

# MPEI – Trabalho Prático

## INFORMAÇÕES

---

Este projecto está a ser realizado por Mário Alexandre Lopes Liberato (NMEC: 84917) e Jorge Miguel Aires de Matos Oliveira (NMEC: 84983). O projecto pretende utilizar alguns módulos desenvolvidos no âmbito dos guiões práticos da unidade curricular de Métodos Probabilísticos para Engenharia Informática, convertidos para Java quando conveniente. Quando não o for, pretende-se utilizar a biblioteca JavaOctave ou outra que sirva o mesmo propósito, a fim de manter o projecto maioritariamente escrito em Java. Em último caso, utilizar-se-á apenas um script Octave.

## CONCEITO INICIAL

---

Programa que sirva para gerir aplicações de pessoas a um determinado emprego. O lado do potencial trabalhador fornece um questionário que depois de completo adiciona os dados a um ficheiro ou base de dados simples (Exemplo: USER ID, SKILLSET ID, EDUCATION LEVEL ID, TIMESTAMP), guardado em CSV. Este ficheiro é processado pelo programa do empregador, que cria, para cada emprego diferente, um modelo de funcionário com o skillset ideal, com o nível de educação ideal, etc. O programa depois filtrará todos os empregados pelo seu índice de semelhança e opcionalmente eliminar aqueles que não possuem semelhança alguma(ou baixa semelhança).

## DESCRIÇÃO BREVE DO PROGRAMA

---

Três programas são utilizados. Um deles será o programa utilizado pela pessoa que submete a aplicação à empresa. Este é denominado por de programa B no documento. O segundo, o programa A, é utilizado pelo empregador para gerir todas as aplicações aos empregos disponíveis. O último programa, dito programa C, serve para gerar dados simulados para utilização pelo programa A.

### DESCRIÇÃO BREVE DO PROGRAMA B

Este programa contém uma interface simples, com um questionário com escolhas múltiplas. Quando o questionário é submetido, este gera dados em CSV, os quais são adicionados ao ficheiro principal, no formato seguinte:

- USER\_ID, SKILL\_ID, EDUCATION\_LEVEL\_ID

O programa gera ainda dados CSV para um segundo ficheiro no formato:

- USER\_NAME, USER\_BIRTHDATE, TIMESTAMP

Estes dados serão armazenados em dois ficheiros CSV. Idealmente, em vez disso, seriam armazenados numa única base de dados que use duas tabelas. Dependendo do progresso do projecto até à data de entrega, esta funcionalidade poderá ou não ser implementada.

### Descrição dos campos das tabelas

- USER\_ID – ID do utilizador, atribuído por ordem crescente. Na segunda tabela, os dados da linha X correspondem ao utilizador com ID X+1.
- SKILL\_ID – ID da capacidade.
- EDUCATION\_LEVEL\_ID – Nível de educação (0 – Nenhum, 1 – Primário, 2 – Secundário, 3 – Licenciatura, 4 – Mestrado, 5 – Doutoramento).
- USER\_NAME – Nome completo do utilizador.
- USER\_BIRTHDAY – Data de nascimento do utilizador.
- TIMESTAMP – A timestamp de criação dos dados.

### DESCRIÇÃO BREVE DO PROGRAMA A

Este programa trata do processamento dos dados obtidos. Essencialmente, lê os dados da primeira tabela, constrói uma lista dos skillsets de cada pessoa, insere os “ideais” como novas pessoas no fim da lista, calcula a distância de Jaccard entre todos os utilizadores e filtra todos os que estiverem abaixo de um determinado nível de semelhança. Depois apresentará os utilizadores mais relevantes para cada emprego, por ordem de nível de escolaridade. Esta parte do programa utilizará predominantemente MatLab.

A interface é simples, e consiste apenas numa opção que serve para iniciar o processamento. Eventualmente, esta interface poderá ser expandida e aumentar em complexidade, mas não faz parte do plano inicial.

### DESCRIÇÃO BREVE DO PROGRAMA C

Este programa gera dados aleatórios para os ficheiros utilizados pelo programa A. Este programa pede apenas o número de utilizadores que se deve criar e gera esse número de utilizadores nas duas tabelas. É gerado usando um programa escrito em Java.