

Bengali POS Tagging & Word2Vec Project Resources

Research Papers/Documentations:

-Bengali POS Tagging

1. **A Comprehensive Parts of Speech Tagger for Automatically Checked Valid Bengali Sentences.** *2018 21st International Conference of Computer and Information Technology (ICCIT), 21-23 December, 2018.* Authors:
 - **Nahid Hossain**, Computer Science and Engineering, United International University Dhaka, Bangladesh. nahid@cse.uiu.ac.bd
 - **Mohammad Nurul Huda**, Computer Science and Engineering, United International University, Dhaka, Bangladesh, mnh@cse.uiu.ac.bd
2. **Development of POS tagger for English-Bengali Code-Mixed data.** Authors:
 - **Tathagata Raha**, IIT Hyderabad, tathagata.raha@research.iiit.ac.in
 - **Sainik Kumar Mahata**, Jadavpur University, Kolkata. sainik.mahata@gmail.com
 - **Dipankar Das**, Jadavpur University, Kolkata. dipankar.dipnil2005@gmail.com
 - **Sivaji Bandyopadhyay**, Jadavpur University, Kolkata. sivaji_cse_ju@yahoo.com
3. **BNLP: Natural language processing toolkit for Bengali.** Authors:
 - **Sagor Sarker**, Begum Rokeya University, Rangpur, Bangladesh, brursagor@gmail.com
4. **Deep Learning Based Parts of Speech Tagger for Bengali.** *2016 5th International Conference on Informatics, Electronics and Vision (ICIEV).* Authors:
 - **Md. Fasihul Kabir**, Department of Computer Science and Engineering, United International University, Dhaka, Bangladesh, aboutrafi@gmail.com
 - **Khandaker Abdullah-Al-Mamun**, Department of Computer Science and Engineering, United International University, Dhaka, Bangladesh, mamun@cse.uiu.ac.bd
 - **Mohammad Nurul Huda**, Department of Computer Science and Engineering, United International University, Dhaka, Bangladesh, mnh@cse.uiu.ac.bd
5. **Part-of-Speech Tagging for Bengali.** Department of Computer Science and Engineering, Indian Institute of Technology, Kharagpur. Authors:
 - **Sandipan Dandapat**,
 - **Prof. Sudeshna Sarkar**,
 - **Prof. Anupam Basu**,

6. **Implementation of Automated Bengali Parts of Speech Tagger: An Approach Using Deep Learning Algorithm.** 2020 IEEE Region 10 Symposium (TENSYP), 5-7 June 2020, Dhaka, Bangladesh. Authors:
 - **Asraf Hossain Patoary**, department of CSE, Dhaka International University Dhaka, Bangladesh, asrafhossain197@gmail.com
 - **Md. Jahid Bin Kibria**, department of CSE, University of Development alternative Dhaka, Bangladesh, jahidbinkibria@gmail.com
 - **Abdul Kaium**, department of CSE, Dhaka International University Dhaka, Bangladesh, kaium0603@gmail.com
7. **POS Tagging For Code-Mixed Indian Social Media Text : Systems from IIIT-H for ICON NLP Tools Contest.** Authors:
 - **Arnav Sharma**, LTRC, IIIT Hyderabad, arnav.s@research.iiit.ac.in
 - **raveesh.motlanig@research.iiit.ac.in**, LTRC, IIIT Hyderabad, raveesh.motlani@research.iiit.ac.in
8. **Towards POS Tagging Methods for Bengali Language: A Comparative Analysis.** Department of Computer Science & Engineering, Chittagong University of Engineering & Technology, Chittagong-4349 Bangladesh. Authors:
 - **Fatima Jahara**, fatimajahara@ieee.org
 - **Adrita Barua**, adrita766@gmail.com
 - **MD. Asif Iqbal**, asifiqbalsagor123@gmail.com
 - **Avishek Das**, avishek.das.ayan@gmail.com
 - **Omar Sharif**, omar.sharif@cuet.ac.bd
 - **Mohammed Moshiul Hoque**, moshiul_240@cuet.ac.bd
 - **Iqbal H. Sarker**, iqbal@cuet.ac.bd
9. **A Memory Based POS Tagger for Bengali.** Computer Science & Engineering Department, Jadavpur University, Kolkata – 700 032, India, Authors:
 - **Kamal Sarkar**, JUKAMAL2001@YAHOO.COM
 - **Arup Ratan Ghosh**, GHOSH.MAILME@GMAIL.COM
10. **Bengali Parts-of-Speech Tagging Using Global Linear Model.** Center for Educational Technology, Indian Institute of Technology Kharagpur Authors:
 - **Sankar Mukherjee**, sankar1535@gmail.com
 - **Shyamal Kumar Das Mandal**, sdasmandal@cet.iitkgp.ernet.in

11. Approaches of POS Tagging Algorithm for Bangla Corpus. Authors:

- **Maksuda Sultana**, College of Computer Study Department, AMA University, Philippines, mak.sultana@gmail.com
- **Francis G. Balazon**, College of SGS Department, AMA University, Philippines, fbalazon@yahoo.com

12. Part-of-speech (POS) Tagging of Bengali Written Text Corpus. Authors:

- **Niladri Sekhar Dash**, Linguistic Research Unit, Indian Statistical Institute, Kolkata, India, ns_dash@yahoo.com

13. Part of Speech Tagging for Bengali with Hidden Markov Model. Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur. Authors:

- **Sandipan Dandapat**, sandipan@cse.iitkgp.ernet.in
- **Sudeshna Sarkar**, sudeshna@cse.iitkgp.ernet.in

14. Bengali Part of Speech Tagging using Conditional Random Field. Authors:

- **Asif Ekbal**, Department of CSE, Jadavpur University, Kolkata-700032, India, asif.ekbal@gmail.com
- **Rejwanul Haque**, Department of CSE, Jadavpur University, Kolkata-700032, India, rejwanul@gmail.com

-Bengali Word2Vec Model:

1. Bengali Word Embeddings and It's Application in Solving Document Classification Problem. Authors:

- **Adnan Ahmad**, Researcher, Search Engine Pipilika, Department of Computer Science and Engineering, Shahjalal University of Science and Technology, Sylhet, Bangladesh, adnan.ahmad@student.sust.edu
- **Mohammad Ruhul Amin**, PhD student, Computer Science Department, Stony Brook University, NY 11790, USA, moamin@cs.stonybrook.edu

2. BnVec: Towards the Development of Word Embedding for Bangla Language Processing. Authors:

- **Md. Kowsher**, Noakhali Science and Technology University, Noakhali 3814, Bangladesh, ga.kowsher@gmail.com
- **Md. Jashim Uddin**, Noakhali Science and Technology University, Noakhali 3814, Bangladesh

- **Anik Tahabilder**, Western Carolina University, Cullowhee, NC 28723, USA
 - **Nusrat Jahan Prottasha**, Daffodil International University, Dhaka 1207, Bangladesh
 - **Mahid Ahmed**, Noakhali Science and Technology University, Noakhali 3814, Bangladesh
 - **K. M. Rashedul Alam**, Noakhali Science and Technology University, Noakhali 3814, Bangladesh
 - **Tamanna Sultana**, Noakhali Science and Technology University, Noakhali 3814, Bangladesh
3. **Towards Bengali Word Embedding: Corpus Creation, Intrinsic and Extrinsic Evaluations**, Department of Computer Science & Engineering, Chittagong University of Engineering and Technology, Chittagong-4349 Bangladesh. Authors:
- **Md. Rajib Hossain**, rajcsecuet@gmail.com
 - **Mohammed Moshikul Hoque**, moshikul 240@cuet.ac.bd

Web-Resources:

Bengali POS Tagging web links

1. Bengali Natural Language Processing (BNLP)-
<https://bnlp.readthedocs.io/en/latest/>
2. BNLP repository: <https://github.com/sagorbrur/bnlp>
3. Bengali POS (Parts-Of-Speech) Tagging using Indian corpus-
<https://medium.com/analytics-vidhya/bengali-pos-part-of-speech-tagging-using-indian-corpus-e85f47d3ad65>

Bengali Word2Vec Model web resource links:

1. How to train a Bengali Word2Vec Model: <https://www.kaggle.com/csoham/how-to-train-a-bengali-word2vec-model>
2. Bengali Text Visualization using Word2Vec Model: <https://medium.com/analytics-vidhya/bengali-text-visualization-using-word2vec-211e2ed9fa30>
3. NLP with bangla semantic similarity with word2vec and generating bangla text with lstm- <https://sandipanweb.wordpress.com/2021/01/09/nlp-with-bangla-semantic-similarity-with-word2vec-and-generating-bangla-text-with-lstm/>

4. Bangla Word2Vec Training and Visualization- <https://www.kaggle.com/nuhashafnan/bangla-word2vec-training-and-visualization>
5. Bangla Word2Vec Pretrained Model- <https://www.kaggle.com/iftiben10/bangla-word2vec?select=README.md>

Corpus, Datasets & Codes:

Bengali POS Tagging

1. BNLN Basic Tokenization:
<https://colab.research.google.com/drive/1lxRCx2dqaQY7yc8tL27ILHUM7BcUWTBU?usp=sharing>
 - Required Training sets:
<https://github.com/sagorbrur/bnlp/tree/master/model>
2. BNLN Pretrained model Testing on large texts ([Collected from BNLN github library](#))-
<https://colab.research.google.com/drive/1AsHTQITCHymQnTKpT7gWivo38OWJ-NnJ?usp=sharing>
 - Required Training sets:
<https://github.com/sagorbrur/bnlp/tree/master/model>
3. English/Bengali POS Tagging using NLTK:
<https://colab.research.google.com/drive/1vhu-ggZ7YXtL4KYbKyOZbtnloJK99Qae?usp=sharing>
4. Bengali POS Tagging using NLTK on Indian Corpus:
<https://colab.research.google.com/drive/16SshYWxOQD751KczPmSilOwm8P8ObKMm?usp=sharing>

Using Word2Vec Model:

1. BNLN Word2Vec: https://colab.research.google.com/drive/1GOvx6P-AkAZSOUiGzAIJsr_TmWLk4wHC?usp=sharing
 - Required Training sets of BNLN Word2Vec :
<https://drive.google.com/file/d/1cQ8AoSdiX5ATYOzcTjCqpLCV1efB9QzT/vi-ew?usp=sharing>
2. Bengali Text-Visualization using Word2Vec Model:
https://colab.research.google.com/drive/1IQSu97K5VRO62YRSiAWpvDROpZ_wqPN-?usp=sharing
 - Required Font-Style: <https://github.com/Rajspeaks/Bengali-Text-Visualization-using-Word2Vec-Model/blob/main/AmarBangla.ttf?raw=true>

