Preprocessing

1. Load the data (located in the data folder)
2. Perform tokenization
3. Delete stop words
4. According to the data, remove any extra characters (for example: tags, punctuation, etc.) in the data and clean the data.
5. Perform stemming

Indexing

1. After collecting all the tokens, Build an inverted index

Note

- Code in any programming language is acceptable
- Each HW has to be done individually
- For assignment that are similar to each other, the grade is 0.
- deadline: 1402/1/27

Good luck, Fathollahi