CSCE 5290/4290: Natural Language Processing Project Proposal

# **Names:**

Oriana Borges-Salinas, Zachary Chenausky, Anas Mohammed Nayeem, Dagar Rehan

# Project Title:

Machine Translation Techniques Analysis

# Github: <https://github.com/oriAleph/Translation>

# Project Proposal:

For the project there will be four machine translation techniques/models that will be analyzed for accuracy of translation against each other and other industry techniques used by companies. Furthermore there will be an in-depth analysis provided for the results of each of the techniques and how the models can be improved for the future use in regards to precision. Each of the team members will be focussing their resources on a particular technique of implemented translation and analyzing the outputs with reference to the other, we’ve settled on a Spanish dictionary. The expected models should be able to translate words/phrases from Spanish to English and vice versa.

# Description:

1. **Motivation** - The motivation of the project is to compare the accuracy of different techniques and models used in machine translation of one language into another. The reason for this project is to get metrics to compare and contrast the techniques to see what is currently viable, what is already working, how can something that is not that great at translation be improved for better performance. This can better help understand the differences between the approaches and where the applicability of each technique lays.

#Rubric: What is the problem? why do you plan this project?

1. **Significance** - This project is a major aspect of natural language processing and has many different real world applications. It is one of the most researched areas in computer science and used by many companies. These aspects of Machine Translation make it a great learning tool for our team.

#Rubric: Why is project is important? Why was it imperative to accomplish the project purpose?

1. **Objective**

* Compare the accuracy of the techniques against each other
* Compare the accuracy against other models and techniques used in industry
* Analyze the limitation of each technique
* Analyze the benefits of each technique
* How can each technique be improved for better accuracy
* What tradeoffs are required for the technique to be improved

#Rubric: you plan to achieve by the end of your project, what is your goal or result?

1. **Features**

Project features (functions, libraries, frameworks, tools) and Data features (data details, like type of picture it is, eg. rgb, etc)

* Natural Language Toolkit (NLTK)
* spaCy
* TextBlob
* English Dictionary
* Spanish >:D
* REGEX
* GPT Corpus
* Keras
* Collections
* Tensorflow
* Numpy
* Matplotlib

1. **Visualization**

| **Name** | **Task** |
| --- | --- |
| Dagar | Statistical Machine Translation - Word Based Model |
| Zachary | Rules-Based Machine Translation |
| Anas | Statistical Machine Translation - Phrase Based Model |
| Oriana | Neural Machine Translation |

Rubric: some figures: workflow, tables. Then can explain your project.

# Resources:

[4 Types of Machine Translation in NLP | Analytics Steps](https://www.analyticssteps.com/blogs/4-types-machine-translation-nlp)