

# Sistema de Consulta e Análise de Publicações Científicas

---

Segunda parte do semestre: 2024-11-04 a ...

Luís Filipe Cunha e José Carlos Ramalho

## Sinopsis

Neste documento, descreve-se o enunciado do projeto a desenvolver na edição 2024 da unidade curricular "*Algoritmos e Técnicas de Programação*".

## Descrição do Projeto

Neste projeto, estarás responsável por desenvolver um sistema em Python que permite criar, atualizar e analisar publicações científicas. Com base num dataset de publicações, o sistema deverá possibilitar a pesquisa de artigos usando filtros relevantes, tais como a data de publicação, as palavras-chave, autores, etc. Devem ainda ser gerados relatórios (mostrando gráficos ilustrativos com estatísticas) detalhados para a análise de métricas dos artigos e dos seus autores.

## Requisitos do Sistema

1. **Carregamento da Base de Dados:** O programa no arranque deverá carregar para memória o dataset que deverá estar guardado no ficheiro de suporte à aplicação;
2. **Criação de Publicações:** O utilizador deve poder criar um artigo especificando um título, resumo, palavras-chave, DOI, uma lista de autores e sua afiliação correspondente, url para o ficheiro PDF do artigo, data de publicação e url do artigo;
3. **Atualização de Publicações:** O sistema deve permitir a atualização da informação de uma publicação, nomeadamente a data de publicação, o resumo, palavras-chave, autores e afiliações;
4. **Consulta de Publicações:** O sistema deve permitir pesquisar publicações. Esta pesquisa deve permitir filtros por título, autor, afiliação, data de publicação e palavras-chave. Deve ainda ser possível ordenar as publicações encontradas pelos títulos e pela data de publicação;
5. **Análise de publicações por Autor :** O sistema deve permitir listar os autores e aceder aos artigos de cada autor da lista. Os autores devem aparecer ordenados pela frequência dos seus artigos publicados e/ou por ordem alfabética;
6. **Análise de publicações por palavras-chave:** O sistema deve permitir a pesquisa e visualização das palavras-chave do dataset. As palavras-chave devem estar ordenadas pelo seu número de ocorrências nos artigos e/ou por ordem alfabética. O sistema deve também permitir visualizar a lista das publicações associadas a cada palavra-chave;
7. **Estatísticas de Publicação:**

O sistema deve apresentar relatórios que incluam os seguintes gráficos:

- Distribuição de publicações por ano.
  - Distribuição de publicações por mês de um determinado ano.
  - Número de publicações por autor (top 20 autores).
  - Distribuição de publicações de um autor por anos.
  - Distribuição de palavras-chave pela sua frequência (top 20 palavras-chave).
  - Distribuição de palavras-chave mais frequente por ano.
8. **Armazenamento dos Dados:** Quando o utilizador decidir sair da aplicação ou tiver seleccionado o armazenamento dos dados, a aplicação deverá guardar os dados em memória no ficheiro de suporte;
9. **Importação de Dados:** Em qualquer momento, deverá ser possível importar novos registos dum outro ficheiro que tenha a mesma estrutura do ficheiro de suporte;
10. **Exportação parcial de dados:** Em qualquer momento, deverá ser possível exportar para ficheiro os registos resultantes de uma pesquisa (apenas o subconjunto retornado pela pesquisa).

## Requisitos Técnicos

- O sistema deve ser implementado em **Python**.
- Utilize estruturas de dados apropriadas para armazenar a informação sobre as publicações, como dicionários, listas ou listas de dicionários.
- Deve desenvolver **duas interfaces** para interação com o sistema:
  - (1) Interface de linha de comando (CLI)
  - (2) Interface gráfica
- Use bibliotecas Python relevantes para funcionalidades como gráficos (por exemplo, a biblioteca `matplotlib` que já tem sido usada nas aulas).
- Implemente um mecanismo de armazenamento persistente para guardar as publicações, por exemplo um ficheiro JSON.

## Linha de Comando

Um utilizador da interface de linha de comandos deve poder executar as seguintes operações:

- **Help:** Imprimir uma mensagem de ajuda com os comandos disponíveis.
- **Criar Publicação:** Dada a informação de uma publicação, o sistema deve permitir a criação da mesma.
- **Consulta de Publicação:** Dado um identificador de uma publicação, o sistema deve permitir consultar a informação detalhada dessa publicação, de forma **clara e organizada**.
- **Consultar Publicações:** Listar as publicações presentes no sistema. Esta listagem deve permitir filtrar publicações com base em critérios relevantes tais como título, autor, afiliação, palavras-chave e data de publicação.
- **Eliminar Publicação:** Dado um identificador de uma publicação, o sistema deve eliminar a publicação correspondente.

- **Relatório de Estatísticas:** Gerar relatórios sobre a frequência de palavras-chave, número de publicações por autor, e número de publicações por ano.
- **Listar Autores:** Listar todos os autores existentes no sistema, bem como as publicações associadas a cada autor.
- **Importar Publicações:** Importar novas publicações existentes num ficheiro novo (com estrutura semelhante à do dataset fornecido) e adicionar estas ao modelo em memória.
- **Guardar Publicações:** Guardar em ficheiro toda a informação em memória.

## Avaliação (alguns critérios)

Este projeto é de realização obrigatória e tem uma nota mínima de 10 valores conforme divulgado no início do semestre.

A avaliação é sempre algo subjetivo, no entanto, vamos fixar um conjunto de critérios mínimos para obtenção da nota 10:

- Entrega do relatório em Markdown;
- Ler a informação do dataset para uma estrutura de dados em memória;
- Guardar a estrutura de dados em memória num ficheiro;
- Inserir um registo novo;
- Apagar um registo;
- Consultar um registo (de acordo com um determinado critério);
- Listar os registos (apenas uma parte dos campos);
- Criar e mostrar em forma de gráfico uma das distribuições pedidas.

Todas as restantes funcionalidades referidas no enunciado serão avaliadas para majorações acima da nota 10.

## Entrega do Projeto

Na entrega do projeto deverás entregar:

- Relatório técnico em Markdown com estrutura semelhante aos dos TPC;
- O código-fonte do sistema em Python;
- A documentação detalhada que explique como usar o sistema (pode estar incluída no relatório);
- Exemplos de execução do sistema (podem estar incluídos no relatório).

Este projeto é uma oportunidade para trabalhares com dados estruturados e aplicares técnicas de programação para desenvolver uma ferramenta de consulta e análise científica.

Boa sorte e bom trabalho!