

IR A3 README FILE

Question 1:

Pre-processing steps used-

- All the text is converted to lower case
- All meta-data is removed from the text
- All email-ids are removed from the text
- Tokens which consist of letters as well as digits like play2 are removed from the text.
- All the punctuation marks are removed from the text
- All stop-words are removed from the text
- Lemmatization is applied on the text