

Mestrado em Informática

UCE30-EL, Engenharia de Linguagens

Projecto Integrado (PI)

Ano Lectivo de 2011/12

1 Objectivos

Com este projecto integrado pretende-se sedimentar os conhecimentos introduzidos nas aulas teóricas dos 4 módulos desta UCE: Engenharia Gramatical, Processamento Estruturado de Documentos, Análise e Transformação de Software e Scripting no Processamento de Linguagem Natural.

2 Resultados da aprendizagem

No final do projecto, os alunos devem ter adquirido um conjunto de competências tecnológicas, específicas e genéricas detalhadas na rubrica "Resultados de Aprendizagem" da página oficial da disciplina, em

<http://wiki.di.uminho.pt/twiki/bin/view/Mestrado/EL>.

Nomeadamente,

- Capacidade para desenvolver especificações da sintaxe/semântica de linguagens com gramáticas de atributos.
- Capacidade para gerar programas usando ferramentas automáticas baseadas em gramáticas de atributos.
- Capacidade para gerar ou utilizar ambientes de desenvolvimento estruturais e orientados à semântica.
- Capacidade para representar, armazenar e manipular Conhecimento com eficiência.
- Capacidade de escrever scripts para automatização de uma variedade de tarefas e transformações.
- Capacidade de resolver problemas usando transformações via Expressões Regulares.
- Capacidade de compreender as vantagens e o funcionamento de sistemas guiados por regras de produção (condição-reacção).
- Capacidade de construir Linguagens de Domínio Específico (DSLs) concretas, bem como produzir com eficiência eficientes processadores.
- Capacidade de construir e usar corpora.
- Capacidade de extrair informação diversa a partir de corpora.
- Capacidade de construir dicionários electrónicos.
- Capacidade de construir pequenos protótipos para modelar linguagem natural.
- Capacidade de construir front-ends poderosos para a análise de linguagens de programação.
- Capacidade de desenvolver software como uma tarefa de transformar, de forma sistemática eficiente, programas.

- Capacidade de utilizar métricas e técnicas de transformação de programas para otimizar programas (e.g. cálculo parcial, detecção de código morto), efectuar debugging de programas (e.g. slicing), melhorar a estrutura dos programas (e.g. refactoring).
- Capacidade de definir testes para software e testar automaticamente programas em diferentes linguagens de programação.
- Capacidade para criar representações visuais adequadas à compreensão clara do conhecimento complexo detido.
- Capacidade de compreender o ciclo de vida dos documentos estruturados, sabendo identificar as várias fases e as tecnologias a utilizar em cada uma.
- Capacidade de especificar uma linguagem de anotação para um conjunto de requisitos.
- Capacidade de implementar transformações de documentos para diversos fins: extracção de conhecimento, publicação na Web, intercâmbio de informação, ...
- Capacidade de utilizar soluções de armazenamento para documentos anotados.
- Capacidade de definir as camadas necessárias para integrar e realizar o intercâmbio de informação entre sistemas de informação distintos.
- Capacidade de implementar um projecto de publicação electrónica recorrendo a normas internacionais abertas: XML, XSLFO e XSL.
- Capacidade de programar a geração automática de sítios Web a partir de um repositório de documentos XML.
- Capacidade de utilizar linguagens de anotação e respectivas ferramentas desenvolvidas por outrem.

3 Organização e Funcionamento

O projecto será desenvolvido em grupos de 2, fora das aulas, em ambiente Linux (este ano será o único ambiente de trabalho oficial).

A única aula presencial de 1h/semana será usada para:

- Apresentação do projecto (no todo e em cada uma das suas fases), discussão do enunciado e identificação dos requisitos, alinhamento de estratégias e planeamento da resolução.
- Formação teórica básica em temas tecnológicos transversais aos 4 módulos e essenciais ao desenvolvimento do projecto, com proposta de temas a investigar.
- Gestão dos projectos; Apresentações Intercalares (avaliação); e Seminários.

No fim de cada semestre e no fim do ano lectivo, cada grupo apresentará à equipe docente o trabalho realizado e os resultados obtidos, devendo entregar um relatório técnico de desenvolvimento devidamente estruturado e fundamentado, escrito em \LaTeX .

3.1 Calendarização

O projecto deve ser executado ao longo de todo o ano estando a **entrega final** agendada para a 25/06/2012.

Para controlo da situação e avaliação intermédia, haverá 3 apresentações intercalares do projecto, que coincidem com o meio do 1º semestre, o início do 2º semestre e o meio do 2º semestre. Mais concretamente nos dias: 12/12/2011, 05/03/2012 e 30/04/2012.

3.2 Avaliação

Quanto à avaliação do PI, ela será feita por fases: 3 intermédias e 1 final. Em cada fase, o trabalho será classificado de 0 a 20 valores. As classificações das etapas intermédias são informativas, para auto-controlo do grupo e para que os docentes possam também ir afinando todo o processo—requisitos, escalonamentos das várias componentes ou tarefas, abordagens, tecnologias—quer globalmente, ao nível da turma, quer especificamente, ao nível de um grupo.

Embora a classificação da PI seja determinada pela nota atribuída na etapa final, tomar-se-á em conta as classificações intermédias e sobretudo o sinal da sua derivada. Para aferição de cada fase, serão considerados os seguintes factores:

- o Resultado absoluto alcançado;
- o Processo de Desenvolvimento (a análise, as soluções arquitectónicas, as propostas tecnológicas, a solução de implementação, a concepção da interface e da interacção, os testes e demos);
- a Investigação realizada para suportar esse desenvolvimento;
- a Documentação produzida, a nível de relatório e das apresentações;
- a Exposição e Defesa do projecto;
- o Funcionamento em grupo e a gestão do projecto.

4 Enunciado

O que se pretende neste projecto é desenvolver um Sistema para Geração Automática de Salas de Exposição em Museus Virtuais com base em Ontologias, que doravante designaremos por GeSEVibO.

O objectivo principal é criar um ambiente que ofereça funcionalidades para cumprir 3 grandes objetivos:

- arquivar o espólio do museu;
- gerar automaticamente os espaços de aprendizagem na forma de salas de exposição do museu;
- guiar as visitas recebendo delas informação.

Arquivar o espólio significa receber meta-informação sobre cada objeto do espólio (eventualmente o próprio objeto se ele for digital), classificá-lo à luz de uma ontologia definida à partida (mas que poderá ir evoluindo) e armazenar a meta-informação após classificação.

Cada sala de exposição de um Museu Virtual (MV) é um sítio Web através do qual se podem ver os objetos em exposição, ou a sua descrição, devidamente relacionados entre si de forma a transmitir o conhecimento correspondente ao tema dessa exposição. Neste caso, esse tema, ou conhecimento a transmitir, será descrito também por uma ontologia (que será uma vista da ontologia global que descreve o arquivo), ontologia essa que define os objetos a expor e as conexões lógicas entre eles. A visita corresponderá a navegar nesse sítio entre os objetos, seguindo as relações determinadas pela ontologia.

Assim, para criar o sitio web que virtualiza a sala de exposição será necessário, por um lado conhecer a ontologia que define o tema da sala de modo a poder-se extrair do arquivo cada objeto mencionado e a criar os links apropriados entre eles, e por outro lado, conhecer detalhes sobre o layout das páginas que formam o dito sítio web.

Por fim, guiar a visita significa apresentar o sítio criado com base na ontologia da exposição e as indicações relativas à disposição dos objetos na sala e permitir que o visitante da página possa deixar comentários ou mesmo fornecer conhecimento que deve ser arquivado no Museu, quer acrescentando novos objetos ou ligações, quer corrigindo informações apresentadas.

O repositório final deverá seguir o modelo de referência internacional OAIIS ("Open Archival Information System") que mostra como deve ser estruturado um repositório deste tipo (fig.1).

O Modelo OAIIS é composto por 3 megaprocessos:

Ingestão - processo que trata da incorporação de nova informação para dentro do sistema;

Administração - processo que trata da gestão da informação dentro do sistema;

Disseminação - processo que trata da exploração e distribuição da informação armazenada.

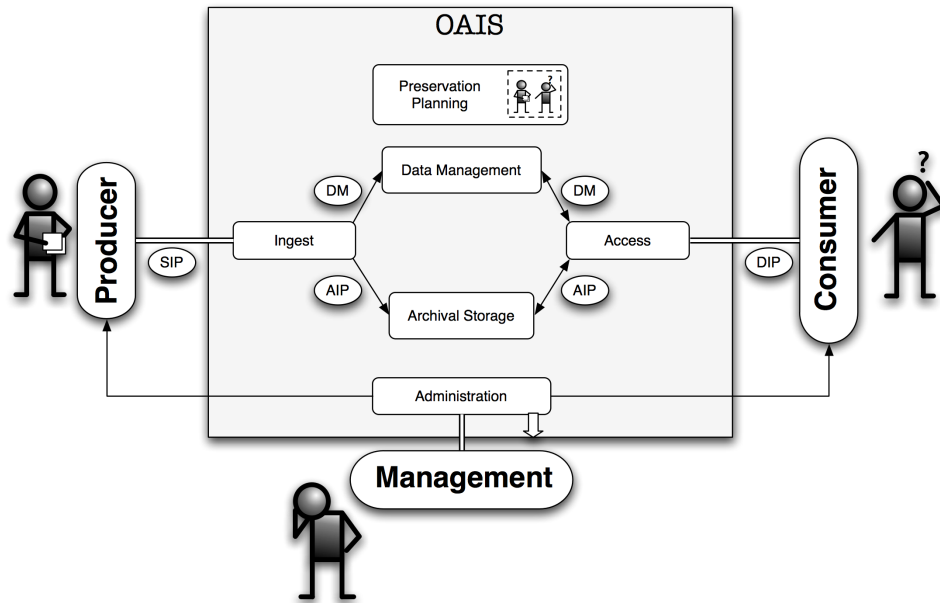


Figura 1: Modelo de referência OAIS

Existirão 3 actores distintos:

Produtor - Utilizador que irá fornecer informação ao sistema. Poderá fazê-lo de forma assistida através de uma interface Web ou em Batch através do envio de SIP ("Submission Information Package") que é um "pacote" de informação com estrutura previamente negociada e especificada entre produtores e repositório;

Administrador - O Administrador irá realizar uma série de operações internas ao sistema: criação e controlo de utilizadores, criação e armazenamento de ontologias para classificação e definição de exposições virtuais, etc.

Consumidor - O Consumidor será o utilizador final do sistema. Irá interactuar com o repositório através duma API específica que lhe permitir realizar uma série de operações: descarregar um conjunto de informação empacotada num DIP ("Dissemination Information Package"), explorar/visitar uma determinada exposição virtual através da navegação num website gerado para o efeito, etc. Cada operação acessível a este utilizador corresponderá à implementação de um disseminador. Não haverá limite para o número de disseminadores.

4.1 Tarefas a Desenvolver

Para concretizar o trabalho é necessário realizar no início as seguintes tarefas, algumas das quais podem prosseguir em paralelo:

1. Escolher um caso de estudo concreto.
2. Começar a especificar o modelo de dados relacional para o repositório e para o SIP. Basear a especificação no estudo e análise dos exemplos das seguintes normas:
 - *Cataloging Cultural Objects (CCO)* - <http://cco.vrafoundation.org/>; ver exemplos, analisar os elementos obrigatórios, etc;
 - Metadados para bibliotecas, arquivos e museus e suas interrelações: <http://firstmonday.org/article/view/1628/1543> (artigo de leitura obrigatória);
 - *Categories for the Description of Works of Art (CDWA)* - http://www.getty.edu/research/publications/electronic_publications/cdwa/introduction.html;

3. Tentar encontrar descrições de colecções na Web: <http://www.openarchives.org/Register/BrowseSites>.
4. Planear com detalhe e rigor as tarefas a realizar para concretizar o projeto, tendo de incluir atividades como:
 - conceber o sistema desenhando a sua arquitectura.
 - modelar os dados e operações do problema.
 - conceber e implementar a Interface do sistema com os utilizadores, cobrindo separadamente cada uma das 3 componentes.
 -

As restantes tarefas serão determinadas pelo plano detalhado que se requer acima (no item 3).