Homework2 Report

Qiwen Zhu
ID:qiwenz

System Archetype

I used the archetype provided by our TA. I just added a type "Evaluation" to the type system to print out the answers according to their scores.

Implementation Method of Each Annotator

AnswerAnnotator(QuestionAnnotator is just using the same method):
I read the tutorial example1 "RoomNumberAnnotator" and use the same way to develop my own AnswerAnnotator.First, I readin the text and use regular expression to match the Answer. By the way I found regular expression is so powerful and amazing.

TokenAnnotator
It is using mainly the same method with AnswerAnnotator. However, the regular expression I used for abstract a token is "[a-zA-Z]+" which regard "doesn't" as two tokens rather than one. I kind of remember that Professor Nyberg suggest us to do so. One error(or opportunity to reduce unrelated word) may occur if "A" is not a comment but a functional word in a sentence since I do not

NgramAnnotator
I used Professor Nyberg's method to achieve NGram. His code could be seen in the class PDF "02 Intelligent Info Systems", I made some changes to make his code work for mine.

AnswerScoreAnnotator
First, I choose all the NGram(1,2,3) within the question and store them in an ArrayList<String>. Then I analysis each answer and store all NGram for each answer in another ArrayList<String>. Finally, I do traversal to the two ArrayList<String> and compare the NGram in them to get the score.

EvaluationAnnotator
I do traversal to "Answer" and get the size and the feature "isCorrect" of it. Then do traversal to "AnswerScore" to get the score of each answer and store them in an double array. Finally I used bubble sort method to rearrange the sequence of the array and record their position change which is helpful for the final printout.

class hierarchy

- java.lang.Object
  - org.apache.uima.analysis_component.AnalysisComponent_ImplBase (implements org.apache.uima.analysis_component.AnalysisCompor
    - org.apache.uima.analysis_component.Annotator_ImplBase
      - org.apache.uima.analysis_component.JCasAnnotator_ImplBase
        - edu.cmu.deiis.types.AnswerAnnotator
        - edu.cmu.deiis.types.AnswerScoreAnnotator
        - edu.cmu.deiis.types.EvaluationAnnotator
        - edu.cmu.deiis.types.NGramAnnotator
        - edu.cmu.deiis.types.QuestionAnnotator
        - edu.cmu.deiis.types.TokenAnnotator
  - org.apache.uima.cas.impl.FeatureStructureImpl (implements java.lang.Cloneable, org.apache.uima.cas.FeatureStructure)
    - org.apache.uima.jcas.cas.TOP
      - org.apache.uima.jcas.cas.AnnotationBase (implements org.apache.uima.cas.AnnotationBaseFS)
        - org.apache.uima.jcas.tcas.Annotation (implements org.apache.uima.cas.text.AnnotationFS)
          - edu.cmu.deiis.types.Annotation
            - edu.cmu.deiis.types.Answer
            - edu.cmu.deiis.types.AnswerScore
            - edu.cmu.deiis.types.Evaluation
            - edu.cmu.deiis.types.NGram
            - edu.cmu.deiis.types.Question
            - edu.cmu.deiis.types.Token
  - org.apache.uima.jcas.cas.TOP_Type
    - org.apache.uima.jcas.cas.AnnotationBase_Type
      - org.apache.uima.jcas.tcas.Annotation_Type
        - edu.cmu.deiis.types.Annotation_Type
          - edu.cmu.deiis.types.Answer_Type
          - edu.cmu.deiis.types.AnswerScore_Type

          - edu.cmu.deiis.types.Evaluation_Type
          - edu.cmu.deiis.types.NGram_Type
          - edu.cmu.deiis.types.Question_Type
          - edu.cmu.deiis.types.Token_Type