

# Perfil de PLC - Processamento de Linguagens e de Conhecimento (MiEI + MEI2018/19)

## Propostas de temas para o Projeto Integrado (PI)

### 1 Um Sistema de Query & Answering (Q&A) melhorado

*Supervisor: José João Almeida + Pedro Rangel Henriques*

*Área: PLN; Bases de Conhecimento; Cálculo de Similaridades*

#### Resumo:

Neste projeto pretende-se retomar o Q&A System desenvolvido no 1º semestre no âmbito da disciplina de GCS e torná-lo de real interesse dotando-o com uma interface para colocar as perguntas em Língua Natural (PT) e fazendo com que o sistema retorne as respostas mais interessantes.

### 2 DSL para geração de Chatbots

*Supervisor: José João Almeida + Pedro Rangel Henriques*

*Área: PLN; Bases de Conhecimento; Rule-based; DSL*

#### Resumo:

Criar um sistema que baseado, numa especificação seguindo uma linguagem específica a definir, gere um chatbot capaz de manter um diálogo, prestar serviços informativos, etc.

Essa ferramenta poderá conter:

- um motor geral de conversação baseado em regras (padrão  $\rightarrow$  função respondedora)
- uma calculadora de regras Chatbots (*álgebra de Chatbots*) que:
  - a partir de uma variedade de recursos (exemplo listas de provérbios, bases de dados, ontologias) construa chatbots;
  - a some, subtraia, etc Chatbots.

### 3 Um Extensor e Aferidor de léxicos afetivos

*Supervisor: Pedro Rangel Henriques + José João Almeida + Ricardo Martins*

*Área: PLN; Emoções*

#### Resumo:

O problema na utilização de léxicos é que eles nunca contém 100% das palavras existentes, e no caso dos afetivos, a situação é um pouco pior, pois sempre são um subconjunto de um léxico principal. Além disso, como a língua é dinâmica, novas expressões surgem muito rapidamente, e com isso, o tempo necessário até *atualizar* o léxico é enorme.

Outro problema ao se utilizar léxicos afetivos é que eles refletem a forma de *alguém* se expressar, entretanto, cada pessoa se expressa de uma forma, e com isso, essa generalização compromete a análise de sentimentos, pois ela se baseará no padrão emocional do léxico, e não no padrão emocional do autor do texto.

Por exemplo, ao utilizar um léxico português, ao analisar a palavra *fino*, no Brasil terá um sentimento diferente de Portugal. É possível até arriscar que na análise entre lisboetas e portugueses já teríamos diferenças emocionais.

O mesmo problema refere-se à intensidade da emoção. Enquanto para portugueses o *fino* desperta uma intensidade alta, para os lisboetas e brasileiros, certamente não despertam emoções muito altas.

O objetivo do trabalho (que faz parte de um projeto de doutoramento na área de *sentiment analysis*) é o desenvolvimento de uma ferramenta que, através de técnicas de PLN, permita estender e personalizar um léxico já existente, a partir de um conjunto de frases e textos de um único autor.

Essa ferramenta deve permitir a criação do novo léxico emocional pessoal em formato XML, além de disponibilizar um Web Service onde possam ser consumidos serviços como:

- entrada de textos
- solicitação de geração de novo léxico
- análise de sentimentos
- recálculo de sentimentos e intensidades

## 4 DSL para Árvores de reação (behaviour trees)

*Área: Jogos, inteligência artificial para jogos, DSL*

*Supervisor: José João Almeida*

Resumo:

As árvores de reação ou de comportamento são um formalismo para descrever o comportamento reactivo de personagens (ou outros elementos) de jogos.

Pretende-se criar uma linguagem textual para Árvores de reação e um compilador (para python, perl, node?) e um caso de estudo.

Ver também wikipedia:

- Behavior tree (artificial intelligence, robotics and control)
- Behavior tree

## 5 Site de avaliação automática de exercícios de programação

*Supervisor: José João Almeida + Rui Mendes*

Resumo:

O objetivo desta proposta de projeto de grande relevo para apoio ao ensino da programação é:

- Proceder à avaliação automática de código:
  - Comparadores mais inteligentes do resultado obtido pelo programa e o esperado;
  - Avaliação parcial;
  - Vários tipos de feedback opcional ao aluno mostrando-lhe os casos em que o programa falhou;
- Coleta de informação:
  - Guardar as submissões juntamente com a data da submissão e outra informação tal como a data da consulta do enunciado;
  - Colecionar estatísticas tais como métricas sobre o código, número de testes falhados, diagnósticos sobre a execução, etc
- Fazer uma ferramenta que permita:
  - A autenticação de utilizadores;
  - A criação de cursos em "back office";
  - A disponibilização de exercícios junto com uma etiquetagem dos mesmos (e.g., seguindo uma ontologia de temas);
  - A resolução dos exercícios;
  - O feedback do resultado da submissão.

## 6 Animador de espaços de aprendizagem

*Supervisor: José João Almeida*

Resumo:

Criar uma ferramenta baseada em DSL para definir e animar sequências de atividades ligados à aprendizagem "autómato de aprendizagem (AA)" (exercícios, slides, PDFs, etc), mantendo um estado tão rico quanto possível das actividades realizadas e do grau de assimilação dos conceitos visitados.

De um modo simplista:

1. Os nós, correspondem ao estado atual de aprendizagem caracterizado pelo que já fez e já aprendeu
2. Os ramos correspondem às atividades (testes bem / menos bem sucedido; slides, pdf, youtubes, etc)
3. a realização de cada atividade por parte de cada estudante altera o seu "estado de conhecimento/reputação", perfil.

## 7 MkbBuilder – Linguagem para criação de teclados de computador musicais

*Supervisor: J. João Almeida*

*Área: Música, audio, DSL, MIDI, SF2*

Resumo:

Definir e implementar um linguagem de domínio específico capaz de:

- definir SF (sound fonts) – conjunto de sons associadas a um id-som
- associar (estática ou dinamicamente) teclas a id-som
- animar musicalmente!

## 8 Ferramenta para especificar música parcial (MPL)

*Área: Música, DSL, abc music notation*

*Supervisor: José João Almeida*

Resumo:

Objectivo: geração de música para acompanhamento de improvisação. ensaio de partes, autoaprendizagem, etc.

Notação musical usada: ABC (e ferramentas abcm2ps, abc2midi, etc. Ver também documento <http://abcplus.sourceforge.net/>, [http://abcplus.sourceforge.net/abcplus\\_en-2015-09-03.zip](http://abcplus.sourceforge.net/abcplus_en-2015-09-03.zip)).

Definir uma domain specific language – MPL – que permita geração fácil de ABC (e consequentemente geração de partituras, midi e mp3) a partir de:

1. sequências de acordes (com atributos opcionais) (mesmo sem melodia)
2. sequências simples de percussão
3. música a várias partes(=vozes) (com atributos de realce, atenuação de partes, *all-but-one*, *just-on*)

A ferramenta a construir pode ter várias abordagens (a decidir) por exemplo:

- MPL (sem mais nada)
- MPL com ABC embutido
- ABC estendido com MPL

```
chords[accordion] [fzcc] {Em|D|Em|Em . D Em|}
```

## 9 Museu da Pessoa – suporte às entrevistas

*Supervisor: José João Almeida*

Resumo:

Pretende-se criar um sistema para suportar/ajudar a entrevistas para o museu da pessoa: (para correr em tablets, androids ou portáteis):

- capaz de fotografar (a pessoa + reproduzir fotos)
- capaz de gravar audio
- capaz de arrumar genealogia
- capaz de relembrar algumas perguntas a fazer (lista de perguntas):

- O que recorda dos seus tempos de escola?
  - aprendizagem?
  - namoro?
  - ofício?
  - ...
- capaz de arrumar os vários ficheiros de modo sistemático

```
-projecto co contrabando Montalegre
  - José-Moura
    - entrevista.mp3
    - entrevista.xml
    - fotos
      - f1.jpg
      - anexo1.jpg
  ...
```

## 10 Um Sistema de apoio à Ingestão e Catalogação de Posts e Comentários

*Supervisor: Pedro Rangel Henriques + Cristiana Araújo*  
*Área: Criação e Catalogação de Repositórios de Conhecimento*

### Resumo:

âmbito do projeto internacional de investigação suportado pela FCT, Netlang (<https://sites.google.com/site/projectnetlang/home>) para criação de corpora paralelo anotado PT/EN que permita estudar formas de insulto e discriminação em redes sociais, pretende-se criar uma ferramenta de uso simples que permita analisar uma coleção de posts e respetivos comentários, recolhidos de online e representados em JSON, e fornecer uma forma semi-automática de criar a metainformação (ou manifesto) de cada grupo e proceder à anotação dos textos (notícia e comentários) em formato TEI para posterior armazenamentonum repositório não-SQL onde posteriormente possa ser consultado.

## 11 Digital humanities Notebook

*Supervisor: J.João Almeida*  
*Área: Ontologias, Literate programming, DSL, Humanidades digitais, bibliotecas digitais*

### Resumo:

Ao escrever no seu caderno as histórias de família, estudos genealógicos, descrever fotografias, recortes de imprensa, resultados de pesquisa, histórias de locais, casas, o investigador amador precisa de ajuda!!!

Faça um DHN:

- LateX based, Ontology based;
- com anotações para gerar automaticamente triplos de ontologias;
- com anotações genealógicas;
- com capacidade de captura de partes de ecrã eventualmente com OCR de capturas (DSL de configuração).

## 12 UTK2-Tatoeba para línguas minoritárias angolanas

*Supervisor: J.João Almeida*  
*Área: Corpora paralelos, crowdsourcing, NLP*

### Resumo:

Pretende-se :

- Criar uma plataforma para (1) aceitar pequenas frases a traduzir (2) aceitar traduções (usando técnicas de gamificação).
- capaz de criar pequenos quizzes de aprendizagem de língua.
- capaz de criar recursos e ferramentas de apoio ao utk2 com base nas bi-frases colecionadas.

## 13 CLAV - Public Information Classification and Evaluation: formal model specification

*Supervisor: José Carlos Ramalho, José Nuno Oliveira*

### Resumo:

CLAV is a platform being developed by DI/UM with the partnership and ordered by the Direção Geral do Livro, Arquivos e Bibliotecas (DGLAB) which aims the classification and evaluation of all documents that move around and across portuguese public institutions.

One view, already visible, of the project is the so called Lista Consolidada, which is a comprehensive catalog of all administrative processes occurring in portuguese public institutions.

The first version is available online: <http://clav-test.di.uminho.pt>. For further information contact the author. There are several articles and presentations about this subject.

Currently, the model is specified in OWL ("Ontology Web Language"), but suffered several changes during the last year and there was no time to study the impact of those changes in the model pre-conditions and invariants.

In this project we want:

1. to formal specify the model from scratch;
2. to load the available information onto the new model;
3. to specify pre-conditions and invariants;
4. to report errors.

## 14 CLAV - Classificação e Avaliação da Informação Pública: desenvolvimento de uma interface reativa

*Supervisor: José Carlos Ramalho, José Francisco Campos*

### Resumo:

A CLAV é uma plataforma em desenvolvimento no DI/UM em parceria e sob encomenda da Direção Geral do Livro, Arquivos e Bibliotecas (DGLAB) que visa a classificação e a avaliação de toda a documentação circulante na administração pública portuguesa. Uma das faces visíveis do projeto é o catálogo dos processos da administração pública portuguesa, designado por Lista Consolidada. A plataforma já se encontra disponível para consulta online: <http://clav-test.di.uminho.pt>. Para mais informação contacte os supervisores. Há muito material escrito sobre o projeto.

Neste projeto, pretende-se desenvolver uma nova interface Web, para as operações de CRUD sobre a lista consolidada. O modelo de dados é complexo, é um grafo, e tem muitos invariantes, o que faz com que a interface tenha de ser complexa e reativa a muitas situações.

A nova interface deverá obedecer aos seguintes requisitos:

1. Ser capaz de garantir os invariantes durante a fase de criação, actualização e remoção dos dados;
2. Ser desenvolvida segundo princípios de Responsive Web Design (capacidade de se ajustar a diferentes tamanhos de écran);
3. Ser desenvolvida em BootStrap 4 e Vue ou, BootStrap 4 e React;
4. Ser modularizada em componentes que possam ser reutilizados;
5. Usar a API de dados disponível no projeto.

## 15 CLAV - Classificação e Avaliação da Informação Pública: desenvolvimento de uma API de dados

*Supervisor: José Carlos Ramalho*

### Resumo:

A CLAV é uma plataforma em desenvolvimento no DI/UM em parceria e sob encomenda da Direção Geral do Livro, Arquivos e Bibliotecas (DGLAB) que visa a classificação e a avaliação de toda a documentação circulante na administração pública portuguesa.

Uma das faces visíveis do projeto é o catálogo dos processos da administração pública portuguesa, designado por Lista Consolidada.

Já se encontra disponível para consulta online: <http://clav-test.di.uminho.pt>. Para mais informação contacte o autor. Há muito material escrito sobre o projeto.

Neste projeto, pretende-se desenvolver uma nova API de dados para consulta de toda a informação com os seguintes requisitos:

1. O modelo de dados encontra-se especificado em OWL e a informação armazenada no GraphDB, por isso a API deverá ligar-se a este sistema;
2. A API deverá ser especificada em Swagger;
3. A especificação em Swagger deverá ser modular e conter exemplos e a especificação dos tipos de entrada e saída;
4. A API deverá fornecer os dados num dos seguintes formatos (indicado em cada pedido): CSV, RDF, JSON e XML;
5. Deverá ser especificada uma bateria de testes à API.

## 16 Ulisses XXI

*Supervisor: José Carlos Ramalho*

### Resumo:

Neste projeto, pretende-se desenvolver um navegador Web para ontologias OWL baseado em SVG (os browsers possuem bons motores de "rendering" para SVG) e JavaScript. Uma Ontologia OWL estruturalmente assemelha-se a um grafo, por isso, o que aqui se pretende desenvolver é um visualizador de grafos em SVG interativo. Mediante os cliques que o utilizador for fazendo sobre os elementos do grafo, este irá mudar apresentando outra informação ou outro ponto de vista sobre a informação. No passado, já foram desenvolvidos navegadores baseados em páginas Web que poderão servir de ponto de partida para o trabalho. As ontologias que servirão de caso de estudo estarão armazenadas em GraphDB e pretende-se usar a API deste sistema para aceder às ontologias.

## 17 Análise Química na Web

*Supervisor: José Carlos Ramalho, Pier Parpot*

### Resumo:

Um problema recorrente nas análises laboratoriais é a comunicação dos dados entre os aparelhos que fazem as análises e outras máquinas que pretendem depois fazer algo com esses dados.

Neste contexto, têm surgido vários formatos ao longo do tempo, uns com mais suporte do que outros.

Um destes formatos é o mzML. Neste projeto, pretende-se que a equipa estude o formato e desenvolva uma aplicação Web capaz de explorar datasets neste formato.

Alguns dos requisitos serão:

1. Melhorar o parser existente em JavaScript para processar o mzML;
2. Criar uma aplicação Web capaz de:
  - permitir o upload de ficheiros mzML;
  - armazenamento desses ficheiros numa BD a estudar;
  - permitir fazer o download de datasets armazenados;
  - permitir a configuração de views sobre os datasets armazenados;

- permitir restringir um dataset para análise e visualização;
- desenvolver vários tipos de views: tabulares e gráficas;
- autenticar os utilizadores.