Final Project Initial Plan
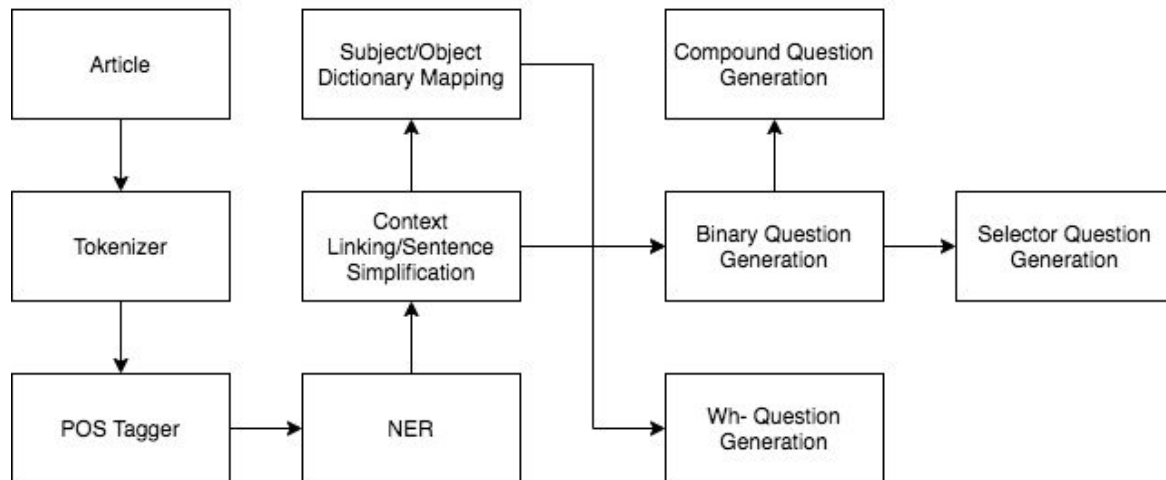
Team <u>404 Name Not Found</u>: Dongtai Du, Xuanxuan Ge, Thomas Li, Jarrett Liang

• How are you going to use the development data to improve your system?

> We plan to use the development data as tests for multiple iterations. For each iteration, we will manually generate several questions including binary questions and wh questions based on the text ones first and compare them with the questions generated by our Q program. Then using the questions to test our A program, we can check if there is anything that can be improved in our programs. Then we can use the input html files to test if our implementation is able to successfully parse the input html contents as well as the correctness of the results.

• Is there going to be a relationship between the asking and answering components of your system, or are you going to implement them independently?

> We are going to implement them independently, given that the requirements for both are also different, and the technical infrastructure is fairly independent. The two segments may share parsing and tokenization components, but the rest will be specialized software.

• What tools are you going to use?

> Nltk (Berkeley Parser), POS, NER, spaCy, Scikit-learn, (possibly) various word-embedding tools. Please refer to the appendix for future details.

• How are you going to share code, data inside your team?

> We will primarily use Slack, Github & a shared Google Drive, as well as AWS if needed.

• How are you going to coordinate development inside the team?

> We will hold weekly meetings and use Slack to exchange information. We will also split into teams of 2, with one team developing the question generating system (Q) and the other working on question answering system (A).

• What technical approaches are you going to take?

> See Appendix for an architectural overview of the proposed pipelines.

Appendix:

1. Question generating system:

Tentative Question Generation Architecture

```
┌──────────┐        ┌──────────────┐        ┌──────────────┐
│ Article  │        │Subject/Object│        │Compound      │
│          │        │Dictionary    │        │Question      │
└────┬─────┘        │Mapping       │        │Generation    │
     │              └──────▲───────┘        └──────▲───────┘
     ▼                     │                       │
┌──────────┐        ┌──────────────┐        ┌──────────────┐        ┌──────────────┐
│Tokenizer │        │Context       │───────▶│Binary        │───────▶│Selector      │
│          │        │Linking/      │        │Question      │        │Question      │
└────┬─────┘        │Sentence      │        │Generation    │        │Generation    │
     │              │Simplification│        └──────────────┘        └──────────────┘
     ▼              └──────▲───────┘
┌──────────┐        ┌──────────────┐        ┌──────────────┐
│POS Tagger│───────▶│     NER      │───────▶│Wh- Question  │
│          │        │              │        │Generation    │
└──────────┘        └──────────────┘        └──────────────┘
```

2. Answer generating system:

Non Deep Learning Approach to Factoid Question Answering