

MPEI 2017-2018

Trabalho Prático

Objectivo Geral

- Desenvolver
- Testar e
- Demonstrar

um programa que demonstre a utilização conjunta de algoritmos probabilísticos para determinar **pertença a um conjunto** (set membership) e determinação de **itens similares** (finding similar items)

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 componentes/ módulos

- Terão de desenvolver grosso modo **2 componentes:**
 - um **Bloom Filter**
 - Processo(s) para descobrir **itens similares**
 - Incluirá uso de **Minhash**
- **As bases dos módulos constituirão os guiões práticos que terão para resolver em Novembro**
 - O primeiro relativo a Bloom Filters

Testar os módulos

- Para além da criação dos módulos **terão de criar um conjunto de testes adequados para cada um dos módulos**
- Será pedido um ou mais testes “obrigatórios” nos guiões da prática, mas podem e devem pensar e implementar outros testes que usem conhecimentos da cadeira ...
 - Será valorizado esse esforço e espírito inventivo

Criar demonstração de uso conjunto

- Esta parte apela particularmente à vossa imaginação
- Exemplo (que não devem replicar por deixar de ser uma ideia nova):
 - Escolher um conjunto de alguns documentos e inserir as palavras que estes contêm num filtro Bloom
 - Procurar um conjunto de páginas web em que mais de metade das palavras não esteja na lista criada com os documentos
 - Processar a lista obtida para determinar as páginas similares

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
- Terão, também, de entregar o código e uma explicação de como o usar (mini relatório)
 - em especial para correr os testes
- Em princípio a entrega final terá de ser alguns dias antes da apresentação do trabalho para que os docentes possam preparar as questões

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
- Resumindo: serão importantes
- 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 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 10
 - Obrigatório ter os 2 módulos a funcionar para ter ≥ 7
 - Nota depende da qualidade da implementação
 - Cuidado com as cópias: se detectadas serão penalizados, dividindo a nota pelo número de cópias
- Parte 2 (teste dos módulos)
 - Classificação até 5
 - Obrigatório ter testes dos dois módulos para ter ≥ 3

...

- Parte 3 (aplicação conjunta)
 - Classificação até 4 valores
- Parte 4 (apresentação)
 - Classificação até 1 valor
- Serão valorizadas soluções que empreguem variantes dos algoritmos base (variantes do Bloom filter, etc).
- Em princípio classificações ≥ 17 obrigarão a participação numa pequena “competição” em que os algoritmos desenvolvidos terão de ser adaptadas a um determinado problema.

Algumas regras

- Trabalho de grupo
 - Máximo de 2 alunos
 - Da mesma turma prática
 - Terão de comunicar os grupos na primeira aula prática sobre Bloom Filters
- Solicitaremos submissão no elearning do estado do trabalho ao longo das semanas

Regras (parte 2)

- Linguagem de programação:
 - Java
 - Matlab (em casos devidamente justificados e se aceites pelo docente das práticas)
- Datas limite para o trabalho
 - Apresentação na 1ª aula prática de dezembro
 - Entrega até 30 de novembro

Têm a palavra...

- Questões ?
- Comentários ?
- Sugestões para melhorar ?