**Project Proposal**

**Team ThePredictive**

1.What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.

Vedant Vipul Jhaveri (vedantj2) - Captain

Sagar Dalwadi (sagardd2)

Sindhu Kopparam (sgk6)

2.What system have you chosen? Which subtopic(s) under the system?

**Theme**: System Extension

**Topic**: Enhance MeTA and Metapy usability

**Subtopics**: Integrate a keyword extraction function in metapy similar to the one existing in SpaCy and integrate chart parser using NLTK.

3. Briefly describe any datasets, algorithms, or techniques you plan to use

* **Keyword extraction**: We will implement or integrate a keyword extraction function in metapy like the one existing in spaCy using the pre-existing en\_core\_web\_sm model for the NLP.
* **Parse tree**: We will use NLTK and integrate the chart parser to show a parse tree for a sentence.
* **Web interface for parameter setting** in MP 2.4 (if advised by instructors): we will use HTML, JavaScript and/or other web programming languages that can interact with python programs in the backend.

4. If you are adding a function, how will you demonstrate that it works as expected? If you are improving a function, how will you show your implementation works better?

* **Keyword extraction**: We will apply the same set of documents/dataset using existing function in spaCy as well as newly integrated function in MeTA and compare the results between the two of them.
* **Parse Tree**: We will parse the same sentence using existing function in NLTK as well as newly integrated function in MeTA and compare the results between the two of them.

5. How will your code communicate with or utilize the system? It is also fine to build your own systems, just please state your plan clearly.

We will use the forked metapy files and edit them to either call functions we implement or call external libraries such as NLTK and spaCy to add the functionality to metapy.

6. Which programming language do you plan to use?

**Programming language**: Python

7. Please justify that the workload of your topic is at least 20\*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

|  |  |
| --- | --- |
| **Task** | **Duration** |
| Keyword extraction - Research | 6 hours |
| Keyword extraction – Test data preparation | 4 hours |
| Keyword extraction - Implementation | 12 hours |
| Keyword extraction – Testing & Validation | 4 hours |
| Parse Tree - Research | 6 hours |
| Parse Tree – Test data preparation | 4 hours |
| Parse Tree - Implementation | 11 hours |
| Parse Tree – Testing and Validation | 4 hours |
| Overall documentation & presentation | 9 hours |