Generating Headline for Nepali News

Datasets:
For model validation
English Gigaword Dataset - the training data consists of 5.5M news articles with 236M words.

For Nepali News
Scrape data from online news portals.

Project Idea:

We propose an application of an recurrent neural network with LSTM units to generating headlines from the text of Nepali news articles. We find that there are state of the art models which are quite effective at concisely paraphrasing English news articles.

Recurrent neural networks have recently been found to be very effective for many transduction tasks - that is transforming text from one form to another. These models are trained on large amounts of input and expected output sequences, and are then able to generate output sequences given inputs never before presented to the model during training.

Software And Tools:

Programming Language: Python3
Libraries Used: Keras, Pandas, NumPy

Team Members and Work Division

Team Members	Work Division
Ashutosh Chapagain (08)	Implement for model validation, tune hyper-parameters, scrape data
Anup Sedhain (43)	Implement on Nepali- Unicode model, scrape data, documentation
Krishu Thapa (58)	Implement on Nepali- Unicode model, scrape data , data cleaning
Aazad Dahal (67)	Implement on Nepali- Unicode model, scrape data, data pre-processing