

Perfil de PLC - Processamento de Linguagens e de Conhecimento (MEI2015/16)

Propostas de temas para o Projeto Integrado (PI)

1 Ferramenta para especificar música parcial (MPL)

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).

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
- MPL com ABC embutido
- ABC estendido com MPL

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

2 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)"(execícios, slides, PDFs, etc), mantendo um estado tão rico quanto possível das actividades realizadas e do grau de assimilação dos conceitos visitados.

De um modo simplicista:

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.

3 Parser de linguagem natural baseado em VISL-CG3

Supervisor: José João Almeida

Resumo:

- partindo de árvores sintáticas (Treebanks, corpora anotado: cetem-publico anotado, floresta sintatica) inferir informação sintática diversa (gramáticas e dicionários)
- estudar o VISL-CG3 (ferramenta para geração de parsers para LN, baseada em regras);
- inferir e testar regras VISL-CG3;
- (...juntar regras manuais).

4 Museu da Pessoa – suporte às entrevista

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 lembrar 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
      - anex01.jpg
    ...
```

5 DSL para Árvores de reação (na continuação do proj. do ano anterior)

Supervisor: José João Almeida

Resumo:

TBA

6 Desenvolvimento dum Site para suporte à Associação Território com Vida (TcV)

Supervisor: Pedro Rangel Henriques

Resumo:

A Associação Território com Vida (TcV), com sede em Arcos de Valdevez, tem um projeto financiado de combate aos incêndios por reabilitação de técnicas tradicionais de pastoreio. Para viabilizar o projeto a TcV vai lançar um sistema de apadrinhado de animais dos rebanhos, pelo que precisa de um sistema de informação baseado na Web para gerir as explorações, o pastoreio e o processo de apadrinhamento.

Neste projeto pretende-se desenvolver o Back-office (BO) e o Front-office (FO) do SIsATCV (sistema de informação de suporte à associação TcV), usando bases de dados, ontologias DSL e tecnologia Web 2.0.

7 Desenvolvimento da Plataforma para suporte ao Ensino da Programação (PEP)

Supervisor: Pedro Rangel Henriques e Paula Tavares

Resumo:

Para minorar as dificuldades que os alunos de cursos de informática sentem nos cursos de introdução à programação, planeou-se uma abordagem que faz uso de técnicas de animação de programas e de sistemas de avaliação automática de programas com vista a aumentar o envolvimento dos alunos a sua auto-confiança e por conseguinte aumentar a motivação essencial à aprendizagem eficaz destas matérias. Nesse contexto será desejável ter um sistema de informação baseado na Web que permita aos alunos fazerem sessões de auto-estudo em casa. Tal sistema deve permitir ao docente carregar os exercícios a usar em cada sessão, fazer o plano da sessão e deve permitir ao aluno praticar um determinado tema do curso animando os exercícios ou resolvendo-os e testando-os com recolha imediata de feedback. Será ainda desejável que o professor possa receber informação sobre a forma como decorreu a sessão de trabalho de cada aluno. Neste projeto pretende-se desenvolver o Back-office (BO) e o Front-office (FO) da PEP (Plataforma para suporte ao Ensino da Programação), usando bases de dados, ontologias DSL e tecnologia Web 2.0.

8 Criação automática de um Museu virtual sobre o Clero Catedralício português Medieval (MvCC)

Supervisor: Pedro Rangel Henriques e Ricardo Martini

Resumo:

Neste projeto pretende-se desenvolver um museu virtual para consultar informação sobre catedrais portuguesas da Idade Média e os percursos de vida dos seus Clérigos, usando uma base de dados construída a partir dos fasti portugueses dessa época e criando uma ontologia para conceptualizar a dita BD. A partir dessa ontologia pretende-se usar a plataforma CaVa para gerar automaticamente as salas de exposição do dito museu virtual.

Uma das tarefas do projeto será estudar a BD e construir a Ontologia em RDF, mapeando os conceitos da ontologia para as respetivas instâncias na base de dados. Por fim será necessário estudar o sistema de geração de espaços de aprendizagem CaVa para descrever as salas do museu desejadas e proceder á sua criação automática.

9 Criação automática de salas de exibição para o Museu da Pessoa

Supervisor: Pedro Rangel Henriques e Ricardo Martini

Resumo:

“O Museu da Pessoa é um museu virtual e colaborativo fundado em São Paulo no ano de 1991. Desde a sua origem, tem como objetivo registar, preservar e transformar em informação, histórias de vida de toda e qualquer pessoa da sociedade”.

Baseado nessa definição, este projeto tem a intenção de criar um museu virtual para aceder às histórias de vida das pessoas comuns entrevistadas de uma forma mais simples (natural e eficaz) do que em documentos anotados em XML

(que é o seu estado atual). Para tal é preciso criar uma base de dados relacional para o Museu da Pessoa a partir dos ditos documentos anotados. Além disso, para que essas histórias de vida possam ser acedidas por um navegador conceptual (que está por trás das salas de exibição virtuais), terá de ser criado um mapeamento entre essa base de dados e uma ontologia (já criada para o propósito).

10 Desenvolvimento dum Sistema para agregação de Informação

Supervisor: Pedro Rangel Henriques

Resumo:

Neste projeto pretende-se criar um sistema que permita agregar fontes de informação a partir de documentos diferentes sobre o mesmo assunto de modo a obter resposta o mais ricas possível.

Para concretizar imagine-se que se pretende obter informação sobre cinema (filmes, atores, realizadores, etc.) e suponha-se que existe uma enciclopédia anotada em formato digital, um compêndio também anotado e um site sobre cinema. O objetivo é receber uma pergunta e procurar a resposta nas várias fontes disponíveis devendo depois ser possível agregar a informação de modo a gerar uma resposta completa.

11 Desenvolvimento dum Sistema de Suporte à Macro Estrutura Funcional (MEF)

Supervisor: José Carlos Ramalho

Resumo:

Com o objetivo de contribuir para a maior eficácia na gestão documental e para o incremento da interoperabilidade semântica nas trocas documentais entre os serviços da Administração Pública, a DGLAB tem vindo a promover a utilização de uma estrutura para classificar e organizar a documentação do setor público.

Para a concretização deste objetivo foram empreendidas duas linhas de atuação:

1. Desenvolvimento de um projeto cooperativo para a criação de uma linguagem comum, estabelecido em duas fases: (a) Criação da Macroestrutura funcional (MEF); (b) Harmonização de classes de 3º nível em planos de classificação conformes à MEF.
2. Apoio a projetos supraorganizacionais ou organizacionais para a criação de planos de classificação que partilhem uma linguagem comum.

De modo a existir um código de classificação unívoco, passível de ser utilizado em diferentes planos de classificação, garantindo a interoperabilidade semântica até ao 3º nível, a DGLAB coordena a atribuição dos códigos de 3º nível para planos de classificação harmonizados e conformes à MEF.

As entidades que se encontram em fase de conclusão dos seus planos de classificação devem solicitar à DGLAB a codificação dos novos processos. As regras para a atribuição dos códigos encontram-se definidas no documento "Orientações básicas para o desenvolvimento dos terceiros níveis em planos de classificação conformes à Macroestrutura Funcional", disponível em http://arquivos.dglab.gov.pt/wp-content/uploads/sites/16/2014/02/2013_Orient-3-niveis_PC-MF.pdf

A DGLAB, no esforço de normalização prosseguido enquanto órgão de coordenação da política arquivística nacional, disponibiliza a lista consolidada dos 3ºs níveis representativos dos processos de negócio estabilizados.

Esta lista agrega os 3ºs níveis aferidos pela DGLAB, propostos por distintas entidades, definidos de acordo com as regras estabelecidas nas "Orientações básicas para o desenvolvimento dos terceiros níveis em planos de classificação conformes à Macroestrutura Funcional".

Até ao momento, toda esta iniciativa tem sido coordenada sobre um documento MSWord/PDF. Neste momento, com a dimensão e complexidade atingidas a manutenção deste sistema de informação num documento de texto deixou de ser viável. Nesse sentido, já se desenvolveram esforços para migrar grande parte da informação para um formato XML adequado. Neste projeto, pretendem-se estudar alternativas ao armazenamento desta informação que permitam uma edição colaborativa mas controlada da norma. No fim e a partir do formato/sistema escolhido para armazenamento devem ser especificadas scripts de transformação que permitam gerar automaticamente:

- um website que permita navegar sobre o sistema de informação;
- um documento PDF que constitua a documentação completa sobre a norma;
- três ontologias nos formatos RDF, SKOS e OWL com toda a informação da norma.

Esta parte de disseminação da informação deverá estar acoplada a um sistema de informação baseado na Web que permita a edição colaborativa da informação.

12 Enciclopédia Musical à Medida (“MusicPedia”)

Supervisor: José Carlos Ramalho

Resumo:

Neste projeto, pretende-se desenvolver um gerador de enciclopédias musicais. O ponto de partida será constituído por várias listas de itens: instrumentos, obras, compositores, etc. A partir destas listas o gerador deverá usar vários recursos da “Linked Open Data” como o DBPedia e o MusicBrainz para tentar agregar o máximo de informação possível sobre cada item da lista. A informação colhida deverá ser armazenada e um sistema genérico de navegação (criado para o efeito) deverá permitir o acesso a esta informação via Web.

13 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.

14 Gestor e arquivo digital de conteúdos (“MIEI Box - MBox”)

Supervisor: José Carlos Ramalho

Resumo:

Neste projeto, pretende-se desenvolver um sistema Web para armazenamento e catalogação de recursos digitais: documentos PDF ou Word, Slides, Filmes, Fotografias, etc. O sistema deverá permitir a utilização de múltiplas taxonomias classificativas (SKOS ou OWL) na catalogação dos recursos e usar essas taxonomias para navegar nos recursos armazenados. O sistema a desenvolver deverá ter as seguintes características:

- funcionar sobre a Web com uma interface para humanos (via formulários) e outra para máquinas (Web Services REST ou SOAP);
- deverá suportar o armazenamento de lotes de recursos via Web Services;
- autenticação;
- gestão de níveis de acesso: os recursos digitais armazenados poderão ser públicos ou privados;
- no portal de consulta só serão visíveis os recursos públicos ou pertença do utilizador que se tiver autenticado.

15 XForms Processor

Supervisor: José Carlos Ramalho

Resumo:

Neste projeto, pretende-se desenvolver um processador de XForms de modo a criar aplicações Web partir de uma especificação em XForms. O resultado final poderá ser XHTML e JavaScript ou outra combinação de linguagens de Scripting. Numa primeira fase, deverá ser desenvolvido um protótipo que suporte as operações CRUD.

16 Cálculo de Pseudo-Palavras

Supervisor: Alberto Simões e Ana Paula Soares (Psi)

Resumo:

O grupo de Processamento de Linguagens tem vindo a colaborar com a escola de Psicologia nomeadamente no que concerne os estudos cognitivos relacionados com a língua e a sua aquisição.

Para algumas das experiências realizadas pelo Psico-Linguística, são necessárias pseudo-palavras: palavras que podiam ser reais, mas não são, e que são parecidas com palavras reais.

Embora existam alguns algoritmos já implementados por outros investigadores, a escola de psicologia pretende a reimplementação do algoritmo no sentido de tirar partido de uma base de dados lexical que foi construída num projeto recente.

Com este trabalho pretende-se:

1. estudar os algoritmos já existentes;
2. re-implementar um dos algoritmos existentes, tendo em conta os detalhes da língua portuguesa;
3. implementação de um serviço web para a geração de pseudo-palavras e a sua avaliação de acordo com diferentes parâmetros da base de dados lexical existente.

17 Editor XForms para Dicionário da Academia das Ciências de Lisboa

Supervisor: Alberto Simões

Resumo:

O Departamento de Informática tem um protocolo com a Academia das Ciências de Lisboa no sentido de prestar auxílio no desenvolvimento de uma base de dados que suporte todo o processo de criação e manutenção da nova versão do Dicionário Contemporâneo da Língua Portuguesa.

Para agilizar o processo de criação do dicionário, a versão impressa em 2005 foi processada (a partir do PDF respetivo) e as entradas do dicionário importadas em documentos XML TEI, hospedadas num serviço noSQL eXistDB.

Com este trabalho pretende-se:

1. implementar uma interface XForms, sobre eXistDB, para a edição das entradas do dicionário;
2. implementar uma interface XForms e JScript/JSON para a consulta do dicionário;
3. implementar ferramentas de manutenção do dicionário, como a deteção de inconsistências, extração de informação, relatórios de estatísticas, etc., usando XSLT e XQuery.
4. desenvolvimento de algumas folhas de estilo CSS associadas à visualização do dicionário, e de XSLT para a exportação em LaTeX.