

## Arquivo de Notícias do Sporting Clube da Covilhã

### Proposta de Projeto. Lic. Eng. Informática

**Orientador:** [Ricardo Campos](mailto:ricardo.campos@ubi.pt) (ricardo.campos@ubi.pt)

### Objetivos

No ano do centenário do Sporting Clube da Covilhã, pretende-se utilizar os recursos do [Arquivo.pt](http://Arquivo.pt) (plataforma de preservação de conteúdos da web portuguesa) para tornar acessível ao público o conjunto de notícias publicadas ao longo da última década nas plataformas digitais dos principais meios de comunicação social da região com foco no principal clube da cidade da Covilhã. Em particular, este projeto visa a disponibilização desses conteúdos através do desenvolvimento de um website dedicado e de um sistema de pesquisa que permita aos utilizadores da plataforma web, recuperar informações acerca do clube a partir dos dados coletados do Arquivo.pt. Pretende-se com este projeto contribuir para a preservação do património desportivo local e a memória coletiva do Sporting Clube da Covilhã.

### Plano de Trabalho

**T1:** Definição e planeamento (2 semanas)

**T2:** Recolha e armazenamento dos dados do Arquivo.pt com recurso ao desenvolvimento de scripts em Python. Aplicação de medidas de similaridade textual com vista à eliminação de notícias duplicadas (4 semanas)

**T3:** Criação de uma imagem Docker com vista à integração dos vários componentes (1 semana)

**T4:** Desenvolvimento do website (4 semanas)

**T5:** Implementação do sistema de pesquisa com recurso ao Elastic Search (base de dados nosql) e ao algoritmo de recuperação de informação BM25 (3 semanas)

**T6:** Testes Finais (2 semanas)

**T7:** Documentação e Apresentação (2 semanas)

A Tabela 1 apresenta a distribuição das tarefas por cada uma das 15 semanas.

Tabela 1: Cronologia das tarefas (T) por semana (S).

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
T1	█	█													
T2			█	█	█	█									
T3							█								
T4								█	█	█	█				
T5										█	█	█			
T6												█	█		
T7														█	█

## Requisitos Técnicos / Académicos

- conhecimentos sólidos em Python (para a recolha de dados) ou disposição para aprender.
- conhecimento em linguagens de programação web (HTML, CSS, JavaScript) ou Flask para o desenvolvimento do website.
- experiência com Elasticsearch (base de dados nosql) ou disposição para aprender.
- familiaridade com a criação de imagens Docker.
- bons conhecimentos de programação, engenharia de software e composição web

## Resultados Esperados

- script python de obtenção dos dados e respetivo dataset coletado
- website responsivo (código a disponibilizar no github do aluno) e disponível online
- relatório
- candidatura ao Prémio Arquivo.pt 2024 (lista de vencedores dos prémios anteriores: <https://sobre.arquivo.pt/pt/colabore/premios-arquivo-pt/>)