

Propostas de Tese para o Mestrado em Engenharia Informática (MEI)

Grupo de Processamento de Linguagens
DI/CAgoritmi
Universidade do Minho

(Pedro Rangel Henriques)
ano lectivo 24/25

1 Um sistema Web para gestão dos processos de doutoramento no IE

Supervisor: Pedro Rangel Henriques (prh@di.uminho.pt; 968412287) + Beatriz Oliveira Pereira (IE)
Área: Sistemas de Informação, Aplicações Web, Bases de Dados

Resumo:

Neste projeto, solicitado pela Presidência do Instituto de Educação (IE), pretende-se desenvolver uma aplicação Web para apoio ao Conselho Científico que possibilite a gestão dos processos de apresentação de propostas de tese, candidatura ao doutoramento e receção e registos dos relatórios anuais que permitam fazer o tracking de cada processo.

2 Um recreio na web para treino personalizado da Atenção

Supervisor: Pedro Rangel Henriques (prh@di.uminho.pt; 968412287) + Cristiana Araújo (decristianaaraujo@hotmail.com)
+ João José Cerqueira (jcerqueira@med.uminho.pt)
Área: Neuroeducação

Resumo:

A incapacidade em manter a atenção sustentada é cada vez mais uma dificuldade manifestada por jovens e adultos, com implicações no seu dia a dia. Esta dificuldade pode, algumas vezes, ser resultado de uma PHDA (Perturbação de Hiperatividade/Défice de Atenção) ou outra patologia, mas é, na maioria dos casos, uma aptidão que pode ser mais ou menos desenvolvida.

Neste projeto pretende-se desenvolver uma ferramenta web que contemple um conjunto de desafios (exercícios) específicos para trabalhar a atenção, de forma a que os utilizadores consigam aumentar progressivamente a sua capacidade. A ferramenta deve contemplar testes de diagnóstico (para avaliar a capacidade de atenção antes do treino) e testes de avaliação (para avaliar a capacidade de atenção depois do treino).

O objetivo final é usar a ferramenta na população em geral, mas também em ambiente escolar e clínico.

3 Plataforma de Caracterização e Visualização de Recursos de Aprendizagem com base na OntoCnE

Supervisor: Pedro Rangel Henriques (prh@di.uminho.pt; 968412287) + Cristiana Araújo (decristianaaraujo@hotmail.com)
Área: OntoCnE/PC; Treino do Pensamento Computacional

Resumo:

No âmbito da tese de doutoramento de Cristiana Araújo, focada na definição do conceito de Pensamento Computacional (PC) e no seu treino, desenvolveu-se uma ontologia designada OntoCnE. Esta ontologia composta por 4 camadas

tem como objetivos: descrever o domínio do Pensamento Computacional e da Programação de Computadores (camada 1); definir o que trabalhar em cada ano escolar e o seu nível de profundidade (camada 2); definir que Recursos de Aprendizagem (RA) que podem ser utilizados para desenvolver o PC (camada 3); definir quais as guias da Neuroeducação que cada RA contempla (camada 4).

Neste projeto pretende-se desenvolver uma ferramenta Web que permita caracterizar os RA com base na OntoCnE. Esta caracterização vai permitir descrever o RA, definir que conceitos de Pensamento Computacional e da Programação de Computadores ele trabalha, em que anos escolares pode ser utilizado e que guias da Neuroeducação contempla. A ferramenta deve contemplar uma interface de pesquisa em linguagem quase natural (uma DSL a definir pelo candidato) que permita ao utilizador final (professores/formadores) encontrar com eficácia e simplicidade os RA que mais se adequam aos conceitos que quer treinar/ensinar dentro do contexto em que vai trabalhar.

O objetivo final é usar a ferramenta em ambiente escolar, ou seja, será utilizada por professores de Informática e de outras áreas.

4 Explorer4GUCO: Um conjunto de ferramentas para analisar um corpus online sobre a Guerra na Ucrânia

Supervisor: Pedro Rangel Henriques (prh@di.uminho.pt; 968412287) + Cristiana Araújo (decristianaaraujo@hotmail.com)
Área: Data Analysis; Língua Natural; Análise de Sentimentos; Corpus

Resumo:

GUCO é um corpus sobre a Guerra na Ucrânia, que foi construído com notícias de jornais online e do Reddit e respetivos comentários. O GUCO pode ser consultado em: <https://guco.epl.di.uminho.pt/corpus>.

Neste projeto pretende-se disponibilizar ferramentas que facilitem a análise do Corpus, como por exemplo: linha temporal; análise do tipo de discurso presente nos comentários; construção de narrativas com os factos extraídos (*storytelling*); estudar o impacto económico no mundo; etc.

Para isso devem ser utilizadas técnicas de Processamento de Língua Natural e Análise de Sentimentos para anotar o Corpus antes de se proceder à extração de dados e respetiva análise estatística.

O objetivo final destas ferramentas é serem utilizadas por investigadores das Humanidades Digitais, nomeadamente, Linguistas, Sociólogos, Historiadores, etc.

5 Anotação automática de Documentos

Supervisor: Pedro Rangel Henriques (prh@di.uminho.pt; 968412287) + Maria João Varanda Pereira (mjoao@ipb.pt)
Área: Ontologias, Linguagens de Anotação, Geração e Automatização

Resumo:

Considere que está a trabalhar um determinado domínio de conhecimento (neste caso para criar um arquivo de postais familiares) e que já anotou manualmente em XML alguns textos, obtendo a anotação que se exemplifica a seguir

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE postais SYSTEM "postais.dtd">
  <postais>
    <postal foto="praia de Albufeira">
      <abertura>
        Postal enviado a <data norma="20230531">2023-maio-31</data>
        pelo <remetente>Pedro Henriques</remetente> para sua
        <destinatario><nome>Alice Pinto da Silva,</nome>
        <morada>Porto</morada></destinatario>.
      </abertura>
      <mensagem>
        Querida Tia, cá estamos chegados para férias....
        Beijinhos do seu sobrinho Pedro.
      </mensagem>
```

```
</postal>
</postais>
```

O objetivo deste projeto, desafiante e atual, é ser capaz de ler um texto no mesmo domínio e do tipo do anterior, como o exemplo seguinte

```
Postal datado de 05 de junho de 2020 enviado a António Rangel
que vive na Foz do Douro.
Caro António,
aconselho sinceramente a vir aqui...
Um abraço, Pedro.
```

e automaticamente produzir uma anotação congruente com a anterior, devendo obter-se algo do género:

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE postais SYSTEM "postais.dtd">
  <postais>
    <postal foto="">
      <abertura>
        Postal datado de <data norma="20200605">05 de junho de
        2020</data> enviado a <destinatario><nome>António Rangel</nome>
        <morada>Foz do Douro</morada></destinatario>.
      </abertura>
      <mensagem>
        Caro António, aconselho sinceramente a vir aqui...
        Um abraço, Pedro.
      </mensagem>
    </postal>
  </postais>
```

A construção deste anotador automático deve tomar como ponto de partida uma ontologia do domínio de trabalho e uma coleção de documentos corretamente anotados.

Pretende-se explorar o recurso a dicionários e a ferramentas tradicionais de PLN (que implementem o habitual pipeline da tokenização à análise morfológica de textos em português) e comparar com a utilização de modelos de Machine Learning tipo LLM com base no seu treino adequado.

6 Obtenção automática de coquetéis a partir de repositórios abertos (GitHub etc.)

Supervisor: Pedro Rangel Henriques (prh@di.uminho.pt; 968412287) + Alvaro da Costa Neto (nepheus.br@gmail.com) + Maria João Varanda Pereira (mjoao@ipb.pt)

Área: Linguagens de Programação, Escrita de código, Ambientes de apoio incrementais

Resumo:

Neste projeto pretende-se desenvolver um sistema que consistirá em um *Web-crawler* e um Tradutor. Seu objetivo será navegar por repositórios de código-fonte abertos e identificar os coquetéis de programação (conjunto de linguagens, bibliotecas, frameworks e ferramentas) que são usados em cada repositório. O sistema deverá construir, para cada coquetel obtido, um cartão de identidade de coquetel (CIC), tendo por base a ontologia de coquetéis de programação desenvolvida por Alvaro Costa Neto no âmbito do seu trabalho de doutoramento. Esta dita ontologia define um modelo conceitual para descrever o domínio dos coqueteis de programação, o qual não é de todo um universo de conhecimento muito divulgado. Por cada coquetel recuperado do repositório, o CIC a criar será uma instância dessa ontologia.

O sistema final será aplicado e testado em pelo menos um sítio de hospedagem de repositórios (como GitHub, por exemplo). Este sistema fará parte de um projeto maior que foca seus esforços em estudar os coquetéis de programação, incluindo uma tese de doutoramento em curso, já referida.

Como resultados adicionais, pretende-se obter dados estatísticos que possam gerar conclusões diversas, como ingredientes mais usados, combinações mais comuns e possíveis padrões de correlação.

7 Rede social de aprendizado de programação

Supervisor: Pedro Rangel Henriques (prh@di.uminho.pt; 968412287) + Alvaro da Costa Neto (nepheus.br@gmail.com)
Área: Linguagens de Programação, Escrita de código, Ambientes de apoio incrementais

Resumo:

Este projeto consistirá na implementação de um sistema de rede social especializado em apoiar e estimular os estudantes de programação, em diversos níveis, áreas e domínios.

Deve contemplar em um primeiro momento, um estudo sobre as principais características psicossociais que atraem usuários para as redes sociais da atualidade. Com este levantamento e um panorama de quais requisitos perfazem redes sociais de sucesso, uma solução com mecanismos de interação dirigidos para o incentivo, apoio e cooperação entre alunos será implementado por meio de um sistema Web com interface simples.

Seu uso será testado com alunos dos cursos de programação da instituição. Mais do que um sistema funcional, espera-se obter um levantamento bibliográfico estruturado contendo fatores que mais podem influenciar alunos a participarem de uma comunidade de troca e aprendizado de programação, resultados estes que poderão ser usados para além dos limites do projeto.

8 Linguagem de programação baseada em Memes

Supervisor: Pedro Rangel Henriques (prh@di.uminho.pt; 968412287) + Alvaro da Costa Neto (nepheus.br@gmail.com)
Área: Linguagens de Programação, Ambientes de apoio incrementais

Resumo:

Memes, assim como qualquer outro símbolo comunicativo, podem compor novas formas de comunicação desde que seus significados sejam compartilhados entre os locutores.

A ideia por trás deste projeto seria criar uma linguagem visual de programação simples em que as palavras reservadas e operações seriam representadas por Memes. Estes memes serão escolhidos primeiro manualmente pelos engenheiros de linguagens.

Se tudo correr bem, seria interessante que esse processo seletivo fosse realizado automaticamente por uma rede neural treinada para associar um meme (escolhido de uma coleção, ou proposto pelo usuário) a cada construtor de uma linguagem de programação (atribuição, soma, subtração, entrada, saída, condicional, laço etc.).

Assim o programador montaria seu código-fonte encadeando esses memes.

9 Sistema interativo de construção de frases derivadas de gramáticas

Supervisor: Pedro Rangel Henriques (prh@di.uminho.pt; 968412287) + Alvaro da Costa Neto (nepheus.br@gmail.com)
+ Maria João Varanda Pereira (mjoao@ipb.pt)

Área: Linguagens de Programação, Escrita de código, Ambientes de apoio incrementais

Resumo:

Neste projeto, de grande interesse para apoio a professores e alunos nos cursos de Processamento de Linguagens (PL), pretende-se gerar automaticamente ficheiros de input (programas-fonte, ou dito por outras palavras, frases de uma dada linguagem) para testes dos parsers que os alunos desenvolvem em Python a partir da GIC que eles próprios escrevem. O utilizador do sistema, para além da gramática, daria um conjunto de palavras possíveis para os terminais da gramática e depois seriam geradas variantes de frases válidas e não-válidas da linguagem definida por essa gramática.

10 Aplicação de quadro interativo para programação

Supervisor: Pedro Rangel Henriques (prh@di.uminho.pt; 968412287) + Sofia Santos (sofiarsantos31@gmail.com)

Área:

Resumo:

Com o surgimento de novas tecnologias, como ecrãs ultrafinos ou ecrãs táteis, vários dispositivos têm sido concebidos para tirar proveito destas inovações. Um deles é o quadro interativo, cada vez mais comum nas escolas, universidades e escritórios. Estes ecrãs combinam a versatilidade de um quadro branco com as funcionalidades de um projetor, permitindo projetar e interagir com o conteúdo que está a ser projetado através do próprio ecrã, como um 'tablet' gigante.

Nas aulas de programação, estes quadros deixam de ser tão úteis. Por exemplo, se um professor estiver a projetar o seu editor de texto ou terminal, não consegue desenhar diretamente no quadro, precisa de usar um quadro branco à parte ou trocar o programa que está a projetar, perdendo-se assim o conteúdo do editor de texto ou terminal.

Pretende-se assim, com este projeto, conceber e implementar uma ferramenta que permita a escrita, compilação e execução de código-fonte, para além de permitir desenhar livremente 'por cima' do código ou do seu output. Por outras palavras, o objetivo deste projeto é o desenvolvimento de uma aplicação que seja equivalente ao ato de projetar um editor de texto e/ou terminal num quadro branco normal e de desenhar no mesmo com marcadores.

Alguns quadros interativos já possuem esta funcionalidade, mas para os que não possuem, este programa seria extremamente útil. Para além disso, este programa poderia incluir funcionalidades adicionais, como a possibilidade de exportar todo o conteúdo do 'quadro' (código e desenhos) e opções de colaboração entre professores e estudantes.

11 Enhancing transpilers with machine learning techniques

Supervisor: Pedro Rangel Henriques (prh@di.uminho.pt; 968412287) + Tiago Baptista (tiago96baptista@gmail.com)

Área: Transpilação, Compilação, ML, LLM

Resumo:

This thesis aims to explore the integration of machine learning techniques with traditional Abstract Syntax Tree (AST) based transpilers to improve the accuracy and efficiency of cross-language code translation. By leveraging the strengths of machine learning models in handling complex and context-dependent translation.

The thesis aims to study and propose a hybrid approach that combines rule-based AST transformations with machine learning algorithms, evaluating its performance, and comparing it with conventional methods.

This work seeks to demonstrate the potential of machine learning in advancing the field of automated code translation.

To start the research, the following initial documentation is recommended:

<https://tomassetti.me/language2language-transformers-machine-learning-to-build-transpilers/>

<https://dl.acm.org/doi/pdf/10.1145/3660778>

<https://www.diva-portal.org/smash/get/diva2:1868611/FULLTEXT01.pdf>

12 Ferramenta para análise completa de Gramáticas em ANTLR

Supervisor: Alberto Simões / Checkmarx (Alberto.Simoes@checkmarx.com) + Pedro Rangel Henriques (prh@di.uminho.pt; 968412287)

Área: Gramáticas, Análise Estática/Dinâmica de Código

Resumo:

Desenvolver uma ferramenta, C#, para o teste de gramáticas ANTLR, em run-time, com visualização das árvores de parsing, análise de estatísticas de parsing (lookaheads, ambiguidades, etc) e com suporte de predicados em C#.

Sendo possível, a ferramenta deverá tentar replicar algumas funcionalidades de outras ferramentas semelhantes, como a indicação do tipo de cada token e regra da gramática.

13 Vulnerabilidades em ferramentas de CI/CD no contexto de SSCS

Supervisor: Alberto Simões / Checkmarx (Alberto.Simoes@checkmarx.com) + Pedro Rangel Henriques (prh@di.uminho.pt; 968412287)

968412287)

Área: *Software Supply Chain, Security, Detecção de Vulnerabilidades*

Resumo:

No contexto de 'Software Supply Chain Security' pretende-se investigar os tipos de vulnerabilidades existentes em ferramentas de CI/CD como Jenkins, CircleCI, GitHub Actions, TeamCI, Buddy, Travis, Bamboo, etc., quer a nível de ações de compilação ou de plugins.

Posteriormente, pretende-se desenvolver uma aplicação que analise a configuração de uma instalação CI/CD e indique as potenciais vulnerabilidades.

14 Modelos de LLM e vulnerabilidades induzidas

Supervisor: Alberto Simões / Checkmarx (Alberto.Simoes@checkmarx.com) + Pedro Rangel Henriques (prh@di.uminho.pt; 968412287)

Área: *Software Supply Chain, Security, LLM*

Resumo:

No contexto de 'Software Supply Chain Security' (SSCS) pretende-se investigar em que medida a integração de modelos de linguagem de larga escala (LLM) em aplicações pode ser um risco de vulnerabilidade para essa aplicação. Será necessário uma análise detalhada de toda a documentação existente nesta área, para perceber as diferentes formas com que um LLM pode comprometer um sistema.

Posteriormente, pretende-se o desenvolvimento de uma aplicação que seja capaz de detetar problemas/vulnerabilidades em modelos LLM.

15 Smart-Doc: AI-Powered Medical Education through Diagnostic Interactions

Supervisor: Nuno Rodrigues / IPCA (nfr@ipca.pt) + Pedro Rangel Henriques (prh@di.uminho.pt; 968412287)

Área: *AI, ML, LLMs, Healthcare SW*

Resumo:

The training of medical doctors is a critical aspect of healthcare, directly impacting the quality of patient care. Current educational methods often rely on traditional lectures, textbooks, and limited clinical exposure, which may not fully prepare medical trainees for the complexities of real-world medical practice. To address these challenges, this project proposes the development of a novel AI medical tutor designed to enhance medical education through the analysis of exemplary medical cases. By leveraging a Large Language Model (LLM) fine-tuned with data from reputable sources documenting patient case studies, this AI tutor aims to provide immersive and interactive training experiences. The AI tutor will operate in multiple modes, simulating patient interactions and analytical training scenarios to assess and improve trainee doctors' clinical skills comprehensively.

The primary objectives of this Master of Science proposal are:

- **Develop an AI medical tutor:** Create an AI-powered tool that can simulate patient interactions and provide analytical training based on detailed medical case studies.
- **Enhance anamnesis skills:** Enable trainee doctors to improve their patient history-taking and data collection skills through interactive simulations.
- **Facilitate diagnostic reasoning:** Provide trainees with comprehensive patient data to analyze and formulate differential diagnoses and treatment plans.
- **Incorporate advanced AI techniques:** Utilize Retrieval Augmented Generation (RAG) and Directed Acyclic Graphs (DAG) to enhance the AI tutor's decision-making and justification capabilities.
- **Evaluate the effectiveness:** Assess the AI tutor's impact on medical training through feedback and performance evaluation of trainees.

16 Connected Lives XR

Supervisor: Nuno Rodrigues / IPCA (nfr@ipca.pt) + Pedro Rangel Henriques (prh@di.uminho.pt; 968412287)
Área: XR virtual Social Club; AI virtual Assistants; Older Adults

Resumo:

Europe is experiencing a significant demographic shift, with an increasing number of older adults projected to rise substantially by 2050. This demographic change poses challenges to the current healthcare and social assistance services, which often rely on the institutionalization of older adults. While aging in place offers a preferable alternative, it can lead to loneliness and social isolation. The Connected Lives XR (CLives-XR) project aims to address these issues by creating an XR Social Club that blends virtual and real-world interactions to enhance social connections among older adults.

The primary objectives of this Master of Science proposal are:

- To design a virtual social space within the XR Virtual Social Club that resembles a warm and welcoming hotel lounge.
- To implement an AI virtual assistant specialized in providing social interaction behavior support for older adults.
- To evaluate the effectiveness of the virtual social space and the AI assistant in reducing loneliness and enhancing social connections among older adults.

17 Intelligent Optimized Resource Management for Nursing Homes

Supervisor: Nuno Rodrigues / IPCA (nfr@ipca.pt) + Pedro Rangel Henriques (prh@di.uminho.pt; 968412287)
Área: AI/ML; Predictive Analysis; Inventory Management; Procurement

Resumo:

This project aims to develop a cutting-edge, intelligent web application specifically designed for the efficient management of nursing homes (a 'nursing home' is a facility for the residential care of older people, senior citizens, or disabled people). The application focuses on automating the procurement process by intelligently managing the inventory of consumables based on the profile and needs of the residents. By integrating AI models, the application will not only streamline the inventory management but also predict future requirements, and compare it with stock supplies, thus ensuring that the nursing homes operate efficiently, reduce waste, and provide the highest quality of care to the residents.

The primary objectives of this Master of Science proposal are:

- **Develop an Intuitive Multi-Tenant Platform:** Create a user-friendly web interface that allows multiple nursing homes to manage their operations independently under a single umbrella application;
- **Automate Inventory Management:** Implement intelligent algorithms to automate stock registration, tracking, and management of consumables;
- **Integrate AI for Predictive Analysis:** Develop pre-trained AI models capable of learning consumption patterns for each resident and predicting future needs to ensure timely procurement;
- **Optimize Procurement Processes:** Utilize the predictive insights to aggregate the needs of all residents, compare with current stock, and automate the procurement process to prevent shortages;
- **Ensure Scalability and Security:** Design the system to be scalable to accommodate an increasing number of nursing homes and secure to protect sensitive resident data.