Projeto 1

Introdução a Ciência de Dados para Biblioteconomia

Utilizando a base de dados da Web of Science, responda as seguintes perguntas:

1. Faça uma pesquisa sobre o tema “data science” nos últimos 10 anos;
2. Baixe os dados completos do resultado para serem tratados no OpenRefine;
3. Realize os tratamentos de dados de limpeza e normalização que você aprendeu no curso. Anote cada um dos procedimentos que você realizou passo-a-passo.
4. Para esses dados, responda, utilizando gráficos de uma planilha de cálculo:
   1. Quais são os 10 autores mais produtivos?
   2. Os nomes dos autores estão escritos de forma contínua ou separados em nome e sobrenome? Se estiverem separados, junte os nomes todos de forma corrente.
   3. Quais são as 10 universidades mais produtivas?
   4. Quais são os 10 países mais produtivos?
   5. Quais são os 10 artigos mais citados?
   6. Quais são as 10 áreas do conhecimento mais produtivas no tema?
   7. O tamanho médio de caracteres dos resumos?
   8. Utilizando fórmulas condicionais e de tratamento de texto, identifique os 10 países mais produtivos e substitua pelos nomes correntes em língua portuguesa;
5. Entregue a resposta final em formato de relatório em texto com uma resenha dos resultados encontrados.