

INF-0614 – VISUALIZAÇÃO DE INFORMAÇÃO

PROF. DR. CELMAR GUIMARÃES DA SILVA

celmar@unicamp.br

Trabalho 3 - 20/04/2024

Objetivo: Usar visualizações baseadas em mapas de calor (*heatmaps*) para analisar dados multivariados, e visualizações de textos baseadas em projeção multidimensional. Usar conhecimentos da disciplina para visualizar o conjunto de dados fornecido.

Objetos do estudo: (a) Conjunto de dados “wine-normalizado-versao-esparsa-para-heatmap-Tableau” (anexado ao trabalho). (b) Bíblia em versão txt (anexada ao trabalho). (c) Conjunto de dados “Sleep study” (anexado ao trabalho).

Tarefas: Faça as seguintes tarefas, em duplas ou em trios:

1. Analise o conjunto de dados “wine-normalizado-versao-esparsa-para-heatmap-Tableau” utilizando mapas de calor. Crie um mapa de calor e analise quais características das variáveis do conjunto diferenciam as três classes de vinhos. Sugere-se usar o software Tableau para esta análise, configurando apropriadamente filtros e legendas de cor, e trabalhando com reordenações de linhas e de colunas. A entrega deve conter:
 - (a) Uma ou mais imagens de mapas de calor desse conjunto de dados. O mapa de calor deve ter no eixo X as características disponíveis na variável “features”, e no eixo Y os vinhos em si.
 - (b) Um parágrafo com a análise solicitada, com base nas imagens mostradas.
2. O exemplo de visualização de textos dado em sala usou projeção multidimensional para comparar a semelhança entre 4 documentos (4 livros da Bíblia) conforme lista de palavras usadas para criar a matriz de distância usada para a projeção. Com base na versão da Bíblia em txt (em anexo), amplie o exemplo para considerar mais livros (pelo menos mais 4) e gere visualização de projeção dimensional (via MDS) dos 6 livros (os 4 anteriores mais os quatro a serem adicionados). Sugere-se usar o software Orange. Proceda da seguinte forma:
 - (a) Preparação da estrutura de dados:
 - i. Escolha mais 4 livros (ou mais) da Bíblia;
 - ii. Faça a lista de palavras de cada livro (pode ser processamento manual via editor simples de texto);
 - iii. Retire pontuações diversas das listas: () . , ; “ ” ? !
 - iv. Altere a planilha em anexo (que calcula a matriz de tf-idf) para que ela considere também os novos livros, adicionando a ela as listas de palavras geradas;
 - v. Altere as palavras usadas para o cálculo do tf-idf, se julgar necessário;
 - vi. Copie a matriz de tf-idf calculada na planilha para um arquivo csv;
 - (b) Visualização: Abra essa matriz no Orange e crie visualização do tipo MDS. Pode gerar mais de uma imagem se for necessário à análise;
 - (c) Entrega: Entregue a planilha, as imagens geradas, e um breve parágrafo sobre sua análise da imagem gerada.
3. Analise o conjunto de dados “Sleep study” usando técnicas de visualização. Apresente visualizações que tragam informações relevantes sobre esse conjunto. O grupo é livre para escolher técnicas que julgue apropriadas para esta análise, desde que apropriadamente justificadas. As informações apresentadas devem ser apoiadas por cópias de tela das visualizações, mapeamento visual usado, e justificativa sobre como a visualização permite observar essas informações.

Observações:

- Para a Tarefa 2, o arquivo “dados sobre biblia-v2020-10-03.zip” (em anexo) contém a Bíblia em txt, arquivos texto de 4 livros (Gênesis, Êxodo, Mateus e Marcos), arquivos texto com palavras desses livros separadas por linha (já preprocessadas para tirar pontuação), planilha com cálculo de frequência e de tf-idf das palavras, e tabelas em arquivos csv com trechos dessa planilha (prontos para uso no Orange).
- Deve ser entregue 1 arquivo PDF, que deve conter em seu cabeçalho o nome dos componentes do grupo.

Entrega: Via Moodle, até 28/04/2024 (domingo).

Correção: Até 05/05/2024 (domingo).