

Instruções para o trabalho da disciplina

1. O trabalho deverá ser desenvolvido em dupla ou individual.
2. O trabalho envolve o desenvolvimento de Web Scraping focado, que deverá fazer a coleta automática de dados associados a um tema, a uma determinada estrutura, a metadados, ou até mesmo dados de um determinado tipo de arquivo específico.
3. Os estudantes deverão definir este ponto em específico **em concordância com a professora** - por sugestão própria ou sugerido pela professora. Dependendo do que for escolhido:
 - Associado a um tema, a uma estrutura ou a métodos: definir o tipo de arquivo, e as URLs para coleta dos dados;
 - Por tipo de arquivo específico: desenvolver um crawler para procurar na Web os dados do tipo definido.
4. O número de URLs depende do tema escolhido e do número de integrantes do grupo.
5. Ao final do semestre, os estudantes deverão gerar um JSON com uma média de 1.000 instâncias. O JSON deve prever, no mínimo, 5 atributos, sendo que um deles deverá guardar a URL de onde os dados foram extraídos.
6. Poderão ser usadas quaisquer APIs prontas ou linguagens de programação.
7. Até a data limite para definição do trabalho, os estudantes deverão responder ao formulário passado, com:
 - a. um texto, descrevendo o que definiram como dados a coletar;
 - b. um exemplo de aplicativo que iriam utilizar os dados extraídos;
 - c. API, bibliotecas e linguagem que pretendem utilizar.
8. Estudantes que não tem planos, devem entrar em contato com a prof. o quanto antes.
9. Os temas definidos serão analisados pela professora, para que não haja repetição. No caso de haver mais de um trabalho com o mesmo tema, ou de haver repetições de URLs já utilizadas em semestres anteriores, o tema deverá ser alterado.
10. Entrega do trabalho:
 - a. A data limite para finalização do trabalho está descrita no Moodle. Alunos que finalizarem antes podem submeter o trabalho antes.
 - b. Entregar:
 - i. **vídeo** de apresentação, utilizando o *template* fornecido no Moodle;
 - ii. **dataset** extraído - colocar em qualquer repositório público (por ex, GitHub)