



# **Relatório de Ambientes e Linguagens de Programação para Ciência de Dados TP1/TP2**

Feito por:

- José Moreira - A106826
- Lorenzo Perry - A106865
- Vasco Azevedo - A106823

# TP1:

## Bibliotecas utilizadas:

- typer
- requests
- re
- csv
- datetime

## ALÍNEA A)

### Objetivo

O objetivo da alínea A é implementar uma funcionalidade na CLI que permita listar vagas de emprego disponíveis no site itjobs.pt. Essa funcionalidade retorna informações detalhadas das vagas e permite ao utilizador exportar os dados para um arquivo CSV.

### Descrição da Implementação

- 1. Entrada do Utilizador:**
  - O utilizador fornece o número de vagas a serem listadas através do parâmetro limite.
  - O limite deve ser um valor positivo.
- 2. Consumo da API:**
  - O programa faz uma requisição para o endpoint <https://api.itjobs.pt/job/list> utilizando o limite fornecido.
  - Os resultados são validados e processados. Caso não existam vagas disponíveis, o utilizador é notificado.
- 3. Exibição dos Dados:**
  - As informações das vagas (título, empresa, localização, descrição, data de publicação e salário) são exibidas no terminal.
  - O utilizador tem a opção de exportar os dados para um arquivo CSV.

## ALÍNEA B)

A alínea B implementa a funcionalidade de pesquisa personalizada por vagas de emprego, permitindo filtrar por localidade e empresa.

### Descrição da Implementação

#### 1. Parâmetros de Entrada:

- localidade: Localização geográfica das vagas.
- empresa: Nome da empresa.
- numero\_trabalhos: Número máximo de resultados a serem exibidos.

#### 2. Requisição à API:

- Endpoint utilizado: <https://api.itjobs.pt/job/search>
- Filtros aplicados: nome da empresa, localidade e apenas vagas de tempo integral (definido pelo parâmetro `full_time=true`).

#### 3. Processamento dos Resultados:

- Os resultados da busca são apresentados no terminal, incluindo título, empresa, localização e outros detalhes.
- Possibilidade de salvar os dados em um arquivo CSV.

## ALÍNEA C)

A alínea C implementa uma funcionalidade que permite consultar o salário de uma vaga específica, utilizando o ID da vaga como referência.

### Descrição da Implementação

#### 1. Entrada do Utilizador:

- O utilizador fornece o `job_id` da vaga a ser consultada.

#### 2. Consulta à API:

- Endpoint utilizado: <https://api.itjobs.pt/job/get>
- O salário é extraído diretamente dos dados retornados pela API, caso esteja explicitamente informado.

#### 3. Extração Alternativa:

- Caso o salário não esteja especificado na resposta, o programa utiliza expressões regulares para tentar extrair informações salariais de outros campos, como descrições textuais.

#### 4. **Exibição dos Dados:**

- O salário, se encontrado, é exibido no terminal.

## ALÍNEA D)

A alínea D adiciona uma funcionalidade para buscar vagas de emprego que correspondam a uma lista de habilidades específicas, dentro de um intervalo de datas definido pelo utilizador.

### **Descrição da Implementação**

#### 1. **Parâmetros de Entrada:**

- skills: Lista de habilidades desejadas.
- data\_inicial: Data inicial do intervalo de publicação (formato AAAA-MM-DD).
- data\_final: Data final do intervalo de publicação (formato AAAA-MM-DD).

#### 2. **Processo de Busca:**

- Endpoint utilizado: <https://api.itjobs.pt/job/search.json>.
- Busca paginada para cobrir todas as vagas disponíveis.
- Os resultados são filtrados com base nas datas de publicação.

#### 3. **Exibição e Exportação:**

- As vagas encontradas são exibidas no terminal com detalhes.
- O utilizador pode optar por exportar os resultados para um arquivo CSV.

## TP2:

### **Bibliotecas utilizadas:**

- **typer**
- **requests**
- **csv**
- **BeautifulSoup**
- **json**

## ALÍNEA A)

O objetivo da alínea a) é implementar uma funcionalidade na CLI que permita ao utilizador obter informações detalhadas sobre uma vaga de emprego específica, utilizando a API do site itjobs.pt. Além disso, essas informações são enriquecidas com dados sobre a empresa, extraídos do site AmbitionBox.

O código implementa uma funcionalidade de linha de comando (CLI) que permite obter informações detalhadas sobre uma vaga de emprego específica. Ele combina dois tipos de dados:

1. **Dados da API itjobs.pt:**

Ao fornecer o ID de uma vaga de emprego, o código consulta a API do site itjobs.pt para obter informações básicas sobre a vaga e a empresa que a publicou.

2. **Dados do site AmbitionBox:**

Após obter o nome da empresa na API, o código utiliza técnicas de web scraping para acessar o site AmbitionBox e recolher informações adicionais sobre a empresa, como:

- Classificação geral da empresa (rating);
- Benefícios oferecidos;
- Descrição detalhada.

O programa integra essas informações e apresenta o resultado em formato JSON no terminal, enriquecendo os dados iniciais da vaga com detalhes adicionais sobre a empresa.

Essa solução combina:

- **Consumo de API** (itjobs.pt) para obter dados estruturados;
- **Web scraping** (AmbitionBox) para enriquecer os dados com informações públicas do site.

## ALÍNEA B)

**Objetivo:**

O objetivo da alínea b) é implementar uma funcionalidade na CLI que permita ao utilizador gerar um relatório estatístico detalhado sobre as vagas de emprego disponíveis na API do itjobs.pt.

O código da alínea b) implementa uma funcionalidade na CLI que permite ao utilizador gerar estatísticas detalhadas sobre vagas de emprego, agrupadas por zona geográfica e tipo de trabalho. Ele executa as seguintes ações:

1. **Consulta à API itjobs.pt:**

O programa utiliza a API para obter uma lista completa de vagas de emprego, consultando página por página até que todos os resultados sejam carregados.

2. **Criação- de Estatísticas:**

A partir dos dados das vagas, o código organiza as informações por:

- **Zona:** Local onde a vaga está disponível;
- **Tipo de Trabalho:** Nome ou título da posição.  
Ele cria um dicionário que regista o número de vagas para cada combinação de zona e tipo de trabalho.

3. **Exportação para CSV:**

As estatísticas geradas são gravadas em um arquivo CSV com três colunas:

- Zona;
- Tipo de Trabalho;
- Número de Vagas.

Ao final, o programa informa que o arquivo CSV foi criado com sucesso, permitindo ao utilizador acessar e analisar os dados.

## ALÍNEA C)

**Objetivo:**

O código implementa uma funcionalidade na CLI que permite ao utilizador gerar estatísticas detalhadas sobre as habilidades mais requisitadas em vagas de emprego, consultando o site AmbitionBox. Ele realiza as seguintes ações:

**Consulta ao site AmbitionBox:**

- O programa gera uma URL de pesquisa baseada no título de trabalho fornecido pelo utilizador.

- Faz requisições à página de pesquisa de empregos e, em seguida, navega para os detalhes de cada vaga para coletar informações específicas sobre as habilidades listadas.

#### **Criação de Estatísticas:**

- As habilidades extraídas de cada vaga são organizadas em um dicionário que registra a frequência de cada habilidade.
- As 10 habilidades mais frequentes são selecionadas para compor o relatório.

A funcionalidade permite ao utilizador identificar rapidamente as competências mais valorizadas no mercado de trabalho para um título específico.

## **ALÍNEA D)**

#### **Objetivo:**

O objetivo desta alínea é explorar uma nova fonte de dados (Indeed) para enriquecer as informações de uma vaga de emprego, oferecendo uma visão mais completa e detalhada ao utilizador. Isso amplia as possibilidades de análise e torna a ferramenta mais robusta.

O código da alínea (d) implementa uma funcionalidade na CLI que permite enriquecer os dados de uma vaga de emprego com informações adicionais sobre a empresa, obtidas do site **Indeed**. Ele executa as seguintes ações:

1. **Consulta à API itjobs.pt:**

O programa utiliza o ID da vaga fornecido pelo utilizador para buscar informações básicas sobre a vaga e o nome da empresa.

2. **Busca de Informações no Indeed:**

Com o nome da empresa obtido pela API, o programa realiza uma requisição ao site **Indeed** e utiliza técnicas de web scraping para recolher:

- **Rating da Empresa:** Classificação geral dada pelos funcionários.
- **Número de Avaliações:** Total de avaliações registradas sobre a empresa.
- **Descrição da Empresa:** Resumo sobre a empresa, incluindo características e objetivos.

3. **Combinação dos Dados:**

Os dados obtidos da API do **itjobs.pt** são complementados com as informações extraídas do Indeed e organizados em um objeto JSON.

4. **Exportação para CSV:**

O utilizador tem a opção de salvar os dados completos em um arquivo CSV, permitindo análise ou consulta futura.

**Nota:** Apesar da implementação estar concluída, não foi possível executar e validar completamente a funcionalidade da alínea d), pois o site Indeed rejeita as requisições realizadas pelo programa. Isso ocorre devido à configuração dos headers utilizados, que não são suficientes para contornar os mecanismos de proteção do site contra o web scraping.