

Design and Development a Bangla lemmatizer

Under the Supervision

of

Mr. Sowmitra Das

Chairman

Dept. of CSE

Port City International University

Submitted By

Name : Md. Dill Mahbub

Id : CSE 01105808

&

Name : Md. Israfil Athamed

Id : CSE 01105885

Introduction

- ❑ Lemmatization is the process of finding the normalized form of a word.
- ❑ Need to find what word suffix should be removed or added.
- ❑ Our aim will be simplifying the task of Bangla text processing by the lemmatizer.

Some Stats on Bangla

- Speakers: More than 260millions
- 90 millions people in Bangladesh use internet
- Another 75 millions Bengalis use internet across the world, including West Bengal of India.

Objectives

- ▶ To remove technical and implementational limitations that the researchers and scientists were faced previously.
- ▶ To build a model which will help to find the keywords in a sentence.
- ▶ To solve the problem that are faced in term of stemming.

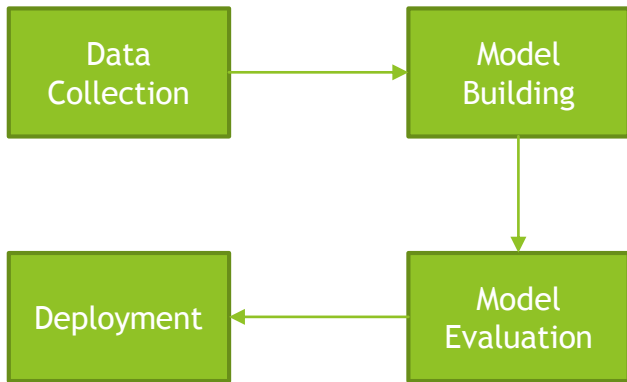
Related Research Review

Title	Author	Publication Date
A Rule based Approach to Word Lemmatization	Plisson J, Lavrac N, Mladenic D	2004
BLTK: The Bengali Natural Language Processing Toolkit	<i>Mr. Saimon Hossain, Mr. Liton Shil</i>	N/A
Automatic Lemmatizer Construction with Focus on OOV Words Lemmatization	Kanis J, Müller L. Automatic	2005 Sep 12
Learning ripple down rules for efficient lemmatization	Juršič M, Mozetič I, Lavrač N	2007
Method of lemmatizer selections in multiplexing lemmatization	Sychev OA, Penskoj NA	2019 Mar

Related Research Review (Contd.)

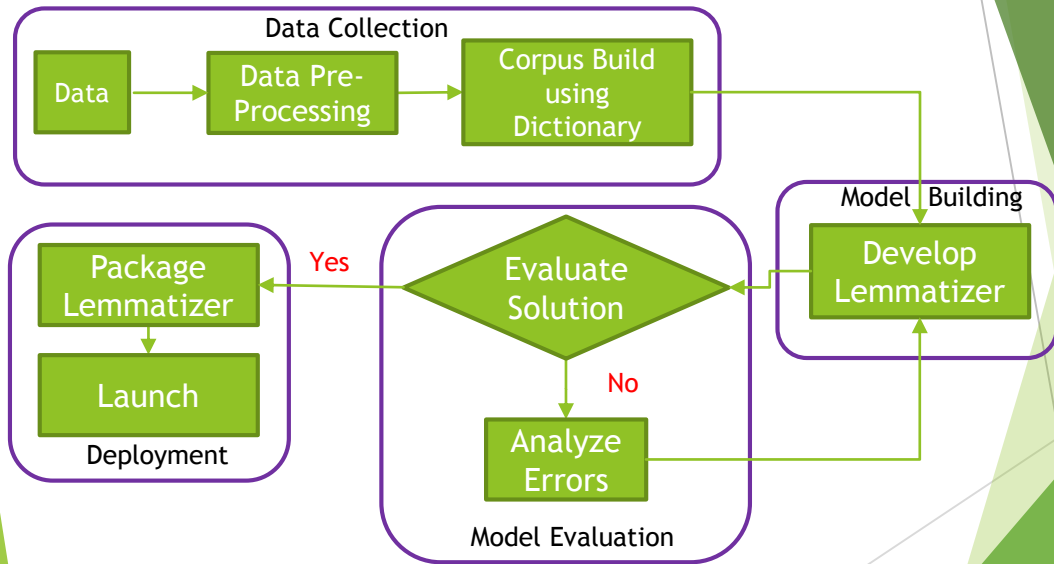
Title	Author	Publication Date
Stemming and lemmatization: A comparison of retrieval performances	Balakrishnan V, Lloyd-Yemoh E	2014
Bangla news classification using naive Bayes classifier	Chy AN, Seddiqui MH, Das S	2014 Mar 8

Research Methodology



System Architecture

6



Tools and Technologies

- ▶ Python
- ▶ TensorFlow

Conclusion

Lemmatization is an essential tool to Bangla Language Processing(BLP), where we can find the Bangla root word. This will help the Reacher's more to get more valuable result in BLP.

References



A Rule based Approach to Word Lemmatization

Plisson J, Lavrac N, Mladenic D. A rule based approach to word lemmatization. In Proceedings of IS 2004 (Vol. 3, pp. 83-86).



BLTK: The Bengali Natural Language Processing Toolkit

<https://pypi.org/project/bltk/>



Automatic Lemmatizer Construction with Focus on OOV Words Lemmatization

Kanis J, Müller L. Automatic lemmatizer construction with focus on OOV words lemmatization. In International Conference on Text, Speech and Dialogue 2005 Sep 12 (pp. 132-139). Springer, Berlin, Heidelberg.

References (Contd.)



Learning ripple down rules for efficient lemmatization

Juršič M, Mozetič I, Lavrač N. Learning ripple down rules for efficient lemmatization. In Proceedings of the 10th international multiconference information society, IS 2007 (pp. 206-209).



Method of lemmatizer selections in multiplexing lemmatization

Sychev OA, Pensky NA. Method of lemmatizer selections in multiplexing lemmatization. In IOP Conference Series: Materials Science and Engineering 2019 Mar (pp. 012091-012091).

References (Contd.)



Stemming and lemmatization: A comparison of retrieval performances

Balakrishnan V, Lloyd-Yemoh E. Stemming and lemmatization: a comparison of retrieval performances.



Bangla news classification using naive Bayes classifier

Chy AN, Seddiqui MH, Das S. Bangla news classification using naive Bayes classifier. In 16th Int'l Conf. Computer and Information Technology 2014 Mar 8 (pp. 366-371). IEEE.