1. Data Collected from 900 news articles (Sayem)???
2. Data labelled - can be put into four separate directories based on label (Sayem)???
3. Corpus read, tokenized using RegEx (Minhaz) – Done
4. Bengali Word Stemming (Minhaz) – Done
5. TF-IDF Score calculation (Minhaz) – pending
6. N-Gram model generation (Minhaz) – pending
7. What else can be done ??? (Farzana)
8. Presentation preparation (Minhaz) – pending
9. Report writing (Farzana) – pending

Rafi Kamal's Stemmer

~~~~~~~~~~~~~~~~~~~~

Originally Developed by `Rafi Kamal`\_. Ported to Python.::

from bengali\_stemmer.rafikamal2014 import RafiStemmer

stemmer = RafiStemmer()

stemmer.stem\_word('বাংলায়')

.. \_`Rafi Kamal`: https://github.com/rafi-kamal/Bangla-Stemmer

Mahmud's Stemmer

~~~~~~~~~~~~~~~~

Originally Implemented by M. R. Mahmud, M. Afrin, M. A. Razzaque, E. Miller and J. Iwashige.

Ported to Python [1] Under development. As of now, only verb stemming has been implemented.

``mahmud2014`` package under ``banglakit``.

----------

References

----------

[1] M. R. Mahmud, M. Afrin, M. A. Razzaque, E. Miller and J. Iwashige, "A rule based bengali stemmer," 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), New Delhi, 2014, pp. 2750-2756.

doi: 10.1109/ICACCI.2014.6968484