary documentary source for the manufacture of artist's materials covering handwritten documents on instructions and workshop notes for, e.g., pigments and paints. A database solution exists which allows its users to access the digitalized versions of these copyright-owned manuscripts together with complementary information including a transcription of the used recipes for the materials, which facilitates search based on diverse criteria. This provides valuable information for researchers in material sciences and conservation sciences, as well as art historians.

Access to this database, first introduced in 2006, has been limited to specific physical locations (desktop computers) for reasons of commercial sensitivity of some of the data, and in practice edition of the contents is not possible in general (outside of the complete version held by the parent company). Recently, concerns have been raised that the limited access and the lacking option of making changes (on the existing data or complementing with data on other documents) is hindering the advances in science, and that a shared and editable database would be preferable, though still requiring access to it can be controlled. The current technical solution is not suitable for that.

In this thesis, the objective is to create a prototype of a web-based solution of such a database. Among the interesting challenges are the faithful conversion of the contents including the sources' images from the previous desktop-based version. This thesis is realized in collaboration with researchers from the Conservation - Restoration Department of NOVA FCT who have access to the contents and collaborate with the owners of the source documents.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

69 Teaching Platform for Computational Thinking with unplugged activities

Proponente(s): João Costa Seco

Contacto: Joao.seco@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
O projeto OctoNOVA decorreu em 2020 e 2021 como uma Academia Gulbenkian de conhecimento (octonova.di.fct.unl.pt).
O principal objectivo era criar recursos próprios para alunos de primeiro ciclo e potenciar as capacidades de professores.
Como principais resultados criaram-se e publicaram-se atividades sem recurso à tecnologia, aplicaram-se e avaliaram-se num conjunto de escolas da zona da grande lisboa.
Nesta tese propõe-se o estudo dos conceitos de computational thinking que estão na base destes ensinamentos: Abstração, Padrões, Lógica, Codificação, Debugging, etc. Os principais recursos são a literatura mais atual, e a sistematização das atividades desenvolvidas pelo projeto.
Para suportar o desenvolvimento de novas atividades, propõe-se o desenvolvimento de uma metodologia, suportada numa plataforma de construção de planos de atividades unplugged para o ensino de conceitos computational thinking no primeiro ciclo.
Existe a possibilidade de atribuição de uma bolsa de estudo.
Keywords: Computational Thinking, Unplugged activities, Learning platforms, User studies.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

70 Application construction User interface for live programming

Proponente(s): João Costa Seco

Contacto: Joao.seco@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Programação Live ou programação em tempo real é um paradigma que permite a construção de sistemas por incrementos coerentes que não necessitam de parar a execução do sistema para que possam ser alterados. O artigo [1] apresenta uma semantica de uma linguagem de programação que merece ser estudada do ponto de vista da interface com o utilizador, num ambiente de programação próprio.
Nesta tese propõe-se a construção de um ambiente de desenvolvimento visual para uma linguagem de programação reativa com suporte para a reconfiguração de programas recorrendo a tecnologias avançadas como touch-screens, gestures, óculos VR, etc.
[1] Domingues, M., Costa Seco, J.: Type safe evolution of live systems. In: Workshop on Reactive and Event-based Languages & Systems (REBLS) at SPLASH. (2015)
Keywords: Programming languages, Live programming, Reactive programming, User interfaces, Development environments

Existe pré-acordo com algum aluno: Não

Nome do aluno:

71 DCR Choreography mapped onto Actyx platform

Proponente(s): João Costa Seco

Contacto: Joao.seco@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

Swarm architectures depict software based on the decentralised coordination of multiple participants, making them well-suited for large-scale distributed systems. Swarms concurrently execute processes in highly distributed environments, where participants can dynamically join and leave the system. Global behaviours emerge from the local interactions between participants. Swarm programming requires a flexible programming model that allows the definition of concurrent processes executed by multiple participants.

In this thesis topic, we propose the application of Dynamic Condition Response (DCR) choreographies to the definition of a data dependent behaviour in a swarm. DCR choreographies define messages and data exchanged between participants and set the control-flow constraints beyond basic data dependencies representing the system's business logic.

The work includes the projection of DCR choreographies to local processes running on the Actyx platform, an event based framework for decentralised systems, allowing for the execution of swarms with weak consistency.

This topic can be awarded a scholarship from the TaRDIS project.

Keywords: Programming languages, Event-driven programming, Distributed systems, Data Consistency, DCR graphs, DCR Choreographies, Business processes

Existe pré-acordo com algum aluno: Não

Nome do aluno:

72 Data consistency in distributed business processes.

Proponente(s): João Costa Seco

Contacto: Joao.seco@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

Swarm architectures depict software based on the decentralised coordination of multiple participants, making them well-suited for large-scale distributed systems. Swarms concurrently execute processes in highly distributed environments, where participants can dynamically join and leave the system. Global behaviours emerge from the local interactions between participants. Swarm programming requires a flexible programming model that allows the definition of concurrent processes executed by multiple participants.

In this thesis topic, we propose the study of data consistency in the soundness of a business process or workflow. The study will be focusing on Dynamic Condition Response (DCR) choreographies that define data dependent behaviours in swarms. DCR choreographies define messages and data exchanged between participants and set the control-flow constraints beyond basic data dependencies representing the system's business logic.

The work includes the development of tools to validate and project DCR choreographies to local processes running on the Actyx or Babel platforms, runtime support frameworks for decentralised systems.

This topic can be awarded a scholarship from the TaRDIS project.

Keywords: Programming languages, Event-driven programming, Distributed systems, Data Consistency, DCR graphs, DCR Choreographies, Business processes

Existe pré-acordo com algum aluno: Não

Nome do aluno:

73 [SI-MORENA] Automatização de tarefas para deteção do estado de faixas de gestão de combustível para incêndios

Proponente(s): Carlos Viegas Damásio, João Moura Pires

Contacto: cd@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

O objetivo desta dissertação é implementar um sistema que a partir de dados de satélite determine o estado das faixas de gestão de combustível para incêndios (FGCI). O sistema terá de obter imagens de satélite a nível nacional, processá-las e aplicar algoritmos de aprendizagem automática previamente desenvolvidos podendo contudo ser melhorados.

Atualmente, existe uma implementação suportada nos serviços da Google Cloud mas pretendemos com este trabalho realizar uma implementação independente.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

74 [SI-MORENA] Visualização para o estado de faixas de gestão de combustível para incêndios

Proponente(s): Carlos Viegas Damásio, João Moura Pires

Contacto: cd@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
No âmbito do projeto Floresta Limpa, pretendemos desenvolver um conjunto de visualizações a disponibilizar na Web para se percepcionar o estado de faixas de gestão de combustíveis para incêndios. As FGCI são em grande número e de diversos tipos necessitando de técnicas específicas para cada uma delas. Adicionalmente, será necessário considerar os aspetos temporais pois elas evoluem ao longo do tempo quer com intervenções humanas quer naturalmente. Estas visualizações poderão ser utilizadas pelas autoridades para efeitos de planeamento de limpeza ou fiscalização do seu estado.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

75 [SI-MORENA] Disponibilização de repositório público de dados de Faixas de Gestão de Combustíveis para Incêndiso

Proponente(s): Carlos Viegas Damásio, João Moura Pires

Contacto: cd@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
No âmbito do projeto Floresta Limpa, pretendemos desenhar um repositório de informação pública/privada com dados anotados de estado das faixas de gestão de combustível para incêndios (FGCI). Estes dados deverão ser acessíveis através de uma API disponibilizando dados em formatos abertos e anotados recorrendo a ontologias e/ou taxonomias para facilitar a sua interoperabilidade.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

76 [SI-MORENA] Modelos para a deteção do estado de vegetação em torno de habitações

Proponente(s): Carlos Viegas Damásio, João Moura Pires

Contacto: cd@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
O projeto Floresta Limpa tem acesso a um conjunto. de dados a nível nacional com dados de vegetação em torno de habitações. O objetivo consiste em construir modelos recorrendo a técnicas de aprendizagem automática para permitir o mapeamento rápido do estado da vegetação em torno de habitações a nível nacional. Esperam-se que sejam construídos dois modelos, um baseado apenas na informação de uma ou relativamente poucos dados de imagens de satélite (para rápida determinação do estado) e outro recorrendo a dados de séries longas de dados de satélite para melhorar a informação contextual.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

77 [SI-MORENA] Fiscalização por deteção remota para a Câmara Municipal de Almada

Proponente(s): Carlos Viegas Damásio, João Moura Pires

Contacto: cd@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
A Câmara Municipal de Almada pretende efetuar um conjunto de operações de fiscalização do ambiente recorrendo a imagens de fotografia aérea/dados satélite que necessitarão de técnicas de inteligência artificial. Em particular, pretende detetar-se construções ilegais, despejo de lixo em áreas não autorizadas, poços, piscinas ou mesmo a retirada de caixotes de lixo dos locais designados para o efeito.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

78 [SI-MORENA] Deteção de plântulas de infestantes a partir de imagem

Proponente(s): Carlos Viegas Damásio

Contacto: cd@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
O objetivo deste trabalho é continuar o desenvolvimento de uma ferramenta Android para o levantamento de campo de infestantes no seu estado inicial de desenvolvimento (plântulas). Atualmente a aplicação identifica 10 tipos de infestantes das regiões do OEste e Ribatejo, mas pretendemos expandir os modelos atuais com mais espécies. Idealmente, gostaríamos de poder efetuar este processo a partir de imagens obtidas a partir de vídeo para facilitar a obtenção de dados e construção de novos datasets.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

79 [SI-MORENA] Operacionalização de serviço de avisos agrícolas a partir de dados de satélite

Proponente(s): Carlos Viegas Damásio

Contacto: cd@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
O objetivo deste trabalho é operacionalizar um sistema de avisos agrícolas baseado em dados de satélite de temperatura ao nível do solo e de previsão meteorológica. Os modelos já existem mas necessitam de ser otimizados para implementação a uma escala nacional ou continental. Em particular, o aluno terá de trabalhar com grandes volumes de dados e propor uma arquitetura para suportar um sistema deste tipo.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

80 Autoformalization

Proponente(s): Nuno Marques (DI, NOVA-FCT) e João Araújo (DM, NOVA-FCT)

Contacto: nmm@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Proofs written by mathematicians are considered informal since they do not enable fully automated checking (by a program for instance). This implies that they can easily contain errors. This has given rise to the field of formalized mathematics and automated theorem provers and proof assistants, which enable fully formalized and automated proof checking. However, these tools require considerable expertise to be used and majority of the existing mathematical research has not been yet formalized and therefore we are interested in methods enabling going from informal mathematical texts to the formal. This process is known as autoformalization.
This dissertation will develop a system that will take on the input informal mathematical text, use large language models to produce a formal representation, and use an automated theorem prover to validate the correctness of the translation. As a source of data, we consider articles from arXive, which have the advantage of being publicly available together with their original latex code and therefore easier to analyze. For automated theorem proving we will use the well known tools such as prover9, vampire, cvc5, or z3.

Existe pré-acordo com algum aluno: Sim

Nome do aluno: Bernardo Atalaia

81 Desenvolvimento de características de acessibilidade e storytelling em aplicação interativa para suporte ao turismo de passeio

Proponente(s): Armanda Rodrigues, Cédric Grueau (Instituto Politécnico de Setúbal)

Contacto: a.rodrigues@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

No turismo de passeio (a pé ou em bicicleta), Portugal apresenta-se como uma excelente alternativa aos países europeus mais frios, uma vez que o clima ameno permite viajar a pé ou de bicicleta praticamente todo o ano. Com o objetivo de promover este turismo durante todo o ano é importante dar a conhecer, disponibilizando dados relevantes, como guias e mapas, para que durante o seu passeio a pé ou de bicicleta tenham a melhor experiência possível.

Existem vários sites e aplicações disponíveis e focadas neste assunto, mas não existe, pelo menos em funcionamento, um serviço integrado que utilize normas de informação geográfica para disponibilização de acesso aos caminhos. No DI deu-se início ao desenvolvimento de uma metodologia de integração de dados atendendo à melhor forma como estes podem ser estruturados, e também a que serviços e normas devem ser usados, nomeadamente OGC (`https://www.ogc.org/`).

Está também em desenvolvimento, no âmbito de uma dissertação de mestrado em elaboração, uma aplicação interativa para acesso, pesquisa e percurso dos caminhos a disponibilizar pelo repositório, com o objetivo de estudar de forma detalhada as necessidades de usabilidade deste tipo de aplicações. Nesta dissertação, pretende-se continuar o desenvolvimento desta aplicação focando em características de acessibilidade e storytelling. O trabalho está a ser desenvolvido em colaboração com Instituto Politécnico de Setúbal e com o Departamento de Ciências do Desporto de Universidade da Beira Interior.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

82 Adaptação Inteligente da Frequência de Leitura em Microcontroladores

Proponente(s): Prof. Nuno Marques (DI, NOVA-FCT) & Doutor Eng. João Marcelino (LNEC)

Contacto: nmm@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Esta tese propõe o desenvolvimento de metodologias para a adaptação automática e inteligente da frequência de leitura em sistemas de microcontroladores utilizando o protocolo MQTT. A pesquisa será conduzida no contexto de aplicações IoT, onde a eficiência na aquisição e transmissão de dados é essencial. Em particular, a tese focará na utilização do algoritmo Ubiquitous Self-Organizing Map (`https://github.com/nmm-fct-unl/ubiSOM2023`), uma variante projetada para análise de grandes fluxos de dados. Adicionalmente valida-se como o UbiSOM permitirá a representação e projeção multidimensional dos dados de sensores e assim reforçar o seu potencial como ferramenta para a exploração interativa e visualização de dados em tempo real.
Objetivo da tese: O objetivo principal é desenvolver e implementar uma metodologia baseada no UbiSOM para ajustar automaticamente a frequência de leitura dos sensores conectados a microcontroladores, otimizando o uso de recursos e melhorando a precisão na deteção de eventos importantes. A programação dos microcontroladores para comunicação via MQTT será realizada, juntamente com a aplicação de técnicas de projeção multidimensional para análise interativa de dados. Serão analisadas soluções que permitirão ao utilizador visualizar os dados de diferentes perspetivas e ajustar dinamicamente as configurações de leitura, oferecendo a escolha entre leituras com frequência constante ou adaptativa.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

83 Sistema Automático para Análise de Imagens e Anotação de Características em Pavimentos Rodoviários

Proponente(s): Prof. Nuno C. Marques (DI, NOVA-FCT) e Doutor Eng. João Marcelino (LNEC)

Contacto: nmm@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

Os pavimentos rodoviários são cruciais para a conexão entre aglomerados populacionais e a integração de países através de redes de transporte. A monitorização desses pavimentos é tradicionalmente feita através de inspeções visuais por técnicos especializados, que recolhem dados para caracterizar e avaliar o seu estado. Esta tese propõe o desenvolvimento de um sistema automático para análise de imagens e anotação de características dos pavimentos rodoviários. O sistema deve ser capaz de receber imagens digitais dos pavimentos, realizar a detecção e análise de diversas características no pavimento e permitir a anotação e caracterização dos dados por especialistas. A integração com ferramentas de sistemas de informação geográfica (GIS) como o QGIS possibilitará a criação de uma base de dados georreferenciada, facilitando a avaliação da evolução das condições dos pavimentos entre campanhas de inspeção sucessivas.

Objetivo da tese: O objetivo é criar uma ferramenta que utilize técnicas de processamento de imagem e aprendizagem de máquina para a detecção e análise automática das condições dos pavimentos rodoviários. A ferramenta permitirá a anotação e caracterização detalhada das imagens por peritos, facilitando a monitorização contínua e a avaliação da evolução das condições do pavimento. Esta solução visa complementar os métodos existentes de inspeção visual, proporcionando um meio mais rápido e eficiente de identificar e avaliar problemas, e, consequentemente, ajudando na definição de medidas adequadas de conservação e manutenção para garantir a segurança e conforto na circulação rodoviária.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

84 Prototype of an Investment Support System Using Large Language Models and Factor-Based Simulation

Proponente(s): Nuno C. Marques

Contacto: nmm@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição

This thesis aims to develop a prototype for investment support by integrating Large Language Models (LLMs) with advanced data visualization and factor-based investment simulation tools. The primary focus is on analyzing fundamental indicators that measure the intrinsic value of companies.

To achieve this, the project will utilize Self-Organizing Maps (SOMs)—a type of artificial neural network used for clustering and visualizing high-dimensional data. SOMs will be integrated with Qrumble - a Python tool for factor-based investment simulation tool based on historical data.

Together with the LLM, the SOMs will reduce and project the most relevant indicators from both factor based investment and company reports into graphical maps, facilitating the identification of similar contexts. These similar contexts can then be used by a Retrieval Augmented Generation (RAG) system to make the LLM generate relevant responses for investors and identifying investment strategies based on investor concerns and similar related historical patterns.

Thesis Objective:

The objective of this thesis is to develop a prototype that integrates advanced data visualization with RAG system. The key components and goals of the prototype include:

Data Visualization: Utilize advanced tools to visually represent financial indicators.

RAG System: Implement a Retrieval Augmented Generation system to combine the analysis of financial reports and statements from a subset of companies with the automated generation of insights through Large Language Models (LLMs). This system will be capable of processing large volumes of data, including text and historical data, using various financial indicators. By creating embeddings, the system will transform the data into a format suitable for analysis.

Dialogue System: Use the LLM to develop a dialogue system that utilizes the processed data and embeddings to present detailed responses about the financial health of companies and market trends.

Investment Recommendations: Evaluate the integration of these technologies to determine if the resulting tool is useful for investors. The LLM should generate well-founded investment recommendations together with intuitive data visualization.

The thesis will thoroughly analyze the integration of these technologies and assess whether this combination can result in a practical and valuable tool for investors.

Links:

`https://github.com/nmm-fct-unl/ubiSOM2023`
`https://github.com/oraibalmegdadi/AI-Based-Multiformat-Document-Chatbot`
`https://qrumble.notion.site/QRUMBLE-3a4b47606b9c4ffd8f40bac0553a97d5?pvs=4`

Existe pré-acordo com algum aluno: Não

Nome do aluno:

85 Exploiting Time in Large Vision and Language Models

Proponente(s): David Semedo

Contacto: df.semedo@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Large Vision and Language Models (LVLMs) such as LLaVa-* and CogVLM, can flexibly address multiple vision and language tasks, in a robust manner. However, it is still unclear how well these models can frame real-world information in time. This is of particular importance in assessing their effectiveness in real-world image-understanding tasks, such as generating a description of an image of an event or predicting a sequence of events.

This will be approached in a two-fold manner: implicitly, by identifying attention-hubs through linear probing, where temporal cues are more salient, and explicitly, by investigating temporal model-prompting schemes to elicit the time dimension, while jointly attending to images and text.

The thesis outcomes will then be evaluated using event-oriented publicly available benchmarks.

Existe pré-acordo com algum aluno: Sim

Nome do aluno: Henrique Paz

86 Video QA For Surveillance with Large Vision and Language Models

Proponente(s): David Semedo

Contacto: df.semedo@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
In the surveillance scenario, understanding and automatically extracting metadata from video footage is paramount to identify potential situations where public safety is compromised, either in real-time or a posteriori analyses. Recent Large Vision and Language Models (LVLMs) have demonstrated significant effectiveness in understanding and answering questions about general videos. However, the surveillance domain brings novel challenges, ranging from low-quality video streams, multiple video perspectives, harsh lighting conditions, and most importantly, novel target action-patterns.
This thesis will build upon the group's work on Surveillance-oriented LVLMs and exploit their capabilities to support both systematic and sporadic Video Question Answering (Video QA). State-of-the-art Video QA approaches based on LVLMs will be investigated and complemented with efficient video moment-of-interest pinpointing strategies. To evaluate the work, the student will set up a thorough and diverse QA benchmark.
In collaboration with UBI and DeepNeuronic.

Existe pré-acordo com algum aluno: Sim

Nome do aluno: Pedro Domingos

87 Clustering-based Indexing of Faces in an Stream of Images

Proponente(s): Hervé Paulino

Contacto: herve.paulino@fct.unl.pt

Tipo de dissertação: Dissertação científica

Descrição
Nesta tese pretende-se implementar um sistema capaz de agrupas caras (na forma de clusters) a partir de um fluxo de imagens. O processo fará uso de machine learning não supervisionado para ir construído clusters das caras das pessoas que vão surgindo num fluxo de imagens, sem saber de antemão quem serão estas pessoas.
O fluxo de imagens é gerado por múltiplos dispositivos móveis. Por exemplo várias utilizadores num evento social: festa, concerto, jogo. O resultado do clustering será utilizado para criar um índice (trabalho já realizado) que é partilhado com os telemóveis que geraram o fluxo das imagens (trabalho também já realizado). Para afinar o processo de clustering pretende-se desenvolver com um mecanismo de feedback, a ser dado pelos utilizadores (por exemplo, a imagem que recebi não tem a cara que eu estava à espera). Tal feedback será incorporado na gestão dos clusters, nomeadamente, na sua dimensão, cisão fusão, entre outros.

Existe pré-acordo com algum aluno: Não

Nome do aluno:

Supervisor Index