# Design

**Tokenization & morphology analysis**

* Sentence ending flag

. ? ! \n

**Syntactic parse**

* **Spacy**
* **Morphology analysis**

**Semantic parse**

* **NER (spacy)**

**Question parse**

* when, who, where, whom, why -> parse directly
* what, which, how -> question focus (the first noun group that is not the word ”name”)

**sentence representation**

* sentence2vec

# Milestone

**Milestone 1**

* overlap(question, sentence\_i)
* sim(question, sentence\_i) based on sentence2vec

**Milestone 2**

* root(question) in (sub)root(sentence)

**Milestone 3**

* extract sentence span

# Data format

**context**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Story #** | **Paragraph #** | **Sentence #** | **Sentence\_raw** | **Sentence\_parse** |
|  |  |  |  |  |
|  |  |  |  |  |

**Question**

|  |  |  |  |
| --- | --- | --- | --- |
| **Question #** | **question** | **focus** | **answer** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |