

# 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 19/20

## 1 Geração automática de Textos para contar histórias de vida

*Supervisor: Pedro Rangel Henriques + Alda Lopes Gançarski*  
*Área: Ontologias, PLN-gramaticas para LN, Museus virtuais*

### Resumo:

O objectivo desta tese é enriquecer a visita conceptual a um museu virtual relacionado com a vida social (como, p.ex., o Museu da Emigração ou o Museu da Pessoa) com geração automática de texto de forma a 'contar histórias' relativas à vida ou emigração de pessoas ou famílias em certas regiões.

A geração deve ser baseada em regras de transformação dos componentes das ontologias em estruturas gramaticais. Estas regras de transformação são específicas e pré-definidas. Cada fragmento da ontologia do ME é assim transformado na árvore de sintaxe correspondente ao texto a gerar. De forma a gerar um texto próximo de um estilo literário, as regras de transformação serão enriquecidas, p. ex., com a capacidade de evitar repetições em frases resultantes de fragmentos de ontologia semelhantes.

Para aumentar a qualidade do texto do ponto de vista do utilizador, o sistema deve propor-lhe a re-escrita ao nível do texto (o que implica a atualização da árvore gramatical correspondente, mecanismo bottom-up), da árvore gramatical ou da própria ontologia.

## 2 Estudo dos Expostos de Fafe no Século XIX

*Supervisor: Pedro Rangel Henriques + Mónica Guimarães + Cristiana Araújo*  
*Área: Humanidades Digitais*

### Resumo:

Recorrendo ao manancial de informação disponível no Arquivo Municipal de Fafe, pretende-se recuperar a informação disponível sobre as crianças abandonadas na roda (expostos) ao longo do sec. XIX no concelho de Fafe e concelhos limítrofes, de modo a poder-se criar maior conhecimento sobre este fenómeno social tão comum em séculos anteriores a 1900.

Nesta proposta de tese pretende-se investigar as várias fontes a usar, criar um repositório integrado para as conter em formato digital, desenvolver as ferramentas suportadas na Web necessárias para a aquisição, tratamento e arquivo dessas fontes e permitir posteriormente a sua exploração com vista a trabalhar o conhecimento que daí se pode extrair. As ferramentas a desenvolver devem ser baseadas na Ontologia do domínio do estudo em causa (a construir), a qual permitirá anotar as fontes documentais, extrair automaticamente os dados e criar o site Web no qual se poderá navegar conceptualmente.

### 3 Online-SoBA: Análise de Textos para estudar comportamentos sociais

*Supervisor: Pedro Rangel Henriques + Cristiana Araújo*  
*Área: Humanidades Digitais; Redes Sociais; Comportamentos*

Resumo:

Neste projeto, Online-SoBA (online Social Behavior Analysis), pretende-se analisar textos curtos que correspondem a comentários escritos em reação a 'posts', colocados em meios de comunicação social (redes sociais ou jornais online), de modo a identificar comportamentos que permitam caracterizar as opiniões sociais de uma dada época.

Para atingir o desafio colocado deverão ser analisados esses textos em Português e extraída informação de acordo com esquemas (a definir) que descrevem os fenómenos a estudar, ou seja estruturas de frases que caracterizam em língua natural os referidos comportamentos.

Como caso de estudo serão fornecidos textos do corpora extraído do Youtube e alguns jornais portugueses no âmbito do Projeto NetLang que estuda o discurso de ódio e perjúrio online.

### 4 NetLangEd, um editor para apoio à Análise de Comentários

*Supervisor: Pedro Rangel Henriques + Cristiana Araújo*  
*Área: Humanidades Digitais; Redes Sociais; Discurso social inadequado*

Resumo:

No âmbito do Projeto NetLang, cujo objetivo é recolher e analisar post/comentários que expressem ódio e usem uma linguagem perjurativa socialmente inaceitável, pretende-se criar um Editor que permita aos investigadores de humanidades e ciências sociais ou psicologia adicionarem as suas reflexões e ideias que vão surgindo enquanto lêem e analisam os documentos que lhes são apresentados online em resposta às pesquisas que fazem ao repositório documental do NetLang.

Esse Editor deve permitir escrever as notas do investigador e associá-las a cada comentário original. Para clarificar o pedido, pense-se nas ferramentas de 'comentários' para documentos PDF do tipo oferecido atualmente pelo Acrobat Reader.

### 5 CompuThinking4SpecialKids: Ensino do Pensamento Computacional para Crianças com Necessidades Especiais

*Supervisor: Pedro Rangel Henriques + Cristiana Araújo*  
*Área: Pensamento Computacional; Ensino da programação; Necessidades educativas especiais; Recursos de aprendizagem*

Resumo:

No âmbito do nosso grupo de Processamento de Linguagens e do Conhecimento temos vindo a trabalhar intensamente num projeto que se chama Computação na Escola (CnE) e cujo grande objetivo é propor como ensinar desde muito cedo crianças e jovens a terem um Pensamento Computacional que os habilite mais tarde a resolver tarefas de resolução de problemas por computador (ou seja de Programação) com maior facilidade e eficácia. É um projeto complexo, mas de absoluta atualidade e aliciante por tudo o que envolve e pelo impacto que irá ter na construção do Cidadão do Sec. XXI.

Nesta proposta de tese pretende-se adaptar essas ideias a uma realidade diferente: o ensino inclusivo da programação para crianças com necessidades educativas especiais. Esta adaptação poderá passar pela definição de uma Linguagem de Programação Física à semelhança da Code Jumper criada pela Microsoft para ensino a cegos no âmbito do projeto Torino.

Assim o candidato a este tema tem que investigar o que é que tem de ser adaptado em termos de características de recursos de ensino/aprendizagem quando se quer ensinar Programação com base no treino do Pensamento Computacional a crianças invisuais, com deficiências auditivas, autistas, dislexia, ou outras com semelhantes necessidades especiais.

Alguns apontadores relevantes para inspirar este projeto:

<https://www.microsoft.com/en-us/research/project/project-torino/>  
<https://codejumper.com/index.html>  
<https://blogs.microsoft.com/ai/project-torino-microsoft-creates-physical-programming-language-inclusive-visually-impaired-children/>  
(alternativa em PT: <https://tecnoblog.net/meiobit/396807/microsoft-project-torino-programacao-deficientes-visuais/>)  
<https://www.perkinselearning.org/technology/posts/project-torino-coding-game-students-visual-impairment>

## 6 Construção de um Jogo com Realidade Aumentada para ensino do CP

*Supervisor: Pedro Rangel Henriques + Lázaro Lima*

*Área: Realidade Aumentada; Ferramentas pedagógicas para apoio ao ensino; recursos de aprendizagem*

### Resumo:

No âmbito do nosso grupo de Processamento de Linguagens e do Conhecimento temos vindo a trabalhar intensamente num projeto que se chama Computação na Escola (CnE) e cujo grande objetivo é propor como ensinar desde muito cedo crianças e jovens a terem um Pensamento Computacional (CP) que os habilite mais tarde a resolver tarefas de resolução de problemas por computador (ou seja de Programação) com maior facilidade e eficácia.

Neste contexto, queremos usar Jogos de Realidade Aumentada (RA) como recurso educacional (RE) no treino do CP, com vista a aumentar a motivação e entusiasmo dos alunos.

Nesta proposta de trabalho de mestrado pretende-se definir com o candidato um pequeno jogo (simples) com RA, o qual deverá ser implementado usando tecnologias atuais com por exemplo WebXR + Three.js + AR.js. Uma vez desenvolvido o artefacto em causa, deve ser concebido e realizado um ensaio que prove a utilidade desse RE.

## 7 Geração de campos customizáveis baseada em Conceitos abstratos

*Supervisor: Pedro Rangel Henriques + Ricardo Martini*

*Área: Ontologias, desenvolvimento de SW*

### Resumo:

Sistemas legados, que já tem uma estrutura consolidada e funcional, dificilmente tem sua estrutura alterada para aceitar novos dados.

Um ponto interessante nesse nicho é o cadastro de novas informações, onde o sistema abre uma série de campos possíveis a serem preenchidos pelo usuário. Porém, os sistemas de hoje em dia, os já consolidados e outros que visam abranger o domínio de forma abstrata, dificilmente dão a liberdade de usuários cadastrarem campos específicos que tratam especificamente de sua área de atuação. Vamos ao problema através de um exemplo: - Imagine um que possui um menu de cadastro de clientes; - Esse sistema serve como gestão documental (abstrato, ou seja, serve para qualquer empresa ou pessoa que necessite gerenciar seus documentos); - O cliente que contratou o sistema de gestão documental trabalha especificamente com o Instituto de Mobilidade dos Transportes (IMT), onde é necessário manter informações de todos os carros que o cliente já teve; - O sistema não permite que o cliente cadastre o campo 'carro', pois não foi modelado para permitir esse cadastro tão específico. Uma das possíveis soluções: Habilitar o cliente a cadastrar o campo desejado (ex: carro - BMW 320i), através de um schema ou mapa conceitual (ontologia, thesaurus, etc.) que trate do domínio de carros, podendo então informar a placa, chassi, último dono, etc; Então, no próximo cadastro de cliente, o sistema deve ser capaz de mostrar que já existe um campo 'carro' e informar os valores possíveis a esse campo. Esses valores podem ser texto livre, baseado nos outros já cadastrados ou baseado em um banco de nomes de carros. A ideia é o usuário saber que pode cadastrar um carro ou abrir novas possibilidades de geração de campos conforme sua necessidade, tudo isso através de um modelo (schema, ontologia, thesaurus, etc.) que mapeie o domínio desejado.

A ideia aqui é identificar uma forma de contornar o problema descrito através do uso de ontologias, dicionários, etc. para descrever o domínio a ser trabalhado e então, dar a possibilidade do usuário final de um sistema já utilizado ou a ser implantado em sua empresa, poder gerar novos campos customizados através do domínio descrito.

Como resultado pode-se ter: - Uma API que habilite o usuário com experiência a tratar do assunto; - Uma interface amigável ao usuário final sem experiência tratar do tema; - Uma linguagem de domínio específico (DSL) amigável para o usuário final sem experiência poder gerar os campos customizáveis.

## 8 VisualPrivas - Anonimização automática de bases de dados

*Supervisor: Pedro Rangel Henriques + Maria João Varanda*

*Área: Privacidade, Segurança de Dados, Bases de Dados, Processamento de Linguagens*

### Resumo:

O sistema Privas, desenvolvido recentemente no contexto de uma tese de mestrado por Joana Miguel ([https://ep1.di.uminho.pt/~gepl/GEPL\\_DS/PRIVAS/](https://ep1.di.uminho.pt/~gepl/GEPL_DS/PRIVAS/)), assegura a privacidade individual na partilha de bases de dados usando uma DSL que permite definir o processo de anonimização usando um conjunto limitado de técnicas.

A ideia subjacente a esta proposta consiste em criar uma versão visual do sistema Privas que permita identificar graficamente os campos da base de dados que se pretende ocultar, generalizar ou anonimizar indicando-se o tipo de ofuscação a aplicar.

Como primeiro objetivo do projeto que aqui se propõe, pretende-se completar o conjunto de técnicas de anonimização disponibilizadas. O segundo objetivo consiste em criar uma interface visual que permita identificar mais facilmente os campos, tabelas e relacionamentos da base de dados que se pretende anonimizar.

## 9 QA Mentoring System - robot anfitrião

*Supervisor: Pedro Rangel Henriques + Maria João Varanda*

*Área: Query and Answering Systems, Robótica, Processamento de Linguagens*

### Resumo:

A ideia será construir um novo QA System usando o sistema ACQA, em desenvolvimento pelo aluno de doutoramento do grupo Renato Azevedo (<https://acqa.di.uminho.pt/>), que possa ser implementado num robot anfitrião (RA) que se passeia nos corredores de uma Instituição com o objetivo de dar apoio informativo a qualquer visitante num ambiente escolar. O sistema QA, a ser incorporado no RA, deve procurar no seu repositório de conhecimento (RC) respostas a pedidos de informação tais como: horários, salas, docentes, prazos, recursos, etc.

O sistema ACQA permita a criação automática de um QA system tendo por base um conjunto de possíveis questões e respetivas respostas. Neste caso, a complexidade deste trabalho não se baseia no processamento das questões nem na obtenção de ranking de respostas (porque isso já está implementado pelo sistema ACQA), mas sim na construção de uma base de conhecimento (RC) dinâmica que alimente o QA System gerado e que a cada tipo de questão associe as correspondentes respostas criadas no momento. Isso iria permitir a consulta de outras bases de dados como, por exemplo, a base de dados dos horários dos cursos ou dos serviços administrativos que fazem a gestão dos alunos ou das instalações e demais recurso.

## 10 Análise biométrica de comportamentos em jogos terapeuticos

*Supervisor: Pedro Rangel Henriques + Maria João Varanda*

*Área: Ciência dos dados*

### Resumo:

No âmbito dos projetos de 4º ano (LEI), três grupos do MiEI/MEI desenvolverem 3 jogos sérios, a pedido do Centro de Neurociência de Braga da Dr<sup>a</sup> Ana Paula Azevedo, para reconhecimento de formas e de emoções e para treino da visão central e periférica. Esses instrumentos servem para terapia da memória, desconcentração, dislexia e outros problemas que afetam a aquisição de conhecimentos em processos de aprendizagem. Se se medirem os dados do paciente durante a realização do jogo (traçado do percurso ocular, equilíbrio e outros marcadores biométricos) é possível também usar os mesmos recursos para avaliar o nível da doença e dosear a terapia.

Nesta proposta pretende-se combinar os dados recolhidos durante a realização do jogo (tempo, tentativas, pontuação obtida) com os dados dos sensores biométricos disponíveis de modo a poder aplicar algoritmos de análise de dados para

se descobrirem novas relações entre os dados e através da adequada visualização se possa criar novo conhecimento.