

MPEI 2019-2020

Trabalho Prático

Objectivo Geral

- Desenvolver
- Testar e
- Demonstrar

uma aplicação que demonstre a utilização conjunta de:

- Deteção de itens similares
- Deteção de pertença a um conjunto

Como vão conseguir isto ?

- Dividindo em partes
 - Como em muitas outras situações
- Que partes ?
 1. Desenvolver os componentes/ módulos
 2. Testar os módulos
 3. Criar demonstração de uso conjunto
 - E testá-la
 4. Demonstrar aos Professores das Práticas

Desenvolver os módulos

- Desenvolver pelo menos **2 módulos**:
 - **Bloom Filter**
 - Processo(s) para descobrir **itens similares baseado em MinHash**
- Alguns dos guiões práticos que terão para resolver em novembro servem de base para a compreensão dos algoritmos
 - O primeiro relativo a Bloom Filters

Testar os módulos

- Para além da criação dos módulos **terão de criar e aplicar um conjunto de testes adequados para cada um dos módulos**
 - Testes de validação e demonstração da correção dos seus resultados
 - Testes de desempenho
 - Testes do efeito de parâmetros no desempenho
 - **O resumo dos testes e resultados deve integrar o relatório**
- A qualidade e quantidade dos testes efectuados será avaliada

Aplicação usando os 2 módulos

- Têm de criar uma aplicação fazendo uso dos 2 módulos para a resolução de um problema concreto
- O tema de cada grupo será atribuído na aula prática de entre uma lista de temas [Semana de 18 nov]
 - Numa primeira fase cada grupo indica o seu tema preferido
 - De seguida é feita a análise e aplicado o seguinte:
 - Todos os temas apenas com um grupo interessado ficam atribuídos
 - Aos grupos que tenham “colisões” com outros é dada a oportunidade de escolherem temas ainda sem grupo
 - Nos casos em que subsistam problemas e os grupos não cheguem a acordo, será efectuado um sorteio pelo docente

Criar demonstração de uso conjunto

- Nas TP e nas práticas podem e devem “discutir” ideias
- Nas TPs haverá tempo no final das aulas para conversar sobre este assunto desde que o solicitem
 - Por vezes podem apenas ter resposta na aula seguinte ...

Demonstrar aos Professores das Práticas

- Na **última semana dedicada a este trabalho**, nas práticas, terão de mostrar **o que fizeram** e responder às questões do docente
- Entrega final será alguns dias antes da apresentação do trabalho para que os docentes possam preparar as questões
 - Terão de entregar o código e uma explicação de como o usar (mini relatório)
 - em especial para correr os testes
 - Até ao final do dia de **Sexta 6 de Dezembro 2019**

Papel das aulas TP e Prática

- Nas TPs serão:
 - apresentados os conceitos base;
 - mostradas utilizações;
 - dadas dicas de como implementar partes dos módulos;
 - discutidas as vossas ideias, questões e problemas
- Nas práticas irão ter ajuda na resolução de dois guiões que garantem uma parte importantes dos pontos 1 e 2 (criar os módulos e testá-los)
 - Resumindo: essenciais para que tenham trabalho para entregar no final
- No entanto será preciso mais do que ir às TPs e Práticas
- Trabalho de casa

Versão provisória das cotações de cada parte

- Parte 1 - Módulos
 - Classificação até 8 (3 para Bloom, 5 MinHash)
 - Obrigatório ter os 2 módulos a funcionar para ter ≥ 6 nesta parte
 - Cuidado com as cópias: se detectadas serão penalizados, dividindo a nota pelo número de cópias
- Parte 2 - **Teste dos módulos**
 - Classificação até 2
 - Obrigatório ter testes de ambos os módulos
 - Bónus de até +20% para quem testar com dados relativos à aplicação conjunta

Versão provisória das cotações de cada parte (continuação)

- Parte 3 - Aplicação conjunta (Classificação até 5 valores)
- Parte 4- Criação e obtenção de dados para desenvolvimento e teste da aplicação conjunta (até 2 valores)
 - Será em particular tido em conta o tamanho dos datasets
- Parte 5 (apresentação + relatório)
 - Apresentação (1 valor) : recomenda-se Powerpoint (ou equivalente)
 - Demonstração “live” (1 valor)
 - Relatório (1 valor)
- Serão valorizadas soluções que empreguem variantes/evoluções dos algoritmos base (variantes do Bloom filter, LSH).
- classificações ≥ 18 poderão exigir uma apresentação extra

Algumas regras

- Trabalho de grupo
 - Máximo de 2 alunos
 - Da mesma turma prática
 - Terão de comunicar os grupos na primeira aula prática após a divulgação do trabalho
- O não funcionamento de um grupo tem de ser comunicado por email ao docente da prática (com CC para o colega de grupo e para o Regente ajst@ua.pt) até ao final de dia 23 de novembro
 - O email tem de justificar o não funcionamento do grupo
 - Ambos os elementos passam a fazer o trabalho individualmente

Regras (parte 2)

- Linguagem de programação:
 - Java
(eventuais exceções apenas para alunos ERASMUS e casos especiais)
- Datas limite para o trabalho
 - Apresentação na 2ª aula prática de dezembro
(semana de 9 de dezembro)
 - Entrega até sexta da semana anterior (6 dezembro)

Regras (parte 3)

- Submissões
 - Obrigatória submissão do estado do trabalho
 - Cada não submissão terá a penalização de 1 valor na nota final do trabalho (descontado(s) após a avaliação)
 - Submissão no Elearning
 - Submissões até ao final dos dias 22/11, 29/11 e 6/12

Documentos a entregar

- Programas (separados e claramente identificados)
 - Classe(s) com os módulos desenvolvidos
 - Testes dos módulos
 - Demontração conjunta dos vários algoritmos
- Relatório
 - Máximo 5 páginas
 - Deve incluir:
 - Descrição de como correr os vários programas (testes, demonstração conjunta, etc)
 - Apresentação e análise dos resultados obtidos nos vários testes
 - Descrição da aplicação de uso conjunto bem como das vantagens e limitações das soluções propostas
 - Descrição dos dados de teste da aplicação
- Apresentação
 - se criarem slides para a apresentação devem submeter o respetivo documento (não é obrigatório fazer um conjunto de slides)

Lista de problemas

1 – Filtro de submissão de trabalhos

- Sistema tipo Git que impeça a submissão de ficheiros com nomes já existentes e que alerte o responsável do repositório de que foi submetido um ficheiro com conteúdo similar a um já existente.

2 - Gestão de comentários online

- Ajuda à gestão de comentários (ex: no Tripadvisor)
 - Procura de comentários similares
 - Conjunto similar de locais comentados por um utilizador
 - Existência de um comentário de um utilizador para um determinado local

3 – Apoio à escrita

- Sistema que permita a um aluno que está a escrever uma Dissertação saber se uma frase (ou parte inicial) que escreveu já existe ou não num conjunto de documentos e obter uma lista de frases similares nesses documentos
- As frases que escreve devem ser adicionadas e consideradas em novas procuras

4 – Gestão de listas de compras

- Por exemplo de supermercado...
- Capacidade de determinar se uma lista com esses mesmos items já existe na aplicação
- Capacidade de descobrir listas similares a uma que tenhamos criado
- Novas listas devem ir sendo adicionadas

5 - Catálogo Criminal

- Base de dados com dados de criminosos
- A base de dados de cada criminoso deve incluir pelo menos crimes cometidos e características físicas
- A aplicação deverá ter as seguintes funcionalidades mínimas:
 - 1: Determinar a existência de um nome na lista de criminosos
 - 2: Determinar criminosos parecidos segundo um conjunto de características seleccionadas

6 – Hábitos de compra

- Encontrar utilizadores semelhantes em compras de produtos
 - Incluir inserção de produtos e utilizadores sem repetição
 - Capaz de identificar clientes com compras similares

7 - Gestão otimizada de percursos

- Sistema de ajuda na gestão dos trajetos /percursos (ex: Google Maps) por forma a que:
 - Apenas se guardem novos trajetos se forem diferentes de todos os que temos guardados
 - Se possa solicitar ao sistema recomendar percurso similar de entre os que já temos guardados

8 – Processamento de reviews de jogos

- Por exemplo de GOG.com
- Funcionalidades mínimas:
 - Saber se um determinado jogo tem reviews
 - E quantas
 - Determinar reviews similares e utilizadores que os fizeram
 - Possível utilização em deteção “spam”

9 – Letras de músicas

- Capaz de responder a questões como:
 - Existem letras com uma determinada palavra ou sequência de palavras (k-gramas)
 - Qual a letra mais similar a uma determinada letra
 - Quais os pares de letras musicais com similaridade superior a um determinado valor

10 – Procura de notícias repetidas

- Sistema capaz de obter e processar notícias de um conjunto de jornais e sites de notícias online por forma a:
 - Devolver uma lista de notícias repetidas com títulos iguais
 - Devolver lista de notícias com títulos similares
 - Devolver pares de notícias com conteúdos similares

11 – Padrões climáticos

- Procurar sequências de medições (ex: temperatura) iguais
- Procurar dias com medições similares
 - Ex: considerar medições hora a hora
- Procurar locais com medições diária/mensal/annual similares

12 – Gestão de biblioteca

- Funcionalidades:
 - Ver se livro existe no acervo (pelo nome)
 - Ver se livro está requisitado
 - Livros com títulos similares
 - Livros com conteúdo similar

Têm a palavra...

- Questões ?
- Comentários ?
- Sugestões para melhorar ?