CSE 390 Assignment 1 README

How to run

Use java 1.8.

- **Language Model Builder**

  1. "Type in file path:" typing in a ABSOLUTE path of train.txt file

     (e.g. /Users/jx/Desktop/2016SP/CSE390/train.txt)

  2. It will print out "languageModel.txt done!" after languageModel.txt is created and print out "Top_20_Joint_Probability.txt done!" after Top_20_Joint_Probability.txt file is created. Those two files is located at where you saved the LanguageModelBuilder.java.

- **Bigram Query Application**

  1. "Language model txt file path"

     typing in a ABSOLUTE path of languageModel.txt file (languageModel.txt was created by LanguageModelBuilder application) e.g. /Users/jx/Documents/workspace/CSE390HomeWork/languageModel.txt

  2. "Choose estimate desired: 1. MLE 2. Laplace 3.Katz backoff (use number only)"

     input number (1 or 2 or 3). E.g. input "1" to use MLE estimation method.

  3. "Type in a pair of words (x, y). Separate by ','"

     E.g. input "w1,w2" (without brackets, w1 and w2 is any words.)

  4. The application will print out a number which is Pr(w1,w2). E.g. 0.001923232. based on what kinds of estimate method you choose at step 2.

- **Language Model Evaluator**

  1. "Language model txt file path"

     Typing in a ABSOLUTE path of languageModel.txt file (languageModel.txt was created by LanguageModelBuilder application) e.g. /Users/jx/Documents/workspace/CSE390HomeWork/languageModel.txt

2. “Test file path:”

   Typing in a ABSOLUTE path of test.txt file (test.txt was provided by instructor) e.g. /Users/jx/Desktop/2016SP/CSE390/test.txt

3. “Choose estimate desired: 1. MLE 2. Laplace 3.Katz backoff (use number only)”

   Input number (1 or 2 or 3). E.g. input “1” to use MLE estimation method.

4. After about 5 seconds, the application will return the perplexity based on what kinds of estimate method you choose at step 3. E.g. “PP is 355”