# Projeto 7 Snemp - ferramenta para identificação de disparidades em compras públicas

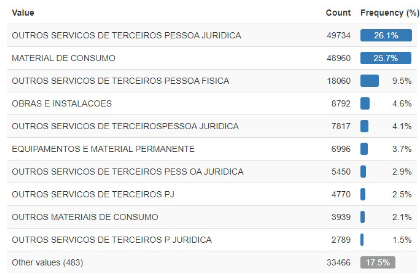
Uma nota de empenho é um documento utilizado para registar as despesas orçamentárias realizadas pela Administração Pública em seu primeiro estágio. Esse documento identifica, dentre outras informações, o nome do credor, a especificação, a importância da despesa e a célula orçamentária, deduzindo o saldo da dotação aprovada.

O volume de dados gerado diariamente na forma de notas de empenho pode ser alvo para análises. Em particular, uma aplicação relevante nesse conjunto de dados diz respeito a identificar disparidades entre preços praticados para itens (produtos ou serviços) similares em diferentes notas de empenho. Um complicador nesse contexto é que a tarefa de identificar quais itens são similares em um conjunto de notas de empenho não é trivial, uma vez que a várias informações que descrevem um item de nota de empenho são registradas como texto livre.

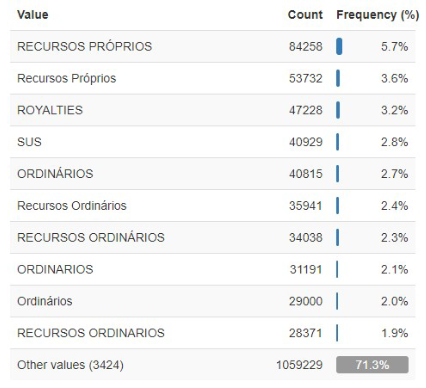
Para facilitar o entendimento do problema, considere as duas tabelas a seguir. As duas tabelas apresentam informação agregada relativa a uma das colunas da base de dados de notas de empenho. Essa coluna é denominada ElemDespesaUG[[1]](#footnote-0).

* A primeira tabela apresenta informação agregada a partir da base de dados original. Repare, por exemplo, que a primeira e terceira entradas nesta tabela apresentam valor na primeira coluna “MATERIAL DE CONSUMO” e “Material de Consumo”, respectivamente. Essas duas entradas represente o mesmo elemento da despesa, mas estão descritos de forma ligeiramente diferente. Para citar outro exemplo, o mesmo acontece com a quinta e a sexta entradas.
* A segunda tabela apresenta informação agregada após terem sido realizadas algumas transformações de normalização (e.g., tornar o texto maiúsculo, eliminar a acentuação etc.) sobre o conteúdo da coluna ElemDespesaUG. Repare que, após a transformação, a entrada com valor normalizado “MATERIAL DE CONSUMO” corresponde a 25.7% de todos os registros da base de dados.





Como outro exemplo, considere a tabela a seguir, que representa informação agregada sobre outra coluna da base de dados de notas de empenho. O nome dessa coluna é DEFonteUG. Essa tabela exibe pelo menos 06 (seis) formas de especificar o valor “recursos ordinários”. Essas variações textuais (acentuação, disposição da fonte entre maiúsculas e minúsculas etc) dificultam a tarefa de encontrar itens de notas de empenho semelhantes.



O objetivo do presente projeto é desenvolver a Snemp, uma ferramenta para identificação disparidades em preços praticados em compras públicas registradas como notas de empenho.

**Requisitos funcionais**

1. A ferramenta deve permitir ao usuário fornecer as características de um item (produto ou serviço) para que a identificação de potencial disparidade de preços seja disparada. Como resposta, a ferramenta deve apresentar uma lista com informações de todos os itens de nota de empenho que sejam similares ao item caracterizado pelo usuário. Para cada item encontrado, a ferramenta deve também apresentar seu preço praticado.
2. Com relação ao resultado descrito no RF1, a ferramenta deve permitir filtrar os resultados pelos seguintes campos: data do empenho (data inicial e data final), unidade, credor.
3. Ainda com relação ao resultado descrito no RF1, a ferramenta deve apresentar resumos estatísticos (média, moda, variância, desvio padrão, mínimo, máximo, quantidade de nfes, quantidades e média ponderada) relacionados ao valor praticado para o conjunto de itens similares, com a possibilidade de geração de gráficos (e.g., box-plot).
4. A ferramenta deve permitir importar um conjunto de notas de empenho a partir um arquivo CSV (cujo formato está para ser especificado).
5. A ferramenta deve permitir exportar o resultado da análise em formato PDF e Excel (cujo formato está para ser especificado).

**Requisitos não funcionais**

1. A ferramenta deve ser implementada como uma aplicação Web.
2. A ferramenta a ser desenvolvida deve trabalhar sobre uma base de dados de notas de empenho, a ser fornecida para o grupo de desenvolvimento.
3. Já existe um módulo de identificação de produtos iguais (mas com descrições distintas). Esse módulo foi implementado em linguagem Python, no contexto da ferramenta Snelf. Esse módulo deve ser aproveitado durante o desenvolvimento da Snemp, para implementação do módulo de identificação de itens similares em notas de empenho.
4. A ferramenta deve armazenar os dados de notas de empenho importados de um arquivo em formato CSV para um banco de dados relacional.
5. Além do código fonte da aplicação, também deverão ser incluídos como artefatos entregáveis:
   1. Documento de instalação e configuração da ferramenta.
   2. Diagrama da arquitetura (tecnologias)
   3. Diagramas de classes
   4. Testes unitários para cada parte relevante do código do código-fonte.
   5. Testes funcionais para partes relevantes com selenium IDE ou selenium WebDriver.
6. O desenvolvimento da aplicação deve, desde o início, ocorrer de forma evolutiva e versionada no github.
7. O repositório github deve ser compartilhado para que seja possível acompanhar a evolução do desenvolvimento.
8. O artefato *(a)* deve ser entregue na forma do arquivo README.md do repositório no github.
9. Os artefatos *(b)* e *(c)* devem ser entregues por meio de um diretório denominado *docs* na raiz do repositório do github.

**Informações adicionais**

* Projeto de mineração de dados em Notas Fiscais Eletrônicas (NFE): [repositório github](https://github.com/MLRG-CEFET-RJ/e-gov-data-mining).
* Relatório de iniciação científica do Augusto Fonseca. Arquivo de nome “IC\_2019\_AugustoFonseca.pdf” disponível na seção “Arquivos” do time da disciplina Projeto e Construção de Sistemas (no Microsoft Teams).
* Para fins de inspiração para a equipe de desenvolvimento já há uma ferramenta similar utilizada pelo estado da paraíba nos seguintes endereços: <https://precodahora.pb.gov.br/produtos/> e <https://precodereferencia.tce.pb.gov.br/>

1. Essa coluna tem por finalidade identificar os objetos de gasto, tais como vencimentos e vantagens fixas, juros, diárias, material de consumo, serviços de terceiros prestados sob qualquer forma, subvenções sociais, obras e instalações, equipamentos e material permanente, auxílios, amortização e outros que a Administração Pública utiliza para a consecução de seus fins.  
   Fonte: https://antigo.plataformamaisbrasil.gov.br/ajuda/glossario. [↑](#footnote-ref-0)