**Assignment 3 – Design Document**

We split the problem into **3 stages** of Map-Reduce:

**Job 1 + Job 2:**

Calculate the **4 association metrics** for each **<Lexeme,Feature>** pair.

Job 1 Mapper Keys and Values (input): **<LineId>, Line**

Job 1 Reducer Input Keys and Values:

We made a **custom key** class **WordAndTagKey** that holds **a** **word and a tag**, to be able to calculate the different count metrics: , in the same job.

Each **input record** of the **reducer** is in the form: **<WordAndTagKey, LongWritable>**

There are **4 key options**:  
**<Lexeme, 'Lex'>, count\_l  
<Lexeme Feature, 'Pair'>, count\_lf  
<\*, 'L'>, count\_L**  
**<\*, 'F'>, count\_F**