

Neste trabalho final, da unidade curricular de análise e visualização de dados criou-se um script de python que para todas as obras fornecidas em formato Markdown:

- Analisa a obra usando NLP
- Extrai vários elementos como ‘pessoas’, ‘locais’ e ‘datas’
- Extrai os elementos de lematização como ‘Advérbio’, ‘verbo’ e ‘Interjeição’
- Analisa a obra por capítulo usando um modelo sentimentalizador
- Extrai os elementos de sentimento ‘negativo’, ‘neutro’, ‘positivo’ e ‘composto’
- Cria três datasets isolados organizados com as informações extraídas
- Cria e exporta gráficos para cada um dos datasets

Algumas das *features* interessantes do projeto são:

- Scripts inteiramente automáticos não sendo preciso intervenção humana em nenhuma das diferentes fases, o que permite automatização total.
- Os scripts automaticamente limpam os dados e corrigem erros que possa surgir, contanto não só com extração seletiva dos dados a partir do fornecimentos dos dados, como também por uma limpeza afim de impedir erros na base de dados.
- Organização dos datasets e gráficos numa estrutura hierárquica e clara
- Possibilidade de ajustar facilmente a dimensão do dataset no script via variável de extração
- Possibilidade de reescrever/regravar o output completo do script, permitindo alterações ao código “on the fly”, sem ter de apagar manualmente o output completo.
- Filtragem automática de coocorrência de nomes de personagens quando providenciada uma lista com os nomes de personagens principais
- Possibilidade de correr os 3 scripts encadeados, ou apenas os scripts dos gráficos, o que permite ajustar apenas os gráficos
- Criação de gráficos que se auto-ajustam aos datasets fornecidos
- Gráficos criados com um aspeto visual limpo e apelativo no formato SVG (Vector Graphics) e EPS (Imagem com camadas editáveis)

A utilização deste projeto é muito simples: Colocar as obras para análise (previamente formatadas em Markdown) na pasta obras; se pretender, é possível editar o script para extrair mais ou menos elementos; executar o script principal (Main\_Script.py) e esperar que seja criada a árvore do output:

