

Aplicação para Análise de Colocações no Ensino Superior

Relatório Final do Projeto Integrador

up202008569 – Ana Rita de Oliveira Carneiro

up202006814 – Hugo Filipe Rodrigues de Almeida

up201905451 – José Carlos Cardiano Mota Gonçalves da Cunha

up202005330 – Raquel Marques Carneiro



Licenciatura em Engenharia Informática e Computação

Proponente/Tutor na U.Porto: João Pascoal Faria

23/6/2023

Conteúdo

1 Introdução.....	2
1.1 Enquadramento.....	2
1.2 Objetivos e resultados esperados	2
1.3 Estrutura do relatório.....	3
 2 Metodologia utilizada e principais atividades desenvolvidas	4
2.1 Metodologia utilizada	4
2.2 Intervenientes, papéis e responsabilidades.....	4
2.3 Atividades desenvolvidas	4
 3 Desenvolvimento da solução	6
3.1 Requisitos	6
3.2 Arquitetura e tecnologias.....	6
3.3 Solução desenvolvida	7
3.4 Validação	9
 4 Conclusões	10
4.1 Resultados alcançados	10
4.2 Lições aprendidas	10
4.3 Trabalho futuro	10
 5 Referências	12

1 Introdução

1.1 Enquadramento

A análise sofisticada e esclarecedora dos resultados das colocações no ensino superior envolve considerar diversos aspetos além das notas dos últimos colocados. Os responsáveis utilizam técnicas estatísticas e ferramentas de visualização de dados para extrair insights relevantes.

Por exemplo, é importante analisar a relação entre o número de vagas e o número de candidatos para avaliar a competitividade de cada curso e determinar se há oferta suficiente em relação à procura. Também é interessante comparar a preferência dos candidatos a cursos semelhantes em diferentes instituições de ensino, o que pode revelar diferenças de reputação ou qualidade de ensino.

Além disso, a análise das tendências ao longo do tempo permite identificar padrões e prever possíveis mudanças na procura futura. É relevante considerar não apenas as notas de candidatura, mas também a posição do candidato na lista de ordenação e opção escolhida.

1.2 Objetivos e resultados esperados

O objetivo do projeto é criar uma aplicação web que seja capaz de extrair dados de candidaturas e colocações no ensino superior diretamente do site da DGES. Esta aplicação permitirá aos utilizadores explorar os dados de forma flexível, selecionando parâmetros específicos, como cursos, instituições e anos de interesse. Com base nesses dados, a aplicação fornecerá estatísticas e visualizações dinâmicas, auxiliando na compreensão da adequação da oferta em relação à procura, identificando tendências ao longo do tempo e permitindo comparações entre diferentes cursos e instituições de ensino.

Tendo tudo isto em conta, as expectativas de resultados alcançados no final do projeto são o desenvolvimento de uma aplicação operacional, comprovada através de testes utilizando dados do período 2018 a 2022, implementada num servidor apropriado, com código fonte disponível num repositório adequado, bem como a elaboração deste relatório.

1.3 Estrutura do relatório

O relatório inicia com a introdução que contextualiza a importância da análise das colocações no ensino superior e apresenta os objetivos e resultados esperados do projeto. De seguida, é descrita a metodologia utilizada, destacando as atividades desenvolvidas ao longo do projeto e os intervenientes envolvidos, bem como a utilização da plataforma GitHub para facilitar a colaboração.

O desenvolvimento da solução é abordado, mencionando os requisitos funcionais e não funcionais do projeto, a arquitetura em camadas adotada e as tecnologias utilizadas, como Python, Flask e bibliotecas para web scrapping e geração de gráficos. A solução desenvolvida, incluindo a interface de utilizador e a validação realizada, também é incluída.

Por fim, são apresentadas as conclusões do projeto, abordando os resultados alcançados, as lições aprendidas ao longo do processo e as sugestões para trabalhos futuros, como melhorias na interface, expansão de funcionalidades, entre outras. Esta estrutura do relatório abrange de forma concisa os principais aspetos do projeto desenvolvido.

2 Metodologia utilizada e principais atividades desenvolvidas

2.1 Metodologia utilizada

Em relação à metodologia utilizada, optou-se por adotar uma abordagem iterativa, dividindo o trabalho em sprints mensais. Esses sprints funcionaram como unidades de trabalho bem definidas, permitindo uma gestão mais eficiente e um acompanhamento mais preciso do progresso alcançado.

Para garantir uma colaboração eficaz entre os membros do grupo, utilizou-se a plataforma GitHub. Esta plataforma oferece recursos que facilitam a organização do código-fonte, o controlo de versões e a colaboração simultânea num mesmo projeto. Através do GitHub, os elementos do grupo puderam compartilhar o código, gerir tarefas e manter um histórico detalhado das alterações feitas ao longo do desenvolvimento.

2.2 Intervenientes, papéis e responsabilidades

O tutor e proponente do projeto é o professor João Pascoal Faria, que desempenhou um papel de orientação e fornecimento de feedback ao longo do projeto. O grupo responsável pelo desenvolvimento do projeto é composta por quatro membros: Ana Carneiro, Hugo Almeida, José Cunha e Raquel Carneiro.

No processo de desenvolvimento de código, cada membro desempenhou um papel ativo no projeto, contribuindo em várias etapas não tendo sido realizadas atribuições formais de tarefas específicas a cada membro. Em vez disso, o grupo trabalhou de forma colaborativa, compartilhando ideias, conhecimentos e contribuindo coletivamente para as diferentes etapas do projeto.

Na fase final, para maximizar eficiência o grupo dividiu as tarefas de escrita do relatório, realizado pelos membros José Cunha e Raquel Carneiro, e elaboração do póster, realizado pelos membros Ana Carneiro e Hugo Almeida.

2.3 Atividades desenvolvidas

Durante o período de desenvolvimento do projeto, foi estabelecido um cronograma de reuniões semanais com o mentor. Essas reuniões desempenharam um papel crucial ao proporcionar um ambiente para obter feedback, esclarecer requisitos e garantir que o grupo estivesse alinhado com os objetivos a serem alcançados. Através

destas interações, foi possível orientar o trabalho de forma consistente e manter um progresso contínuo.

Além das reuniões com o tutor, o grupo também se reunia regularmente para discutir o trabalho em andamento e fornecer apoio mútuo para avançar nas tarefas. Estes encontros adicionais foram valiosos para garantir uma comunicação eficaz e uma colaboração harmoniosa entre os membros.

Dado que não foram estabelecidas entregas intermédias, optou-se por realizar apresentações informais do trabalho desenvolvido ao longo do tempo. Todos os envolvidos fizeram esforços para cumprir as metas estabelecidas e seguir a distribuição de tarefas sugerida. Essa distribuição proposta previa que o mês de fevereiro fosse dedicado à análise de requisitos e seleção das tecnologias adequadas. Já os meses de março e abril foram designados para o desenvolvimento iterativo dos módulos da aplicação. Em maio, o foco estava nos testes, refinamento e disponibilização da aplicação, enquanto junho foi reservado para a redação do relatório final do projeto e a criação do poster de apresentação. Esta abordagem ajudou a manter um progresso constante e uma clara direção para o trabalho desenvolvido.

3 Desenvolvimento da solução

3.1 Requisitos

Ambos os requisitos funcionais e não funcionais desempenham um papel fundamental no desenvolvimento de uma aplicação eficaz e bem-sucedida, garantindo que as necessidades dos utilizadores sejam respeitadas e que o sistema opere conforme os padrões desejados.

Os requisitos funcionais são as capacidades e funcionalidades específicas que a aplicação deve fornecer. No contexto da solução desenvolvida, os requisitos funcionais são:

- Extrair dados de candidaturas e colocações do site da DGES;
- Armazenar os dados extraídos;
- Filtrar os dados com base em parâmetros especificados pelo utilizador (cursos, ano, estudantes colocados ou estudantes candidatos em 1ª opção)
- Gerar gráficos para visualizar os resultados da análise através do filtramento de dados (gráfico de retas em que a reta das ordenadas é a média e a da abcissas o número de alunos).

Por outro lado, os requisitos não funcionais dizem respeito às características gerais do sistema, sendo estes os que foram identificados:

- Capacidade de lidar com uma grande quantidade de dados;
- As consultas à base de dados e geração de gráficos devem ser rápidas e responsivas;
- Minimizar a ocorrência de erros durante a extração e análise de dados;
- Opcionalmente, tornar a interface adaptável para dispositivos móveis;
- Permitir a adição de funcionalidades adicionais ou expansão para análises mais complexas no futuro.

3.2 Arquitetura e tecnologias

A arquitetura adotada para a implementação do projeto foi baseada na arquitetura em camadas. Esta divide o sistema em camadas distintas, tendo cada camada uma responsabilidade específica. Neste tipo de arquitetura a comunicação entre camadas é unidirecional, em que a camada superior chama a camada inferior. Para além disto, ajuda a separar preocupações e facilita a manutenção e evolução do sistema.

Em relação às tecnologias utilizadas, a linguagem de programação em que o projeto foi desenvolvido foi Python, dada a variedade de bibliotecas para web scrapping.

O framework utilizado foi Flask porque fornece uma estrutura para criação de rotas, manipulação de solicitações HTTP e renderização de templates. Para além disto, foram incluídas as seguintes bibliotecas: matplotlib (criação de gráficos), sqlite3 (para realizar operações com a database), csv (para ler ficheiros CVS), requests (criar pedidos HTTP para aceder às tabelas do site da DGES) e BeautifulSoup (retirar as informações dos gráficos do site). A base de dados foi criada em sqlite3 e para criar a interface de utilizador foram criados ficheiros HTML e CSS.

Uma das dificuldades técnicas com que nos deparámos foi a estrutura complicada do site da [DGES](#) (o site que está em hiperligação é o de 2022) que pode alterar e por isso é importante entendê-la e adaptar o código de scrapping conforme necessário. Também foi preciso garantir que a base de dados fosse dimensionada corretamente para que as consultas com volumes de dados extensos sejam otimizadas. Para além disto, o grupo experimentou diferentes abordagens para criar gráficos, recorrendo a diferentes bibliotecas até os resultados desejados serem alcançados.

3.3 Solução desenvolvida

A solução desenvolvida é uma aplicação web que permite extrair dados de candidaturas e colocações no ensino superior a partir do site da DGES. Com base na informação do site, a aplicação gera estatísticas e automatiza a produção de gráficos, fornecendo uma análise expedita e interativa. Com uma interface web amigável, esta solução proporciona uma ferramenta eficaz para explorar e compreender os resultados das colocações no ensino superior, auxiliando na tomada de decisões para os responsáveis pela gestão académica. Esta aplicação ficará alojada numa máquina virtual, mas até lá apenas será disponível numa máquina local que tenha o projeto.

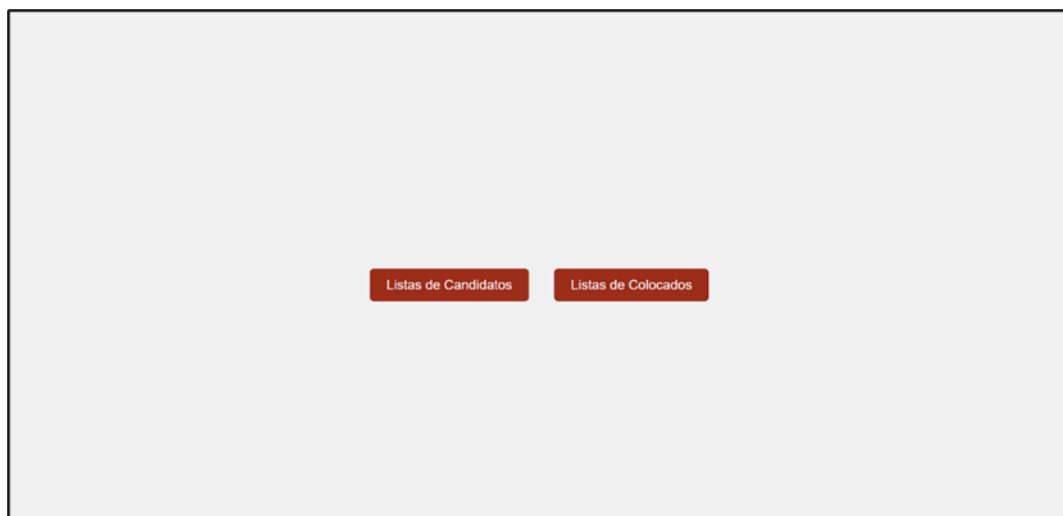


Fig.1: Página inicial onde o utilizador seleciona que tipo de dados procura

Na imagem acima está representada a página web que mostra dois botões: “Lista de Candidatos” e “Lista de Colocados”. Estes botões servem para o utilizador escolher se quer procurar informações sobre as candidaturas ou colocações dos estudantes no ensino superior, e ser direcionado para as respetivas páginas da aplicação em que poderá visualizar o tipo de lista escolhida.



Seleciona o(s) curso(s)

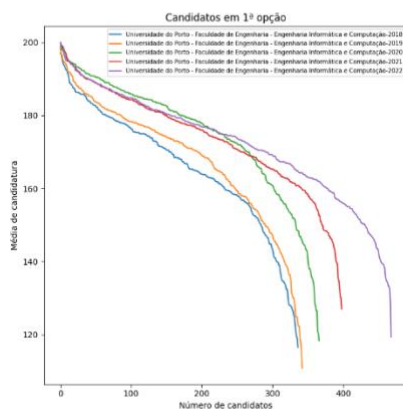
Gerar Gráfico

- ☐ Universidade dos Açores - Faculdade de Ciências Agrárias e do Ambiente - 2018 - Ciências Agrárias
- ☐ Universidade dos Açores - Faculdade de Ciências Agrárias e do Ambiente - 2019 - Ciências Agrárias
- ☐ Universidade dos Açores - Faculdade de Ciências Agrárias e do Ambiente - 2020 - Ciências Agrárias
- ☐ Universidade dos Açores - Faculdade de Ciências Agrárias e do Ambiente - 2021 - Ciências Agrárias
- ☐ Universidade dos Açores - Faculdade de Ciências Agrárias e do Ambiente - 2022 - Ciências Agrárias
- ☐ Universidade dos Açores - Faculdade de Ciências Agrárias e do Ambiente - 2018 - Ciências Farmacêuticas (Preparatórios)
- ☐ Universidade dos Açores - Faculdade de Ciências Agrárias e do Ambiente - 2019 - Ciências Farmacêuticas (Preparatórios)
- ☐ Universidade dos Açores - Faculdade de Ciências Agrárias e do Ambiente - 2020 - Ciências Farmacêuticas (Preparatórios)
- ☐ Universidade dos Açores - Faculdade de Ciências Agrárias e do Ambiente - 2021 - Ciências Farmacêuticas (Preparatórios)
- ☐ Universidade dos Açores - Faculdade de Ciências Agrárias e do Ambiente - 2022 - Ciências Farmacêuticas (Preparatórios)
- ☐ Universidade dos Açores - Faculdade de Ciências Agrárias e do Ambiente - 2018 - Medicina Veterinária (Preparatórios)
- ☐ Universidade dos Açores - Faculdade de Ciências Agrárias e do Ambiente - 2019 - Medicina Veterinária (Preparatórios)
- ☐ Universidade dos Açores - Faculdade de Ciências Agrárias e do Ambiente - 2020 - Medicina Veterinária (Preparatórios)
- ☐ Universidade dos Açores - Faculdade de Ciências Agrárias e do Ambiente - 2021 - Medicina Veterinária (Preparatórios)
- ☐ Universidade dos Açores - Faculdade de Ciências Agrárias e do Ambiente - 2022 - Medicina Veterinária (Preparatórios)
- ☐ Universidade dos Açores - Faculdade de Ciências Agrárias e do Ambiente - 2018 - Natureza e Património
- ☐ Universidade dos Açores - Faculdade de Ciências Agrárias e do Ambiente - 2019 - Natureza e Património
- ☐ Universidade dos Açores - Faculdade de Ciências Agrárias e do Ambiente - 2020 - Natureza e Património
- ☐ Universidade dos Açores - Faculdade de Ciências Agrárias e do Ambiente - 2021 - Natureza e Património
- ☐ Universidade dos Açores - Faculdade de Ciências Agrárias e do Ambiente - 2022 - Natureza e Património
- ☐ Universidade dos Açores - Faculdade de Ciências Sociais e Humanas - 2018 - Arquitetura (Preparatórios)
- ☐ Universidade dos Açores - Faculdade de Ciências Sociais e Humanas - 2022 - Comunicação e Relações Públicas
- ☐ Universidade dos Açores - Faculdade de Ciências Sociais e Humanas - 2018 - Educação Básica

Fig.2: Página em que o utilizador escolhe os cursos cujas informações quer comparar

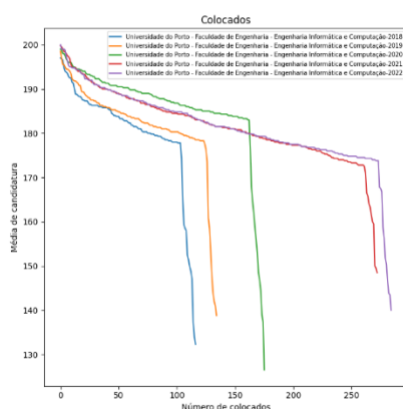
Na imagem acima está representada a página web que permite ao utilizador selecionar cursos para gerar um gráfico, ou de colocados ou de candidatos. Esta página exibe uma lista de cursos (instituição – ano - curso) com caixas de seleção que o utilizador pode marcar conforme o seu interesse. A escolha pode ser feita ao clicar no botão “Gerar Gráficos”, apresentado também na imagem, ou na tecla “Enter”. Quando a informação é submetida, o gráfico é gerado e o utilizador é redirecionado para uma página que apresentará o mesmo.

Resultado



Anexo 3: Página que mostra o gráfico de Candidatos em 1ª opção de L.EIC de 2018-2022

Resultado



Anexo 4: Página que mostra o gráfico de Colocados de L.EIC de 2018-2022

Na imagem acima está representada a página web que exibe os resultados do gráfico gerado. Como se pode observar, o gráfico apresenta um título e legenda de dados para o utilizador poder compreender a informação que lhe está a ser apresentada. Para além disso, também está disponível a opção de retornar à página inicial ao clicar no botão “Voltar à Página Inicial” no canto superior esquerdo do ecrã.

3.4 Validação

Como forma de verificar a eficácia da solução desenvolvida, realizamos reuniões semanais com o proponente do projeto, que nos forneceu feedback valioso e conselhos pertinentes para o progresso e aprimoramento do trabalho. Além disso, comparávamos regularmente os resultados obtidos com os gráficos finais fornecidos pelo proponente em arquivos Excel, a fim de validar a precisão e a qualidade da nossa solução.

Para avaliar a usabilidade da interface, convidamos estudantes da FEUP a testarem a plataforma e a fornecerem feedback informal sobre a sua experiência. Os resultados obtidos foram consistentemente positivos, o que indica que a solução apresentada é intuitiva e de fácil utilização. Esta abordagem de testes com utilizadores reais contribuiu para aperfeiçoar a interface e garantir uma experiência satisfatória para os utilizadores finais.

4 Conclusões

4.1 Resultados alcançados

Os resultados alcançados incluem a exibição de informações sobre cursos e candidatos, juntamente com a geração de gráficos relacionados aos candidatos ou candidatos colocados nos cursos selecionados. O produto é exibido no site complementar também implementado. A solução atingida deriva da concretização das metas propostas.

O trabalho realizado neste projeto foi dividido de forma equitativa entre os quatro membros do grupo, garantindo que cada elemento contribuisse com 25% do esforço total. Esta abordagem colaborativa permitiu que todas as responsabilidades fossem distribuídas de forma justa, resultando numa contribuição equilibrada de cada membro para o sucesso do projeto.

4.2 Lições aprendidas

O projeto selecionado proporcionou vários momentos de aprendizagem integrais para o nosso futuro. Aprendemos a integrar Flask, um framework web em Python, com uma base de dados SQLite. Compreendemos como estabelecer conexões, executar consultas e recuperar dados da base de dados. Tivemos oportunidade de aprofundar os nossos conhecimentos sobre rotas, templates, requisições HTTP e geração de gráficos. Estas competências são cruciais para a construção de aplicações web funcionais, para além de serem fundamentais na visualização de dados de forma significativa e comunicação de insights.

Um aspeto que tentámos reforçar foi a organização do código, utilizando ficheiros e diretórios para separar funcionalidades específicas. Esta abordagem promove a legibilidade e manutenção do projeto. Para além disto, ao trabalhar em grupo, dividir tarefas e partilhar resultados, conseguimos maximizar a eficiência e criamos um ambiente colaborativo. Tudo isto contribui para o nosso crescimento profissional, pelo aumento das nossas capacidades de desenvolvimento de software.

4.3 Trabalho futuro

Um dos aspetos que está sempre aberto a melhoria é a interface do utilizador para tornar a aplicação mais intuitiva, e também pode ser melhorada a parte estética e de layout. Algo que também poderia ser implementado seria a autenticação do utilizador

para haver controlo de acesso, no entanto, não foi uma prioridade por não ser um recurso utilizado no próprio site da DGES.

A arquitetura implementada permite a evolução do projeto em relação à expansão de funcionalidades. Até agora, o projeto foca-se em exibir gráficos de acordo com o número de candidatos ou colocados, mas seria interessante poder ser usado outro critério e realizar diversas comparações de dados.

À medida que o projeto cresce e lida com conjuntos maiores de dados, a otimização do desempenho torna-se crucial. É possível explorar técnicas como o uso de índices nas consultas à base e dados e apurar a estrutura da mesma para melhorar a eficiência e a velocidade de resposta da aplicação.

Outro ponto a que o grupo não deu prioridade foi o desenvolvimento de testes automatizados. Apesar da solução ter sido validada, esta prática é recomendada para garantir a estabilidade e qualidade do software.

5 Referências

Referência 1

(2023) *Python web scraping tutorial*. Acedido em: 10, março, 2023, em: <https://www.geeksforgeeks.org/python-web-scraping-tutorial/>

Referência 2

(2023) *A practical introduction to web scraping in Python, Real Python*. Acedido em: 18, março, 2023, em: <https://realpython.com/python-web-scraping-practical-introduction/>

Referência 3

Marica, S. (2021) *Python web scraping tutorial – how to scrape data from any website with python*. Acedido em: 23, março, 2023, em <https://www.freecodecamp.org/news/how-to-scrape-websites-with-python-2/>