Assignment 1: Table-to-text Generation

In this assignment, you’ll be implementing some part of the table-to-text generation (<https://www.aclweb.org/anthology/W18-6502.pdf>).

# 0 Setup

To get things set up to run the code, you will need to do the following:

1. Create an Anaconda environment that has Python 3.6 or Python 3.7.
2. Using anaconda, install the following packages into this environment:
	1. PyTorch 1.4 (<https://pytorch.org/>)
	2. NumPy 1.17 (<https://docs.scipy.org/doc/numpy/user/install.html>)
	3. Matplotlib 3.1.2 (<https://matplotlib.org/users/installing.html>)
3. Download the code and data from <https://drive.google.com/open?id=1JViBk5myZRNLw0TEbTgu06atOF-u3PV6>

# 1 Preprocessing (2 pts)

1. In order to apply tensor operations, we must ensure that the sentences in a given batch are of the same length. Thus, we must identify the longest sentence in a batch and pad others to be the same length. Implement the *pad\_vector* function in *utils/loader.py*, which shall produce these padded sentences
2. Implement the *vectorize* function in *utils/loader.py*. This function converts lists of words into tensors. You need to first pad the list of ids and then convert them into vectors.

# 2 Encoder (5 pts)

1. Implement the *forward* function for *structure\_generator/EncoderRNN.py*.

# 3 Decoder (5 pts)

1. Complete the *beam\_search* and *getOverallTopk* functions for *structure\_generator/DecoderRNN.py*.

# 5 Results (3pts)

1. (20 pts) After you finished training, report the BLEU, Rouge, and METEOR score for greedy decoding by running *python main.py --cuda --mode 3.* Check the output of the model, analyze errors and provide potential solutions.
2. (20 pts) Report the BLEU, Rouge, and METEOR score for beam search decoding by running *python main.py --cuda --mode 4.* Check the output of the model, analyze errors compare the results against results of greedy decoding.

# Bonus (up to 10pts)

1. Improve inter-row time expression consistency
2. Handle duplication problems better than beam search
3. Handle gender bias problem
4. Improve entity representation using external resources