Discussão parcial

Conseguimos realizar a extração bem sucedida de uma amostra de documentos de patentes, do qual pré processamos e criamos um corpora que poderá ser usado não somente para este trabalho como para outros trabalhos com documentos de patentes. Fizemos um primeiro levantamento dos tópicos usando o modelo LDA e obtivemos 20 tópicos que serão avaliados e rotulados corretamente.

Utilizando termos simples para rotular os tópicos neste primeiro momento, expandimos com termos que são sinônimos, hiperônimos e hipônimos dos termos associados aos tópicos.

Para os próximos passos, desenvolveremos um modelo para classificar os documentos de patentes a partir do dicionario criado.