Conclusão

Ao longo do trabalho demostramos que é possível realizar a classificação dos documentos de patentes a múltiplos tópicos baseado em um dicionario gerado automaticamente a partir do conjunto de documentos de patentes estudado ou que poderá ser baseado em um dicionario construído por um especialista em analise de documentos de patentes. Verificamos que o pre processamento aplicado no tratamento do texto foi satisfatório e o modelo RandomForest teve melhor adequação ao nosso tipo de problema.

Em trabalhos futuros, proponho a melhora da construção de dicionários automaticamente e avaliar se modelos de rede neural teria um melhor resultado de acurácia em relação ao RandomForest.