

## RELATÓRIO DE INFORMAÇÕES TÉCNICAS (RIT)

### 1. MATERIAL AVALIADO

Campo textual com as informações de evolução recebidas do software C2Cmb no banco de dados PostgreSQL.

### 2. EMPREGO DO MATERIAL

O banco de dados PostgreSQL é utilizado para armazenar informações recebidas do software C2Cmb. Nele há um campo textual que lista a evolução das operações. Essas informações são repassadas em texto corrido e por vezes, seguem um padrão de palavras chaves.

### 3. ASPECTOS AVALIADOS

#### a. CONTEXTUALIZAÇÃO DO PROBLEMA

O problema previamente relatado foi a impossibilidade de coletar dados de maneira rápida e fácil dos campos descritivos das evoluções das operações presentes no C2Cmb. O preenchimento dos campos, pelos operadores, ocorre de maneira livre e pouco estruturada no lugar de um formulário que salve as informações em tabela permitindo filtros, pesquisas diversas e fácil visualização gráfica.

#### b. EVIDÊNCIAS ENCONTRADAS

Os campos avaliados encontram-se presentes na tabela *sad.pln\_ordr\_progress*, na coluna *progress\_detail* da operação com *pln\_ordr\_id*=10007200000000003098, a qual contém diversas informações numéricas e textuais acerca da evolução da operação, em algumas operações, o texto encontra-se de maneira direta as seguintes informações: quem, onde, quando, imigrantes, cubanos, haitianos, venezuelanos, outros, residentes permanentes, turistas, residentes temporários, refugiados, destino Boa Vista, destino outras cidades, destino outros países, prisões, carros, caminhões, ônibus, motos, vans, outros veículos, apreensões, ocorrências.

### c. IMPLEMENTAÇÃO DA SOLUÇÃO

Primeiramente a equipe avaliou empregar ferramentas de processamento de linguagem natural (principalmente através da Análise de Dependência entre as palavras), especificamente bibliotecas *python* spaCy, NLTK e udpipe. Após diversos testes, foi verificado que os modelos da língua portuguesa utilizados pelas ferramentas não produzia resultados a contento devido ao formato das informações e a incapacidade de associar corretamente determinados itens a suas quantidades (Ex: “substâncias entorpecentes”). Isso se dá por conta do treinamento desses modelos. Uma solução é coletar informações militares suficientes para se treinar um modelo capaz de reconhecer as palavras recebidas pelas operações militares. Contudo, isso demanda tempo, e foge do escopo da ORF.

A solução cabível para a ORF é de se coletar informações da Operação Acolhida e estruturá-las de forma automática, tendo em vista que seguem um padrão de palavras chaves, no formato “Quem: Onde: Quando: Apreensão: ...”. Para isso foi desenvolvido um script em Python que pega as informações do banco de dados e estrutura em um arquivo .csv, gerado como saída. As informações foram recuperadas usando reconhecimento de expressões regulares e tratamento de dados. O arquivo .csv pode ser inserido em ferramentas de visualização como o *kibana* para gerar gráficos e estatística a respeito das operações. O manual do operador segue em anexo junto ao *script*.

## 4. SUGESTÕES

- a. Reunir dados militares manualmente anotados nos moldes do Processamento de Linguagem Natural a fim de que se treine um modelo capaz de ser usado futuramente na estruturação automática de textos corridos advindos das operações militares.
- b. Como o processo sugerido anteriormente demanda tempo e pesquisa, o problema atual pode ser solucionado adotando uma estrutura de informações para todas as operações. Uma padronização possível seria: Quem: , Onde: , Quando: , Imigrantes: , Cubanos: , Haitianos: , Venezuelanos: , Outros: , Resperm: , Turistas: , Restemp: , Refugiados: , Destino: Boa Vista: , Outras Cidades: , Outros Países: , Prisões: , Carros: , Caminhões: , Ônibus: , Motos: , Vans: , Outros Veículos: , Apreensões: , Ocorrências: , Informações Adicionais: .

## 5. CONCLUSÃO

Foi observado devido a falta de estruturação do campo de evolução do *software* C2Cmb, muitos dados tem o acesso dificultado, necessitando o emprego de esforço humano para recuperá-los. Foi feito um estudo preliminar do emprego de ferramentas de processamento de linguagem natural, porém não existem modelos adequados na língua portuguesa para o problema. Foi feita uma redução do escopo para solucionar o problema específico da Op Acolhida.

Como oportunidades de melhorias foram levantadas a padronização do preenchimento do campo evolução, a própria estruturação do campo por meio do C2Cmb e um possível estudo usando o resultado estruturado para treinamento de um *software* de aprendizado de máquina

para gerar um modelo de estruturação do texto, que pode ser concebido como um tema de Projeto de Fim de Curso de anos futuros.

Manaus-AM, 14 de novembro de 2019

---

1º Ten QEM – ROBERTO TADEU ABRANTES DE ARAÚJO  
Engenheiro de Computação