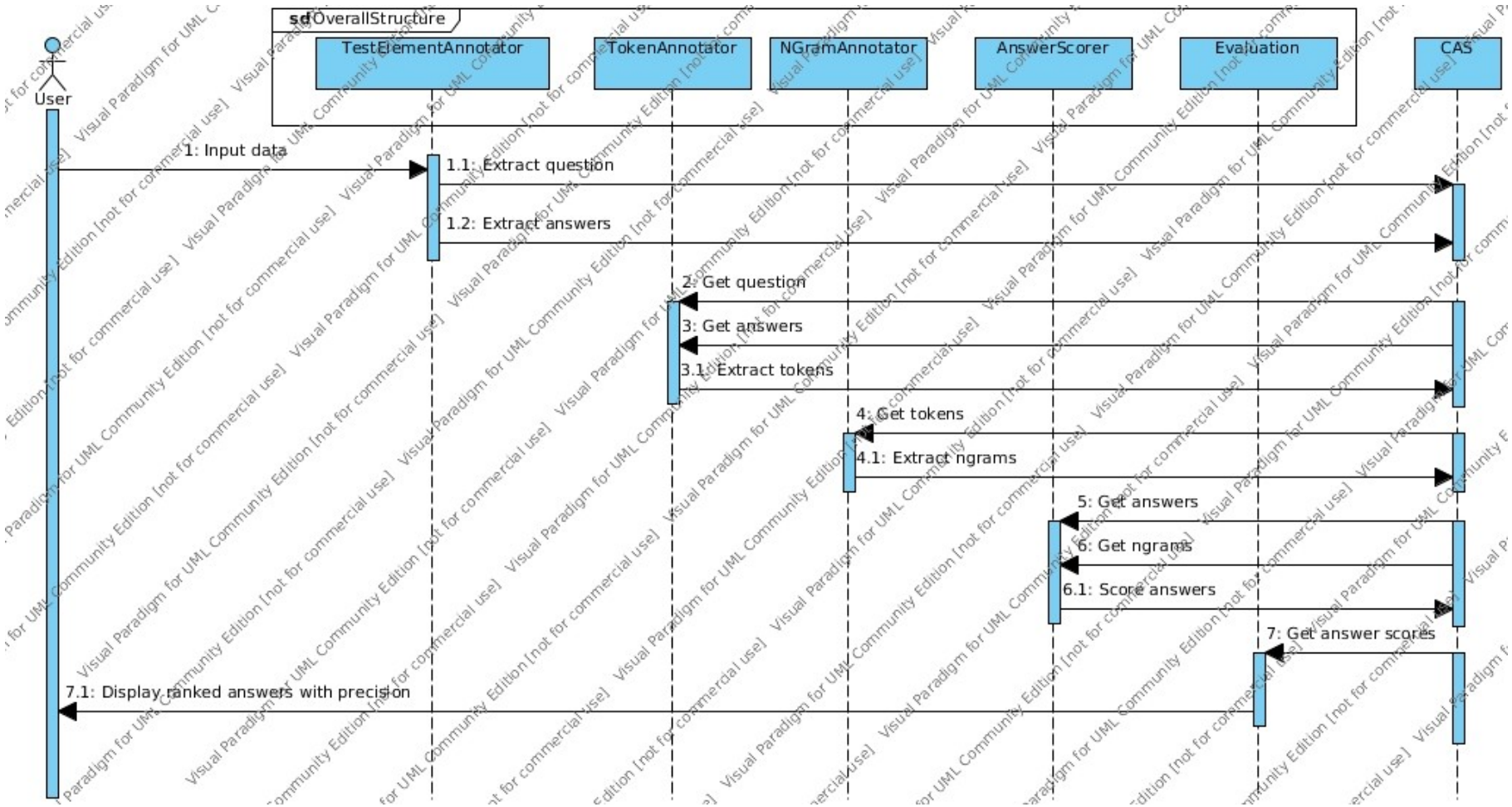
Jonathan Barker

jcb

Hw #2 – Report

The figure below describes the workflow of the system.



The type system provided was not altered or expanded in the design of the system. This system takes a question and answers, scores each answer's correctness, and displays a ranked list of answers and precision.

Example input

```
Q Booth shot Lincoln?

A 1 Booth shot Lincoln.
A 0 Lincoln shot Booth.
A 1 Lincoln was shot by Booth.
A 0 Booth was shot by Lincoln.
A 1 Booth assassinated Lincoln.
A 0 Lincoln assassinated Booth.
A 1 Lincoln was assassinated by Booth.
A 0 Booth was assassinated by Lincoln.
```

Example output

```
Question: Booth shot Lincoln?

+ 1.42 Booth shot Lincoln.
- 1.08 Lincoln shot Booth.
+ 0.72 Booth assassinated Lincoln.
- 0.72 Lincoln assassinated Booth.
+ 0.70 Lincoln was shot by Booth.
- 0.70 Booth was shot by Lincoln.
+ 0.47 Lincoln was assassinated by Booth.
- 0.47 Booth was assassinated by Lincoln.
Precision at 4: 0.50
```

An answer is scored by combining it's question token and ngram overlap scores. Token overlap is simply the number of question tokens found in the answer divided by the number of tokens in the answer. Ngram overlap is calculated the same way except using uni-, bi- and trigrams. After the answers are scored they are ranked in Evaluation and their precision is calculated at N (being the number of correct answers).

Annotator Parameters:

The *TestElementAnnotator* has *questionPattern* and *answerPattern* parameters. The are the regular expression's used to extract questions and answers from the input data.

The *NgramAnnotator* has the parameter *N*. This is a list of the different values of "n" for the ngrams the annotator is extracting.

Annotator Notes:

The *TokenAnnotator* uses the Stanford Tokenizer to tokenize Question and Answers.