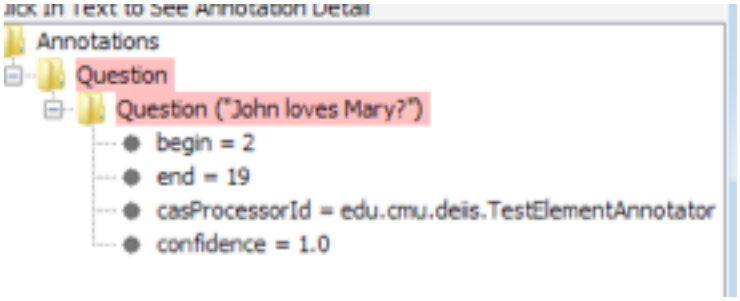
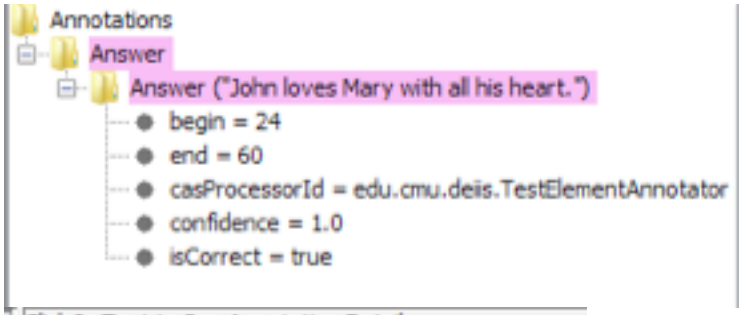
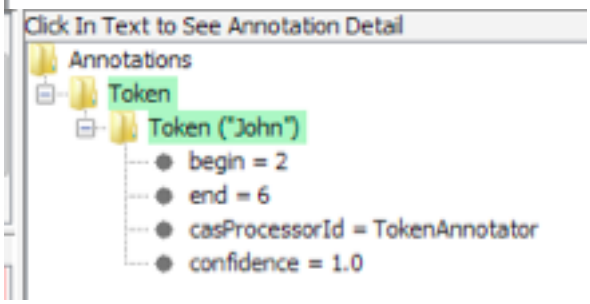
Our annotator will need only one type – `org.apache.uima.tutorial.RoomNumber`. (We use the same namespace conventions as are used for Java classes.) Just as in Java, types have supertypes. The supertype is listed in the second column of the left table. In this case our RoomNumber annotation extends from the built-in type `uima.tcas.Annotation`.

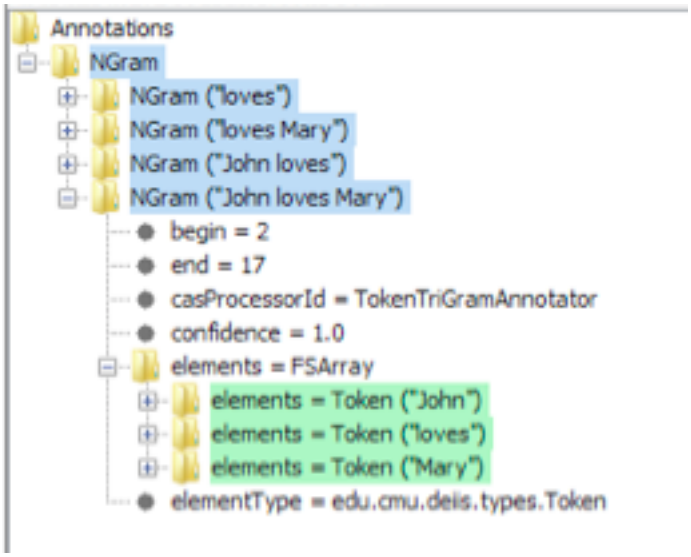
The built-in Annotation type declares three fields (called *Features* in CAS terminology). The features `begin` and `end` store the character offsets of the span of text to which the annotation refers. The feature `sofa` (Subject of Analysis) indicates which document the begin and end offsets point into. The `sofa` feature can be ignored for now since we assume in this tutorial that the CAS contains only one subject of analysis (document).

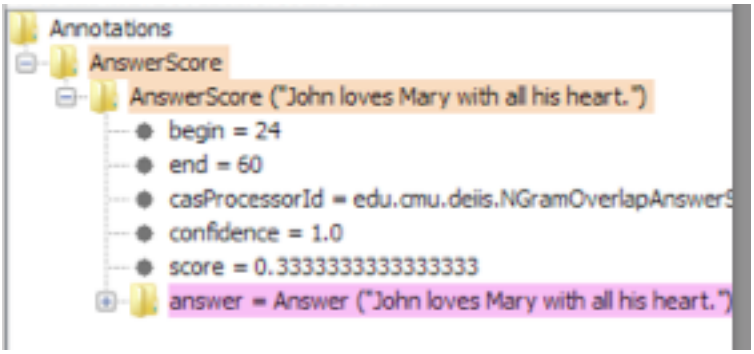
In addition, all annotations must record the name of the component that produced the annotation, and the confidence score assigned to the annotation by the component.




Annotations
Answer
Answer ("John loves Mary with all his heart.")
begin = 24
end = 60
casProcessorId = edu.cmu.deiis.TestElementAnnotator
confidence = 1.0
isCorrect = true


Click In Text to See Annotation Detail
Annotations
Token
Token ("John")
begin = 2
end = 6
casProcessorId = TokenAnnotator
confidence = 1.0


Annotations
NGram
NGram ("loves")
NGram ("loves Mary")
NGram ("John loves")
NGram ("John loves Mary")
begin = 2
end = 17
casProcessorId = TokenTriGramAnnotator
confidence = 1.0
elements = FSArray
elements = Token ("John")
elements = Token ("loves")
elements = Token ("Mary")
elementType = edu.cmu.deiis.types.Token


Annotations
AnswerScore
AnswerScore ("John loves Mary with all his heart.")
begin = 24
end = 60
confidence = 1.0
score = 0.3333333333333333
answer = Answer ("John loves Mary with all his heart.")

- We will evaluate if the type system for the necessary IIS steps and its overall structure are designed and implemented in a proper way, by looking at your codes (Java classes and descriptors) and report. Specifically, each of following items has 5 pts for design and 5 pts for implementation: Overall Structure, Test Element Annotation, Token Annotation, NGram Annotation, Answer Scoring, and Evaluation.

The information processing pipeline will consist of the following steps:

- Test Element Annotation: The system will read in the input file as a UIMA CAS and annotate the question and answer spans. Each answer annotation will also record whether or not the answer is correct.

- Token Annotation: The system will annotate each token span in each question and answer (break on whitespace and punctuation).

- NGram Annotation: The system will annotate 1-, 2- and 3-grams of consecutive tokens.

- Answer Scoring: The system will incorporate a component that will assign an answer score annotation to each answer. The answer score annotation will record the score assigned to the answer.

- Evaluation: The system will sort the answers according to their scores, and calculate precision at N (where N is the total number of correct answers).