

# Projeto 7

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

Uma nota de empenho é um documento utilizado para registrar 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<sup>1</sup>.

- 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.

---

<sup>1</sup> 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>.

Value	Count	Frequency (%)
MATERIAL DE CONSUMO	162951	11.0%
OUTROS SERVIÇOS DE TERCEIROS - PESSOA JURIDICA	81262	5.5%
Material de Consumo	75073	5.1%
Outros Serviços de Terceiros - Pessoa Juridica	69050	4.7%
VENCIMENTOS E VANTAGENS FIXAS - PESSOAL CIVIL	49258	3.3%
Vencimentos e Vantagens Fixas - Pessoal Civil	37672	2.5%
OUTROS SERVICOS DE TERCEIROS - PESSOA FISICA	37575	2.5%
Outros Serviços de Terceiros - Pessoa Fisica	35096	2.4%
OUTROS SERVIÇOS DE TERCEIROS - PESSOA FÍSICA	33465	2.3%
DIÁRIAS - CIVIL	26262	1.8%
Other values (2594)	876268	59.0%

Value	Count	Frequency (%)
OUTROS SERVICOS DE TERCEIROS PESSOA JURIDICA	49734	26.1%
MATERIAL DE CONSUMO	48960	25.7%
OUTROS SERVICOS DE TERCEIROS PESSOA FISICA	18060	9.5%
OBRAS E INSTALACOES	6792	4.6%
OUTROS SERVICOS DE TERCEIROSPESSOA JURIDICA	7617	4.1%
EQUIPAMENTOS E MATERIAL PERMANENTE	6996	3.7%
OUTROS SERVICOS DE TERCEIROS PESS OA JURIDICA	5450	2.9%
OUTROS SERVICOS DE TERCEIROS PJ	4770	2.5%
OUTROS MATERIAIS DE CONSUMO	3939	2.1%
OUTROS SERVICOS DE TERCEIROS P JURIDICA	2789	1.5%
Other values (483)	33466	17.5%

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.

Value	Count	Frequency (%)
RECURSOS PRÓPRIOS	84258	5.7%
Recursos Próprios	53732	3.6%
ROYALTIES	47228	3.2%
SUS	40929	2.8%
ORDINÁRIOS	40815	2.7%
Recursos Ordinários	35941	2.4%
RECURSOS ORDINÁRIOS	34038	2.3%
ORDINARIOS	31191	2.1%
Ordinários	29000	2.0%
RECURSOS ORDINARIOS	28371	1.9%
Other values (3424)	1059229	71.3%

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

- RF1. 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.
- RF2. 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.
- RF3. 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).
- RF4. A ferramenta deve permitir importar um conjunto de notas de empenho a partir um arquivo CSV (cujo formato está para ser especificado).
- RF5. A ferramenta deve permitir exportar o resultado da análise em formato PDF e Excel (cujo formato está para ser especificado).

### Requisitos não funcionais

- RNF1. A ferramenta deve ser implementada como uma aplicação Web.
- RNF2. A ferramenta a ser desenvolvida deve trabalhar sobre uma base de dados de notas de empenho, a ser fornecida para o grupo de desenvolvimento.
- RNF3. 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.
- RNF4. A ferramenta deve armazenar os dados de notas de empenho importados de um arquivo em formato CSV para um banco de dados relacional.
- RNF5. Além do código fonte da aplicação, também deverão ser incluídos como artefatos entregáveis:
- Documento de instalação e configuração da ferramenta.
  - Diagrama da arquitetura (tecnologias)
  - Diagramas de classes
  - Testes unitários para cada parte relevante do código do código-fonte.
  - Testes funcionais para partes relevantes com selenium IDE ou selenium WebDriver.
- RNF6. O desenvolvimento da aplicação deve, desde o início, ocorrer de forma evolutiva e versionada no github.
- RNF7. O repositório github deve ser compartilhado para que seja possível acompanhar a evolução do desenvolvimento.
- RNF8. O artefato (*a*) deve ser entregue na forma do arquivo README.md do repositório no github.
- RNF9. 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](#).
- 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/>