

## **1. Summary of A *Maximum-Entropy-Inspired-Parser***

- a. This paper proposes and tests some new techniques that can be useful in automatically identifying the structure of sentences, the primary one being a new type of statistical model. Like previous models that would generate a parse tree for a set of text, this model takes in a set of probabilities about how parts of speech interact called a grammar, and outputs a parse tree that applies that grammar to a set of text with a likelihood of that being the correct parse. The model also considers characteristics of each piece of the text like the words around it and other words in the text that are influencing its likely role in the parse. The primary difference in the methodology of this paper against previous work, is that this model conditions the influence of different words only on the most relevant other pieces of data, based on known interactions of the grammar probabilities. Additionally, this paper adds a process by which they use the context of a word to guess how it fits into a sentence before running the full model, which brought a further improvement to the model's performance. Overall, this paper used several new techniques which helped it break through the state of the art of automatic sentence parsing technology.

## **2. Regex to identify email addresses and dates**

- a. [https://github.com/tonycolucci/alc460\\_MSIA414\\_2019/tree/homework1](https://github.com/tonycolucci/alc460_MSIA414_2019/tree/homework1)