

## 6 Conclusões

A melhor relação entre acurácia e desempenho foi identificada nos testes do algoritmo *K-Means* com a modelagem utilizando trigramas sem nenhuma dimensão extra ou modificação nos pesos dos caracteres de acordo com sua posição relativa à palavra.

Curiosamente, no início do planejamento e pelos testes preliminares o algoritmo *K-Means* não havia se destacado com nenhuma modelagem, ou melhor, havia sido menos preciso e mais lento que os outros conjuntos algoritmo-modelagem testados.

Além de contradizer as expectativas iniciais em relação ao algoritmo utilizado, outro ponto é que a modelagem mais simples, apenas utilizando trigramas como dimensões, foi a que mostrou o melhor resultado. Novamente havia indícios nos testes preliminares e o estudo preparatório também sugeria que a aplicação de algoritmos fonéticos ou ainda a inclusão de mais dimensões que permitissem uma melhor caracterização do registro aumentaria a acurácia e, em muitos casos, aceleraria a convergência do algoritmo a um resultado satisfatório.

Isso mostra que apesar do domínio estudado se assemelhar a outros estudados anteriormente e encontrados na literatura, seu comportamento frente aos algoritmos não pôde ser inferido a partir dos estudos prévios. Entretanto, o resultado obtido é absolutamente satisfatório atendendo o objetivo traçado de reduzir o tempo de processamento significativamente mantendo a qualidade do resultado final do processamento.

Os resultados indicam que problemas cujo volume de dados seja grande podem recorrer a algoritmos e modelagens semelhantes para acelerar o processo sem perda significativa na qualidade, ou ainda dependendo do cenário, sem apresentar nenhuma perda de qualidade. Esse tipo de mecanismo representa um ganho qualitativo para a arquitetura e para a execução de algoritmos que trabalhem com grandes massas de texto.