

Proposta para Dissertação de Mestrado em Engenharia Informática (2019/2020)

Título: Investigação, desenho e implementação de uma solução web baseada na nuvem para a recolha de informação de eventos desportivos usando técnicas de mineração de dados

Orientador: Prof. Lúcio Studer Ferreira

Co-orientador: Prof. Sebastião Pais

Motivação: Nos dias que correm existem com cada vez maior frequência eventos desportivos. Há sempre uma grande quantidade de informação dispersa na web associada aos vários eventos. Neste contexto, uma empresa de design como a Mixed Creations, que faz cartazes alusivos aos jogos de futebol e de futsal de um alargado conjunto de equipas às quais presta serviço, tem de dispendir uma considerável quantidade de tempo em cliques pelos sítios web das várias associações de futebol existentes, para pesquisar as informações dos respetivos jogos por campeonato.

Objetivos: O principal objetivo deste trabalho é a investigação, desenho e implementação de uma solução web baseada na nuvem que aplique técnicas de mineração de dados, em especial web scraping, para agregar informação referente a eventos desportivos de forma automática e transparente ao utilizador. Tipos de informação a ser agregados serão por exemplo o local, dia, hora em que decorrerão os jogos que envolvem determinadas equipas que estão a ser acompanhadas, assim como informação resultante dos jogos, tal como o resultado final e marcadores, entre outros.

Tarefas a Realizar

- Tarefa 1: Estudo da base tecnológica.
- Tarefa 2: Pesquisa do problema e estado da arte.
- Tarefa 3: Análise de requisitos de uma solução deste tipo.
- Tarefa 4: Desenho e implementação de uma solução real.
- Tarefa 5: Testes de avaliação e aceitação do protótipo desenvolvido.
- Tarefa 6: Escrita da dissertação e artigo científico.

Tecnologias e Ferramentas a Utilizar

- Linguagem de programação Python.
- BeautifulSoup, biblioteca para captura de informações de páginas da web.
- Flask, padrão para desenvolvimento de aplicações web.
- Heroku, plataforma para aplicações na nuvem.

Resultados Esperados

- Uma aplicação computacional validada.
- Uma publicação descrevendo as metodologias, implementação e os resultados da validação.
- Uma dissertação.

Cronograma

	Out 19	Nov 19	Dez 19	Jan 20	Fev 20	Mar 20	Abr 20	Mai 20	Jun 20
Tarefa 1									
Tarefa 2									
Tarefa 3									
Tarefa 4									
Tarefa 5									
Tarefa 6									

Referências

- [1] D.M. Thomas e S. Mathur, “Data Analysis by Web Scraping using Python”, em *Proc. de ICECA 2019: 3rd International conference on Electronics, Communication and Aerospace Technology*, Coimbatore, India, junho 2019. (<https://ieeexplore.ieee.org/document/8822022>)
- [2] P. Mudgil, H. Bansal, J. Pruthi, e P. Batra, “Information Retrieval using Sports Based Ontology”, em *BPIT's International Journal of Technology & Management* Vol. 4, Issue I, Maio 2018. (<http://www.bpitindia.com/pdf/Bodh/Bodh-Vol-4-Issue-I.pdf#page=44>)
- [3] Pengfei Wang and Aleksandar Kuzmanovic, “What’s the Score?: A First Look at Sports Live Data Feed Services”, em *Proc. de IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, Honolulu, USA, abril 2018. (<https://statsweb-wpengine.netdna-ssl.com/wp-content/uploads/2017/08/22What%E2%80%99s-the-Score22-A-First-Look-at-Sports-Live-Data-Feed-Services.pdf>)