

Team JXC Project Proposal - Free Topic

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.

- Xiaoyu Zhu, xiaoyu23 (Captain)
- Cheng Wei, chengw5
- Jiayi Wu, jiayiwu4

2. What is your free topic? Please give a detailed description. What is the task? Why is it important or interesting? What is your planned approach? What tools, systems or datasets are involved? What is the expected outcome? How are you going to evaluate your work?

- Topic Description: we want to build a proof of concept for a children-friendly vocabulary learning app to help children memorize words easily by returning relevant example sentences ranking by difficulty after a user keys in a word. We think it's very beneficial for parents to utilize this app to help educate their children based on feedback from our team member who is a parent.
- How the app works: a user will enter a word, the app will return the definition, synonyms, and relevant example sentences that are age appropriate and children friendly. The result synonyms and example sentences will be ranked based on difficulty as we would like to return the easiest to understand for better learning experiences. We will exclude stop words the purpose of this app given it doesn't provide much educational value. If we have time in the future to build the product, we will consider including user feedback to rank the results.
- Preliminary data sources:
 - Word definition: (<https://data.world/idrismunir/english-word-meaning-and-usage-examples>)
 - Word ranking: (<https://www.kaggle.com/ratman/english-word-frequency>)
 - Children friendly sentence: [Facebook Children Story Project](<https://research.fb.com/downloads/babi/>)
- Planned approach and task:
 - Research for implementation details
 - Clean and compile datasets
 - Index by sentence
 - We plan to use the Metapy library.
 - Ranking function to rank the difficulty of the sentence
 - We plan to use GloVe to rank words for synonyms.

- Compile evaluation dataset
 - Project Report and Demo
- Expected outcome:
 - After the user enters the word, the following will return as results:
 - The word that the user enters
 - The definition and most relevant synonyms
 - Relevant sentences
 - Evaluation:
 - We will leverage the Cranfield Evaluation Methodology and manually create our own evaluation dataset. We will randomly select a few words and manually rank all the sentences that have the matching word returned by the app and compare it against the ranked results.
3. Which programming language do you plan to use?
- We plan to use Python and Javascript.
4. Please justify that the workload of your topic is at least $20 \times N$ 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	Estimated Hours
Research for implementation details	10
Clean and compile datasets	5
Index by sentence	20
Ranking function	30
Compile evaluation dataset	5
Project Report and Demo	10

Total Estimated hours: 80 hours