

2010/11: Projecto

Enunciado

Todos os anos são realizadas no DI as **Jornadas de Informática (JOIN)** que agregam, durante alguns dias, toda a comunidade académica local (alunos e professores), investigadores convidados (nacionais e estrangeiros) e o tecido empresarial (um conjunto restrito de empresas que se associam e patrocinam este nosso evento) --- ver, por exemplo, as [JOIN 2010](#).

As inscrições dos participantes são efectuadas através de um portal, preenchendo-se um formulário *on-line* (disponível [aqui](#)) que irá gerar um ficheiro com a seguinte estrutura:

```
nome,email,curso,numero-de-aluno,universidade,curso-1,filiacao-
empresa,jantar-da-conferencia,almocos-no-restaurant-panoramico,porque-se-
inscreveu-nas-join,outro-1,dt
Joaquim da Silva Pereira,joaq_per@netcabo.pt,LCC,23180,,,,0,0,['Procura
projecto/disserta\xc3\xa7\xc3\xa3o de mestrado', 'Interesse pelas
\xc3\xalreas
tem\xc3\xalticas']",,2009/08/26 13:59:04.993 GMT+1
Francisco Pacheco Vieira, fpv@gmail.com,LEI,a41836,,,,0,0,['Procura
1\xc2\xba emprego', 'Interesse pelas \xc3\xalreas
tem\xc3\xalticas']",,2009/08/26 21:40:35.611 GMT+1
Miguel Lopes Costa, costa.miguel@gmail.com,LEI,40626,,,,0,0,['Procura
1\xc2\xba emprego', 'Interesse pelas \xc3\xalreas
tem\xc3\xalticas']",,2009/08/27 18:17:44.077
GMT+1
João Costa, joaomcc@gmail.com,LCC,a41039,,,,0,0,['Procura 1\xc2\xba
emprego', 'Interesse pelas \xc3\xalreas tem\xc3\xalticas']",,2009/08/28
02:01:44.318 GMT+1
Alberto Silva e Sá, setezore@gmail.com,MEI,12251,,,,0,0,['Interesse pelas
\xc3\xalreas tem\xc3\xalticas']",,2009/08/28 15:38:07.065 GMT+1
João Teixeira, joteixe@alunos.uminho.pt,LEI,42050,,,,0,0,['Procura
1\xc2\xba emprego']",,2009/08/28 15:56:23.495 GMT+1
João Teixeira, joteixe@alunos.uminho.pt,LEI,42050,,,,0,0,['Procura
1\xc2\xba emprego']",,2009/08/28 15:57:30.053 GMT+1
Freitas Marques, freitasm@gmail.com,LEI,47140,Universidade do
Minho,,,,0,0,['Interesse pelas \xc3\xalreas tem\xc3\xalticas']",,2009/08/28
16:04:47.952 GMT+1
João Fernando Albertino Veiga,
tino.veiga@gmail.com,MEI,pg1225,,,,0,0,['Interesse pelas \xc3\xalreas
tem\xc3\xalticas']",,2009/08/28 16:20:18.569 GMT+1
Paulo Pereira, pper@gmail.com,MI,11252,,,,0,0,['Procura 1\xc2\xba
emprego', 'Interesse pelas \xc3\xalreas tem\xc3\xalticas']",,2009/08/28
16:28:21.331 GMT+1
```

```

Nuno Cadete,lncadete@gmail.com,,,Universidade do Algarve,Engenharia
Informática,,0,0,['Interesse pelas \xc3\xalreas
tem\xc3\xalticas'],,2009/09/01 16:48:13.091 GMT+1
Cesar Rodrigues,cesar@di.uminho.pt,,,,DI/UMinho,0,0,[],,2009/09/08
16:23:56.947 GMT+1
Paula Novais,paula@di.uminho.pt,,,,DI - Universidade do
Minho,0,0,[],,2009/09/08 16:24:08.379 GMT+1
Francisco Rodrigues,fr@di.uminho.pt,,,,Departamento de
Informática,0,0,['Interesse pelas \xc3\xalreas
tem\xc3\xalticas'],,2009/09/08 17:43:57.098 GMT+1
Hugo Miguel Silva Teixeira,hsmcost@net.sapo.pt,LEI,28241,,,,0,0,[],Procuro
emprego,2009/09/08 21:24:08.457 GMT+1
Vasconcelos Oliveira,v.oliveira@scert.com,,,,,"SUPERTICERT - Certificação
Electrónica, SA",0,0,[],,2009/09/10 15:28:51.332 GMT+1
João Filipe Pinto Ferreira,
pintojoao@gmail.com,LCC,42410,,,,0,0,"['Procura 1\xc2\xba emprego',
'Procura projecto/disserta\xc3\xa7\xc3\xa3o de mestrado', 'Interesse pelas
\xc3\xalreas tem\xc3\xalticas']",,2009/09/10 17:34:22.134 GMT+1

```

A primeira linha indica os vários campos de uma ficha de inscrição, a saber:

- **nome** -> Nome do inscrito
- **email** -> Email do inscrito
- **curso** -> Curso do aluno (na Universidade do Minho)
- **Universidade** -> Universidade do aluno / universitário
- **curso-1** -> Curso do aluno em universidade externa
- **filiação-empresa** -> Empresa do inscrito
- **jantar-da-conferencia** -> Inscrição no jantar da conferência.
- **almocos-no-restaurant-panoramico** -> Inscrição nos almoços da conferência.
- **porque-se-inscreveu-nas-join** -> Lista de razões pelas quais se inscreveu nas JOIN
- **outro_1** -> outras razões pelas quais se inscreveu nas JOIN
- **dt** -> data de inscrição

As restantes linhas contêm dados separados por vírgulas que seguem a ordem de campos determinada na primeira linha. **Nota:** este formato, muito habitual em migração e processamento de dados, dá pelo nome **CSV** (*'comma separated values'*).

Tarefa 1

A primeira tarefa consiste no reconhecimento (*parsing*) do ficheiro das inscrições e validação das mesmas. Existem três tipos diferentes de inscrições: alunos, universitários e empresas. Os primeiros e segundos referem sempre uma universidade, distinguindo-se pelos campos **curso** ou **curso-1** (um destes preenchido no caso de alunos). Uma inscrição só é considerada válida se forem preenchidos campos relativamente ao seu tipo, p.e. um aluno da UM não deverá preencher o campo em que se regista o curso de um aluno de outra universidade (ver [formulário](#)). Há ainda que ter em conta problemas como, por exemplo

- inscrições repetidas (possivelmente incoerentes)
- dados incompletos

perante os quais o programa a desenvolver deverá agir em conformidade.

Tarefa 2

A segunda tarefa consiste em gerar crachás, em LaTeX, para todas as pessoas inscritas. Deverão ser gerados diferentes crachás para os diferentes tipos de inscrições. No caso de um aluno, os crachás deverão conter o primeiro e o último nome, curso e universidade. Um representante de uma empresa deverá ter o primeiro e último nome e a empresa que representa. O mesmo para um universitário.

No seguinte ficheiro [ZIP](#) os alunos podem encontrar código LaTeX para geração dos formatos normalizados (notar, por exemplo, as diferenças de cor). Deverá ser gerado o maior número de crachás por cada página A4 de impressão, para poupar papel.

Tarefa 3

A terceira tarefa consiste na geração de estatísticas acerca dos inscritos: Número de inscritos, inscritos por curso, interesse dos inscritos em áreas temáticas, inscritos para o jantar, inscritos para os almoços na cantina.

O meio de publicação dessas estatísticas fica ao critério dos alunos, por exemplo gerando um documento LaTeX. O recurso a utilitários externos como, por exemplo, [gnuplot](#) para a apresentação de dados de forma gráfica, será valorizado.

Tarefa 4

A última tarefa consiste em gerar análises multi-dimensionais dos dados das inscrições, por exemplo, tabelas do género

Contagem	curso				
porque-se-inscreveu-nas-join	LCC	LEI	MEI	MI	Totais
Interesse pelas áreas temáticas		2	2		4
Procura 1º emprego	2	1		1	4
Procura projecto/dissertação de mestrado		2			2
Totais	2	5	2	1	10

Para saber mais sobre este tipo de tabelas ver, por exemplo, o artigo [Pivot Table](#) da Wikipédia. Como valorização do trabalho, poder-se-á também apresentar a informação das análises multi-dimensionais de forma gráfica, recorrendo, por exemplo, à representação sob a forma de histograma que é disponibilizada pelo gnuplot (exemplos em http://gnuplot.sourceforge.net/demo_4.2/histograms.html)."

Relatório

Deverá ser escrito um pequeno relatório em LaTeX descrevendo os tipos de dados, funções (e respectivos algoritmos) usados para resolver cada uma das tarefas. O relatório deverá adoptar o formato `article`, com tamanho de letra `10pt`, tamanho de página `a4wide`, e ter no máximo 8 páginas. Devido a esta limitação de espaço, não é recomendável incluir todo o código no relatório: apenas se devem listar extractos que se julguem mais interessantes.

Logística

- O projecto deverá ser realizado em grupos de 3 alunos e apresentado aos monitores na semana de **10-14 de Janeiro de 2011** nos turnos práticos da disciplina.
- Os projectos (relatório e código) devem ser submetidos por email até **8 de Janeiro** para o monitor do turno respectivo (Ver [Funcionamento](#)).
- As apresentações de projectos submetidos em regime de melhoria terão lugar no turno PL7.
- Cada turno terá 3 aulas de apoio à realização do projecto (ver sumários previstos para tal em [Sumários](#)). Contudo, os alunos não deverão esperar por essas aulas para iniciar a realização do trabalho - devem começa-lo já. Eventuais dúvidas devem ser tiradas junto do monitor do respectivo turno.
- Para a apresentação deverão trazer o código desenvolvido e o relatório impresso (idealmente em frente-e-verso).
- Os grupos que não enviarem o projecto até ao dia **8 de Janeiro** não poderão ser avaliados.
- As notas dos elementos do grupo podem ser diferentes, de acordo com a sua prestação durante a apresentação.
- Caso se detectem cópias entre os trabalhos, **todos** os membros dos grupos envolvidos ficarão imediatamente reprovados.
- As cotações máximas que podem ser obtidas em cada uma das tarefas são as seguintes:

Tarefa	Cotação (%)
Tarefa 1	40
Tarefa 2	10
Tarefa 3	10
Tarefa 4	15
Relatório	25
Total	100

FAQs

Q1 - Quando temos inscrições repetidas, apenas a data e hora mudam ou existem outros campos que podem ser diferentes? Ou inscrições repetidas têm pelo menos o mesmo nome?

R: Relativamente a inscrições repetidas existem várias hipóteses que vamos deixar ao critério dos alunos. Se mudar apenas a data e a hora, sugere-se que a inscrição mais recente seja a válida.

Q2 - O curso, quando é aluno da Universidade do Minho, é de preenchimento obrigatório?

R: Mais uma vez, fica ao critério dos alunos. Sugere-se que, neste caso, se apenas faltar o curso a inscrição seja válida mas indicada como "incompleta" algures.

Q3 - Qual o nome do ficheiro que contém as inscrições? É o utilizador que decide o nome do ficheiro que quer abrir?

R: O nome do ficheiro que lê as inscrições deve ser passado como parâmetro.

Q4 - Na tarefa 2, o programa deverá invocar o LaTeX e gerar o PDF com os crachás, ou apenas gerar um ficheiro .csv que o LaTeX usará para os gerar?

R: O programa gera o LaTeX a partir das inscrições consideradas válidas.

Q4.1 - No caso de ser a primeira hipótese, para que sistema o fazemos (Linux/Mac OS/Windows) visto que os comandos de invocar funções do sistema são diferentes?

R: Para o sistema operativo onde vão apresentar o trabalho.

Q5 - No enunciado há códigos hexadecimais. É preciso convertê-los para letras com acentos?

R: Sim.

Q6 - Quando se faz o 'parsing' do formulário, são detectados alguns erros. Devemos actualizar o ficheiro ou criar um novo ficheiro apenas com as inscrições válidas?

R: O ficheiro deverá ser lido sem qualquer problema de 'parsing', isto é, a sintaxe está correcta. Depois é que se validam as inscrições.

Q7 - *Também existem crachás para oradores e jobshops, ou apenas para participantes?*

R: Participantes e empresas apenas.

Q8 - *Tenho a seguinte dúvida sobre a leitura do ficheiro csv: nos campos do almoço e jantar, se as opções não são seleccionadas no formulário dá um 0 ao ficheiro e se forem seleccionadas dá um 1?*

R: Exacto: 1 vai ao jantar, 0 não vai.

Q9 - *Ao copiar o ficheiro da página da disciplina para um editor de texto a 1ª e 3ª inscrição não têm tudo na mesma linha. O programa tem que ler como está na página ou podemos juntar as linhas sem que seja considerado um erro?*

R: Devem juntar as linhas que são consideradas erro.

Q10 - *O ficheiro vem sempre ordenado cronologicamente (i.e. a última inscrição é sempre a mais recente)?*

R: Sim.

Q11 - *Numa das inscrições ("Freitas Marques") está preenchido o campo da universidade com "Universidade do Minho". Caso se não tivesse preenchido este campo a inscrição estaria válida; mas, assim, está preenchido um campo que apenas poderia ser preenchido por alunos externos à Universidade do Minho. O programa reconhece-a como inscrição válida ou inválida?*

R: Válida.

Q12 - *Em que SO vamos apresentar o trabalho? Podemos utilizar um portátil de um dos elementos do grupo?*

R: O trabalho poderá ser apresentado num computador de um elemento do grupo, no respectivo SO. Por omissão, será apresentado num dos Mac's disponíveis no laboratório da disciplina.

Q13 - *As inscrições repetidas devem ser acrescentadas à lista de inscrições inválidas ou serão descartadas?*

R: Ver resposta a **Q1** acima. Como o tratamento das inscrições repetidas fica ao critério do grupo, a justificação das decisões tomadas deverá constar do relatório.

Q14 - Para a tarefa 2 criamos dois ficheiros PDF através do pdfpages, é permitido?

R: Sim, estimula-se a utilização de qualquer uma das packages LaTeX existentes.

Q15 - A conversão para letras com acentos tem que ser feita manualmente (i.e uma função que nos fazemos que percorre o ficheiro a procura dos códigos hexadecimais) ou é algo herdado pelos comandos do Prelude/System.IO? Se for pelos comandos do Prelude, qual é a codificação com que o programa tem que ler o ficheiro?

R: O Haskell não faz essa conversão automaticamente. Podem criar uma função que substitua os códigos hexadecimais pelos caracteres correspondentes, ou podem assumir uma codificação para o ficheiro de entrada, e construir um programa que dê resultados coerentes com essa codificação.

Q16 - É necessário criar um ficheiro só com o formulário válido?

R: (Admitindo que estão a perguntar se é necessário criar um ficheiro com os dados do formulário introduzido, mas apenas com os registos considerados válidos): o enunciado não o pede, mas podem fazê-lo.

Q17 - Para empresas usa-se o crashá dos participantes ou há algum específico?

R: Utiliza-se o crashá dos participantes.

Q18 - Ao exportar do Gnuplot é necessário exportar directamente para LaTeX ou pode-se exportar para outro formato (por exemplo jpeg)?

R: Podem optar pelo formato que entenderem. Outra possibilidade é gerarem um ficheiro jpeg que é importado por um ficheiro LaTeX com o template para a apresentação das estatísticas.

Q19 - Quando alguém escolhe mais que um Motivo, o total vai ser maior que o número de inscritos nesse curso. Assim a última linha da tabela multi-dimensional poderá ter valores superiores ao número de inscritos. Ou a tabela é para indicar o nº de opções escolhidas?

R: A tabela pretende consolidar dados relativos às opções escolhidas, e não às inscrições.