Materiais e métodos

1 – Descrição do sujeito ou do objeto de estudo

Pesquisa de patentes nos sites Y e Z sobre o tema agricultura.

2 – Delineamento da pequisa

Totalizando N patentes, selecionamos uma amostragem de X patentes. Utilizamos um webcrawler para realizar o download dos arquivos (verificar se é webcrawler ou webscraping). Todo o conteudo foi armazenado no banco de dados.

3 – Procedimentos específicos

DESCREVER

Construção do dicionario

- construção de um corpora com o titulo e o abstract das patentes

- processar o corpora (manter em caixa baixa, remoção de caracteres especiais, remoção de numeros, remoção de stop-words)

- conversão em uma matriz de documento-termo (Document-Term Matrix – DTM)

- aplicar o modelo LDA

- utilizar o wordnet para expandir o numero de palavras relacionados aos tópicos com sinonimos

Construção do modelo

…

4 – Analise dos dados

Avaliar o resultado do dicionario, observando se o valor k usado para geração de tópicos tem relação ao corpora utilizado.