A Mini Report

On

NLP Technique Using Word Spell Check using Phython

Poojitha Ram 19WJ5A0527

Under The Guidance Of

Mr. P. Hari Shankar

CONTENTS

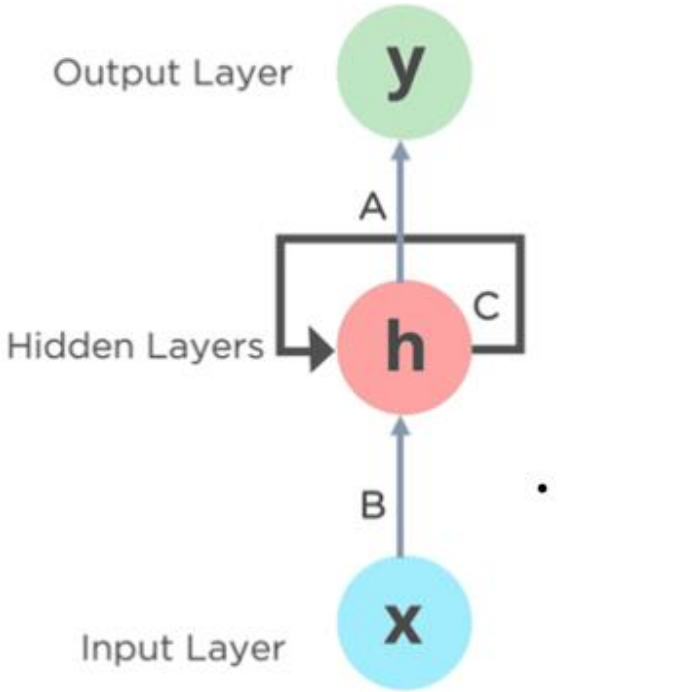
Abstract

- we trained a LSTM with dictionary words and used as an oracle. For a misspelled word, this oracle returns a candidate dictionary word.
- Character level bigram model is used to generate new query words from a misspelled word. These new query words are also given to the trained network for getting more candidate dictionary words. For testing the method’s performance, randomly distorted dictionary words are used. Results showed that the trained network had an acceptable accuracy level.

EXISTING SYSTEM

- Usually websites generally provide spelling correction as suggestions or automatic corrections.
- Correcting obvious errors by auto-correction is acceptable. On the other hand suggestions are more convenient for users to accept a proposed correction without retyping.
- Also, context information is not useful for all cases like general purpose words (e.g. the, this, what). In this project, we focused on developing a spelling correction method which doesn’t require context information

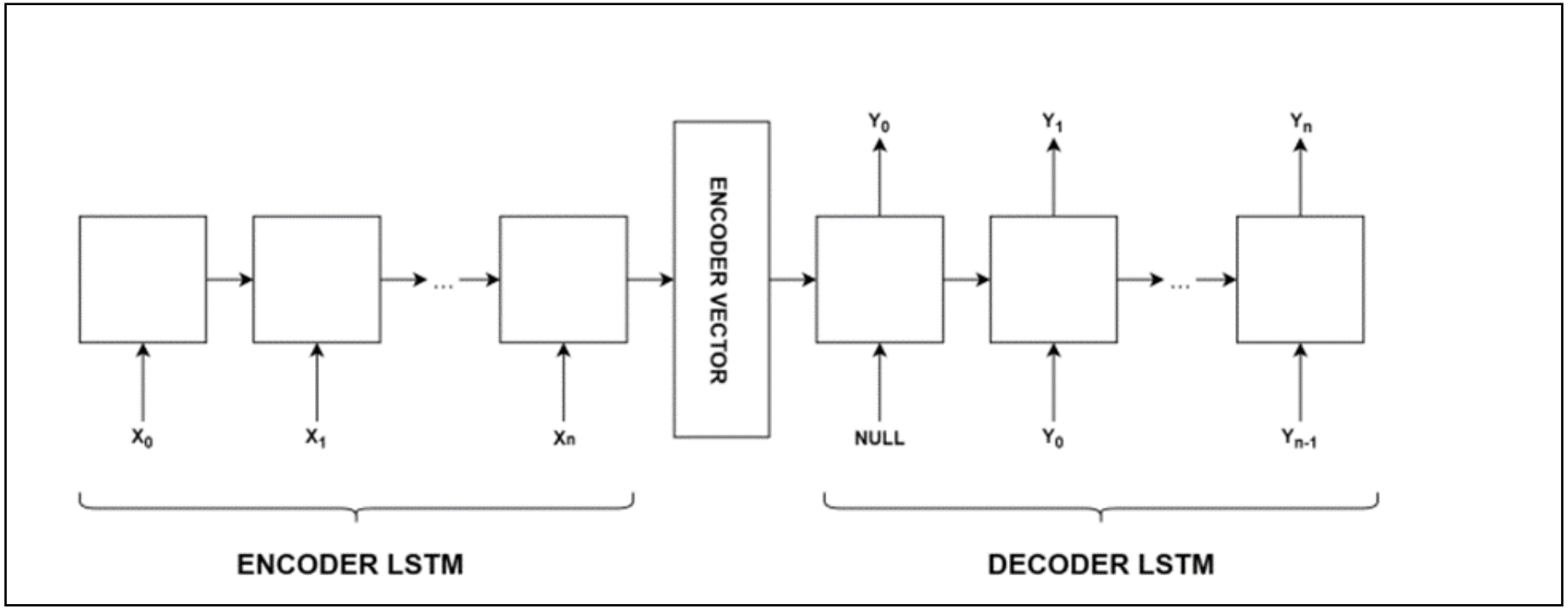
RNN


Output Layer
y
A
C
Hidden Layers
h
B
Input Layer
x
A, B and C are the parameters

PROPOSED SYSTEM

- The proposed method is composed of two parts;
- a trained RNN (LSTM) and
- a character level bigram model.
- Long Short-term Memory (LSTM) has shown to give an extraordinary result in solving sequential problems, including spelling correction. In this project we propose an LSTM model that encodes input word at character level, that also uses word and POS tag contexts as features.

LSTM MODEL


ENCODER VECTOR
Y_0
Y_1
Y_n
X_0
X_1
X_n
NULL
Y_0
Y_{n-1}
ENCODER LSTM
DECODER LSTM

LITERATURE SURVEY-I

Title: Design and Implementation of NLP-Based Spell Checker for the Tamil Language

Author: Pawan Kumar , Abishek Kannan and Nikita Goel

Year: 2020

Description:

A spell checker is a tool used for analyzing and validating spelling mistakes in the text. Recently, the role of a spell checker has diversified, and it is also used to suggest possible corrections to the detected spelling mistakes. Tamil is one of the oldest surviving and international spoken languages of the world, and it is grammatically very rich. Grammar is vital for effective communication and information transmission. However, learning the language rules and the old teaching methodology becomes a challenge for the researchers. The amalgamation of computer and language using natural language processing (NLP) provides a solution to this problem. In this paper, an advanced NLP technique is used to detect wrongly spelled words in the Tamil language text, and to provide possible correct word suggestions and the probability of occurrence of each word in the corpus

LITERATURE SURVEY-II

Title: Swedish Natural Language Processing with Long Short-term Memory Neural Network

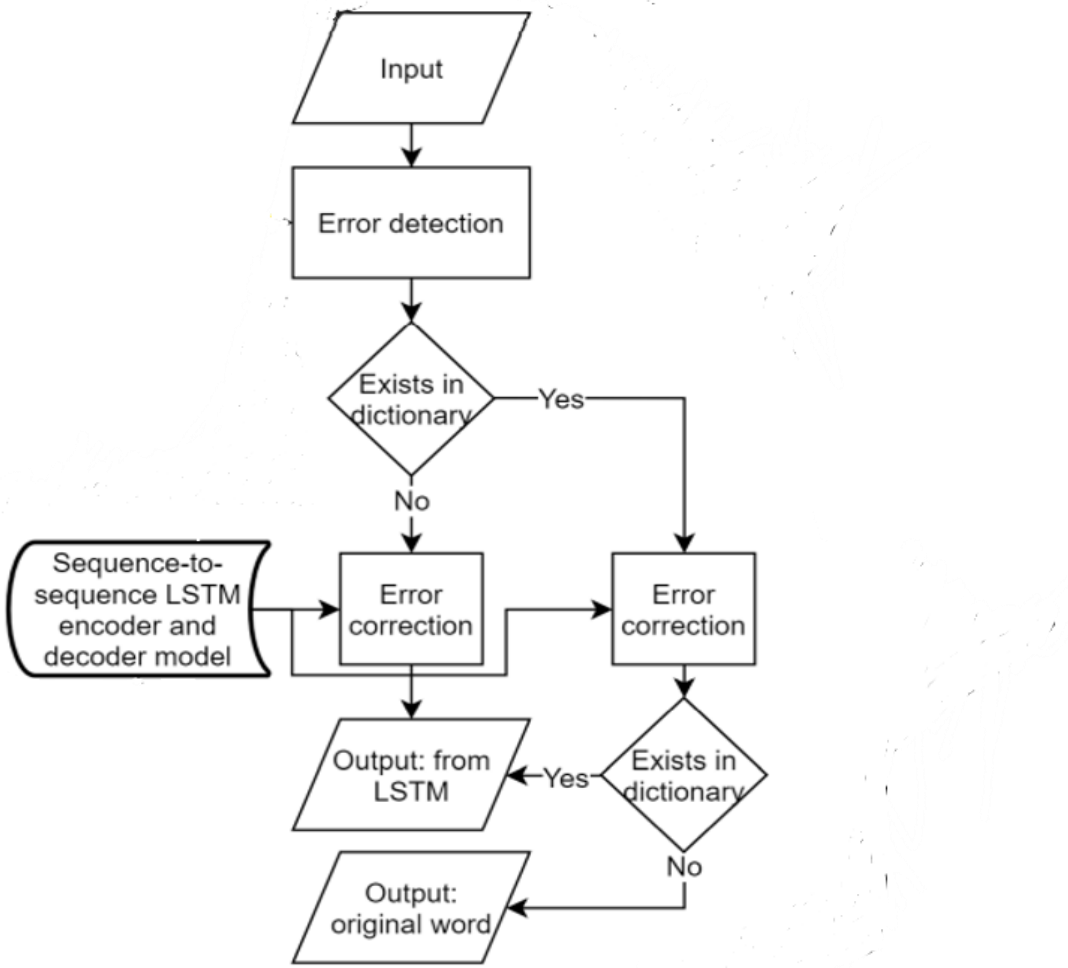
Author: Johan Gudmundsson Francis Menkes

Year: 2018

Description:

Natural Language Processing (NLP) is a field studying computer processing of human language. Recently, neural network language models, a subset of machine learning, have been used to great effect in this field. However, research remains focused on the English language, with few implementations in other languages of the world. This work focuses on how NLP techniques can be used for the task of grammar and spelling correction in the Swedish language, in order to investigate how language models can be applied to non-English languages. We use a controlled experiment to find the hyperparameters most suitable for grammar and spelling correction on the Göteborgs-Posten corpus, using a Long Short-term Memory Recurrent Neural Network. We present promising results for Swedish-specific grammar correction tasks using this kind of neural network; specifically, our network has a high accuracy in completing these tasks, though the accuracy achieved for language-independent typos remains low.

SYSTEM ARCHITECTURE



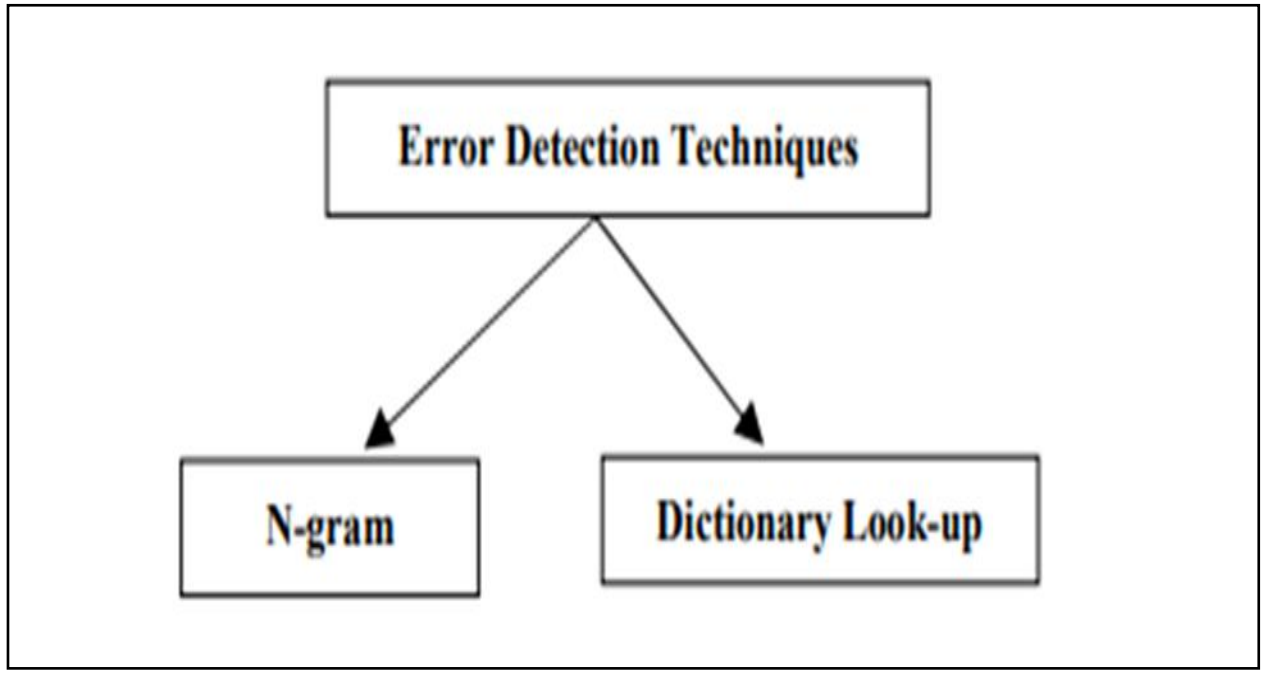
MODULES

- Following are the modules

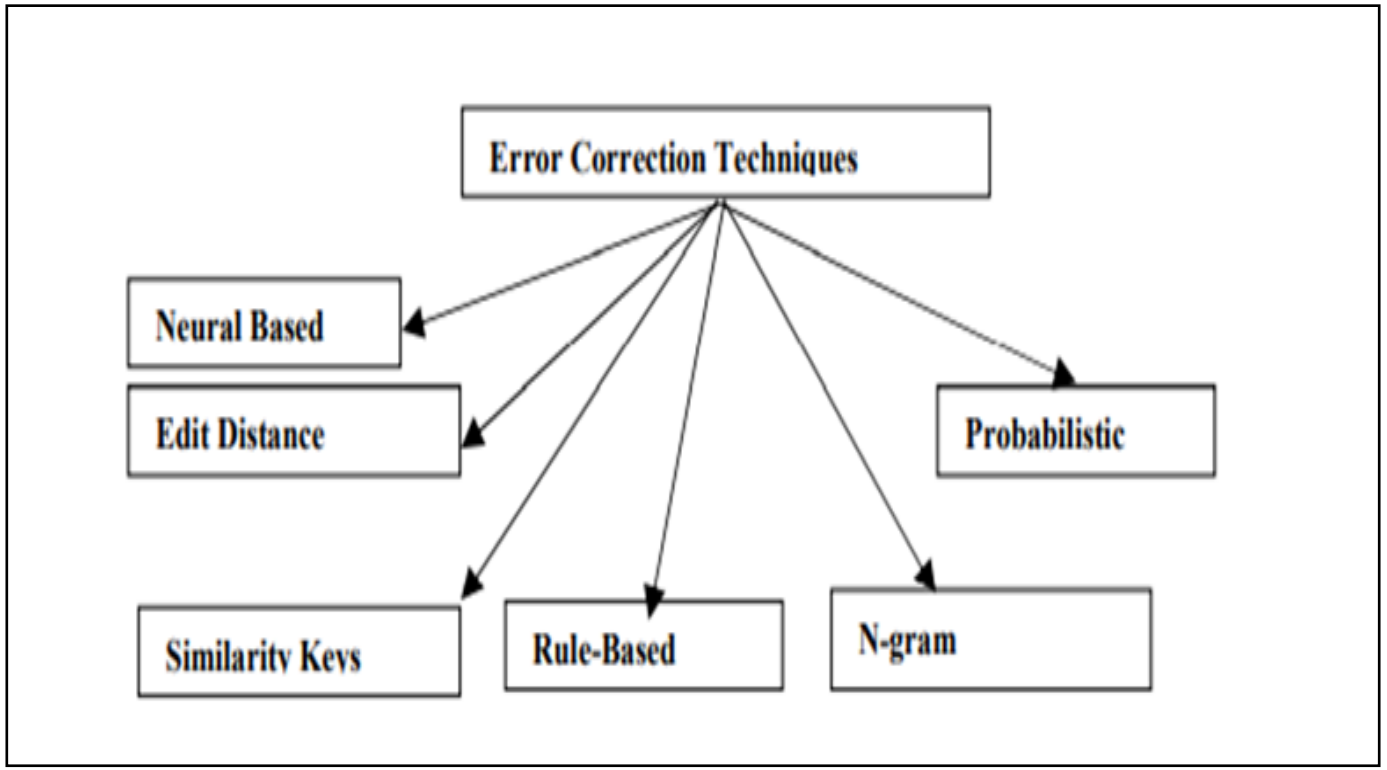
  Error Detection

  Error Correction

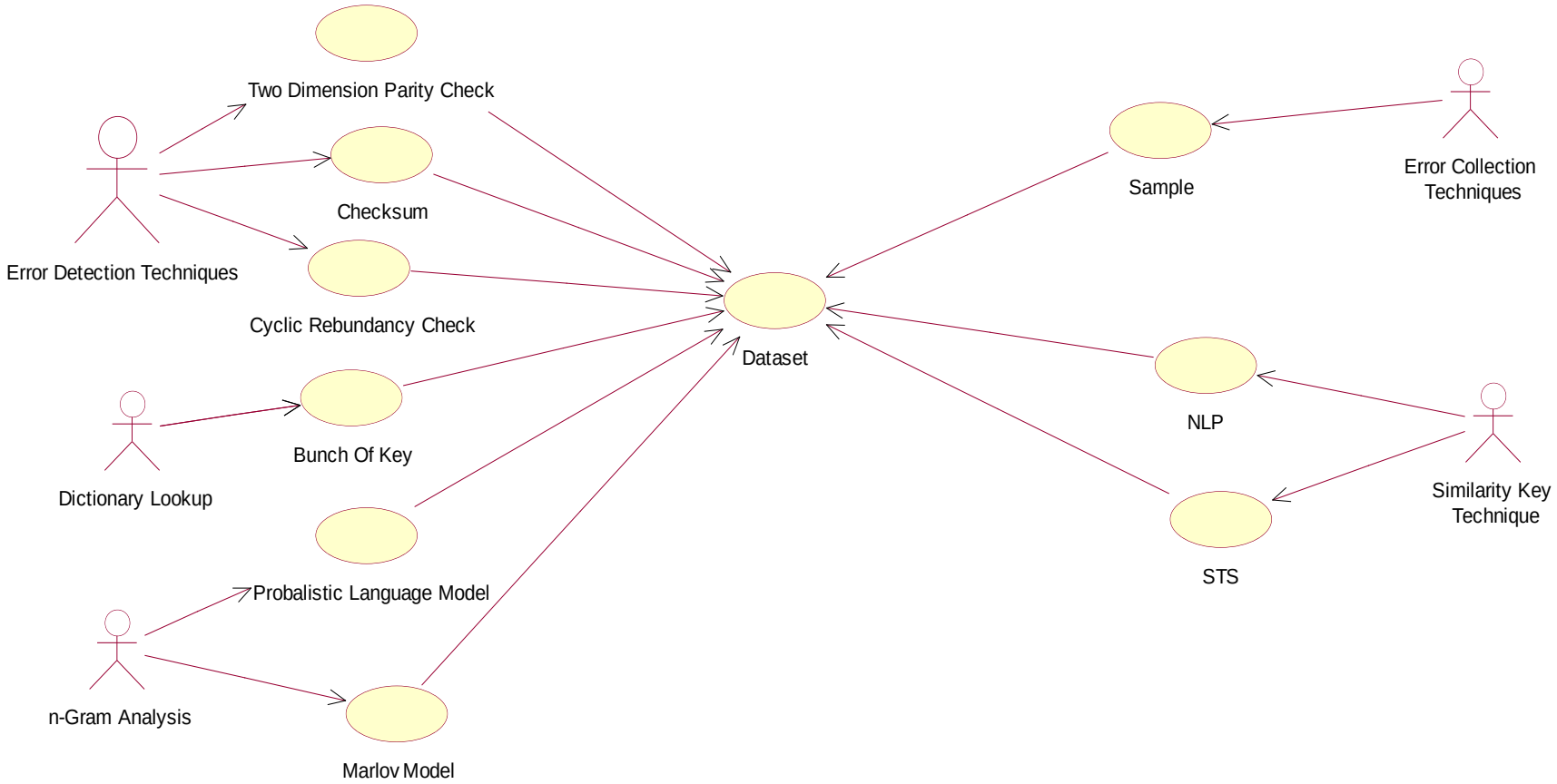
ERROR DETECTION



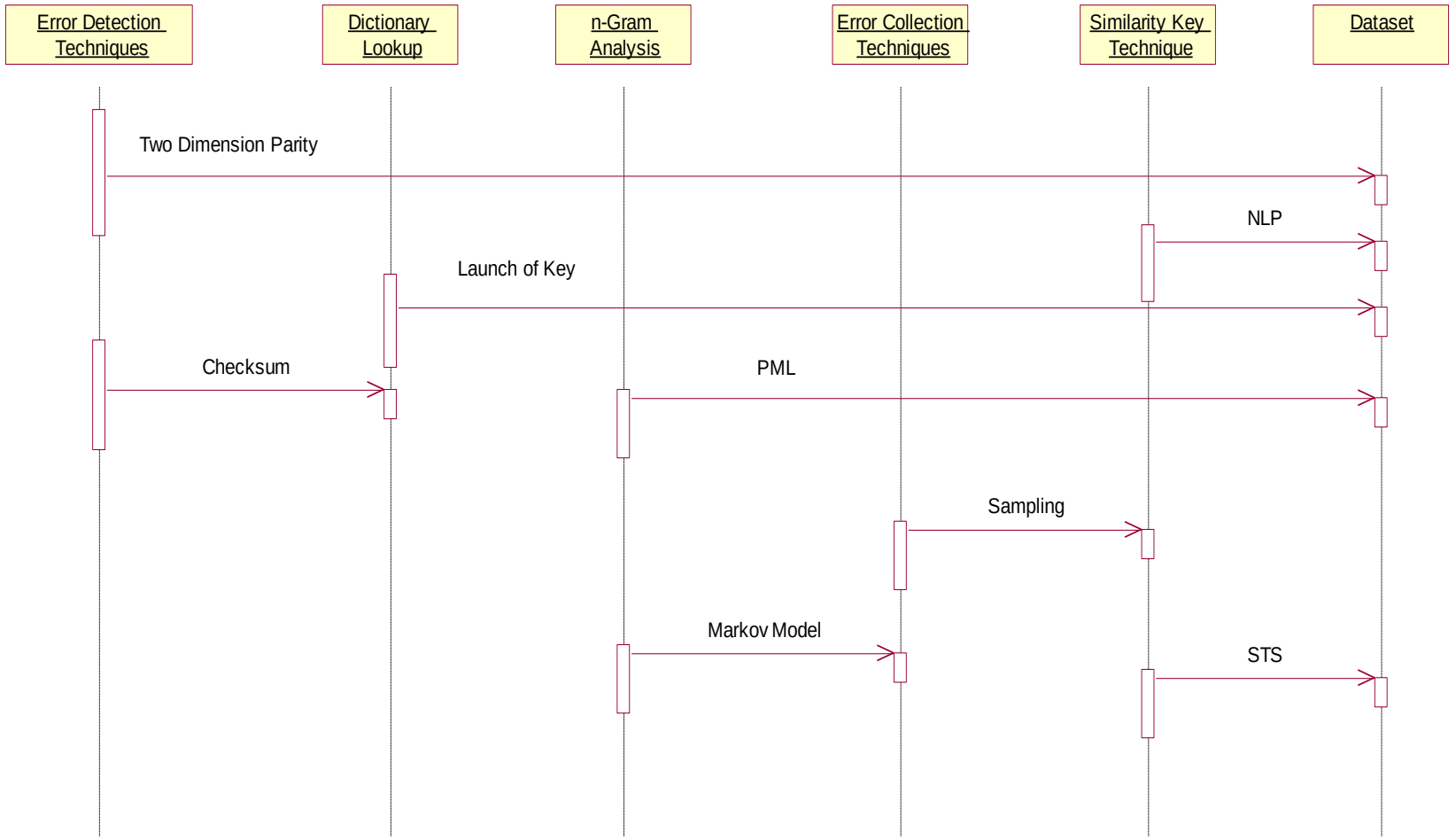
ERROR CORRECTION



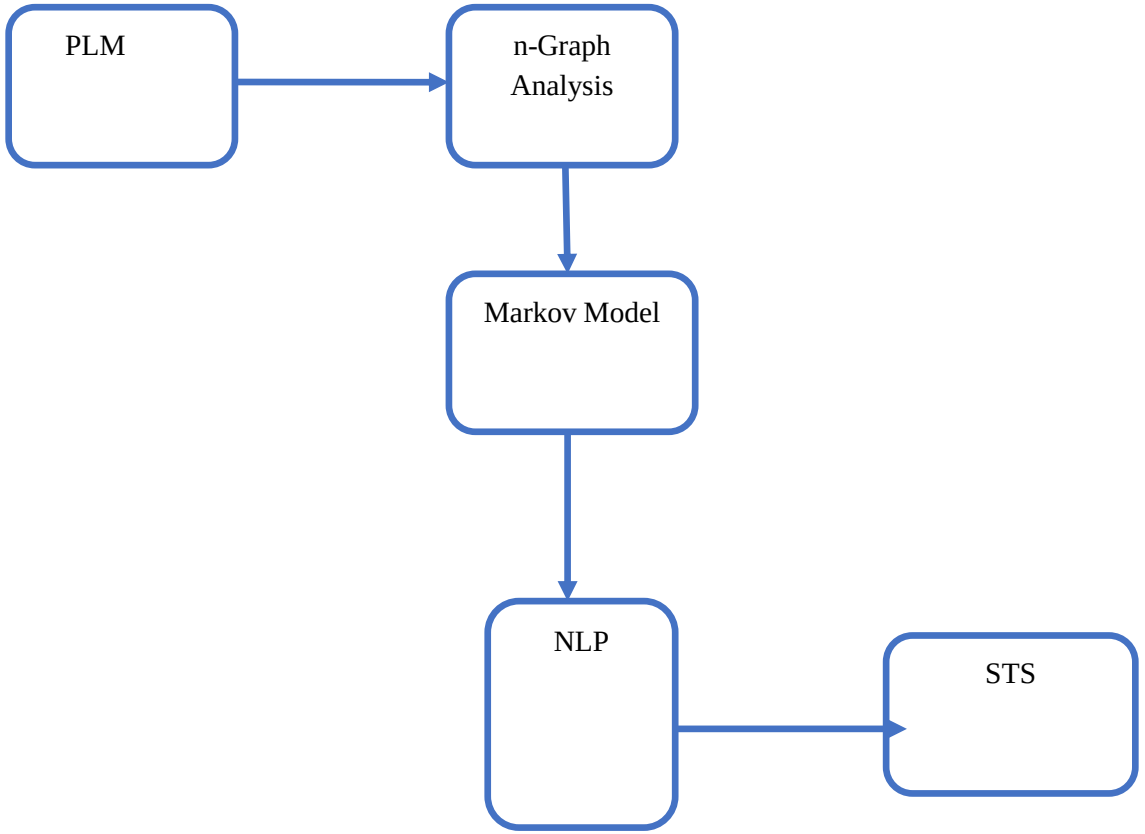
Usecase Diagram



Sequence Diagram



Data Flow Diagram



Snapshots

```
Enter the word to be checked:- hello
original text: hello
corrected text: hello

Try Again? 1 : 0
```

Snapshots

```
Try Again? 1 : 0 1

Enter the word to be checked:- this is implemqntation of naturai language processing techniqus using
word's speli check
original text: this is implemqntation of naturai language processing techniqus using word's speli
check
corrected text: this is implementation of natural language processing technique using word's spell
check

Try Again? 1 : 0 0
```

Conclusion

- The spelling checker in this project is able to detect and correct spelling errors. Since the main purpose of the models is to correct incorrect words, our main evaluation metrics is the recall incorrect. Our Character gives the highest accuracy that is 83.76%. With the given results, we conclude that using LSTM with character as time step feature and additional non-time step features can detect and correct spelling errors in English language. By using more data for training our models, we believe that they will perform better. By using named entity recognition, it should also be able to recognize that named entities should not be evaluated, thus the models can give better accuracies.

-

Future Enhancements

- In this Project, we tried to achieve a less information dependent and context-free spelling correction method. Spelling correction is auto-correction form list.A recurrent neural network model LSTM is used to predict the correct form of a given misspelled word. Using character based bigram model new query words generated in order to increase correction accuracy. Test results are promising but testing with real-life data would give more accurate results about the method’s success.
-

Thank You