



Instruções para o trabalho da disciplina

1. O trabalho deverá ser desenvolvido em dupla ou trio (trabalhos individuais serão analisados).
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 concordância com a professora** - por sugestão própria ou sugerido pela professora.
4. O número de URLs depende do tema escolhido e do número de integrantes do grupo, mas para temas complexos, deve ser no mínimo 3.
5. Ao final do semestre, os estudantes deverão gerar um JSON com uma média de 1.000 instâncias. O JSON deve prever, no mínimo, 11 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; consultas a chats como ChatGPT são permitidas.
7. Não é permitido o uso de APIs que entregam os dados prontos, visto que o objetivo do trabalho é o desenvolvimento de um scraping. Portanto, o que vale é a raspagem de dados em sites que não entregam os dados prontos. Em alguns casos, isso pode ser revisto se o grupo fizer um misto entre o uso de APIs prontas e scraping de páginas.
8. O grupo deverá pensar na utilidade dos dados coletados, portanto, no momento da definição de seu trabalho, deverão informar um exemplo de aplicativo que iriam utilizar os dados extraídos; esse aplicativo deverá ser simulado e alguns prints de tela deverão ser mostrados no vídeo de apresentação do trabalho.
9. 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)
 - iii. **codigo** desenvolvido - colocar em qualquer repositório público (por ex, GitHub)