

Implementation Plan - Student Project SilkMoth

April 2025

Figure 1 shows a more detailed version of our initial project plan. Note that some tasks may take longer or could be completed earlier than this plan assumes, and we are willing to adjust the plan according to our resources. We aim to parallelize the implementation tasks during the project whenever possible. We split the project into three phases as follows.

1. (17.4 - 15.05) - Core Pipeline

- Get a common understanding of the system
- Implement the main components without major optimization
- Prepare small data set to test correctness and larger data sets for evaluation phase
- Goal: Runnable code for at least the base case (single search pass, similarity threshold $\alpha = 0$, similarity function $\phi = \text{Jac}$)

2. (16.5 - 12.06) - Extended Framework

- Improve the core pipeline
- Refinement and optimization
- Support for discovery mode, $\alpha \neq 0$, $\phi = \text{Eds}$ and $\phi = \text{NEds}$
- Goal: Most features should be finalized and ready for expert review

3. (13.6 - 24.07) - Evaluation

- Improve the system from the feedback and finalize the last functionalities
- Implement the applications to conduct experiments
- Visualize experiment results
- Write report/documentation
- Consider bonus improvements e.g. additional data sets like GitTables¹ or additional similarity functions like Hamming similarity²
- Goal: Presentation and submission of the final system

¹<https://gittables.github.io/>

²https://en.wikipedia.org/wiki/Hamming_distance

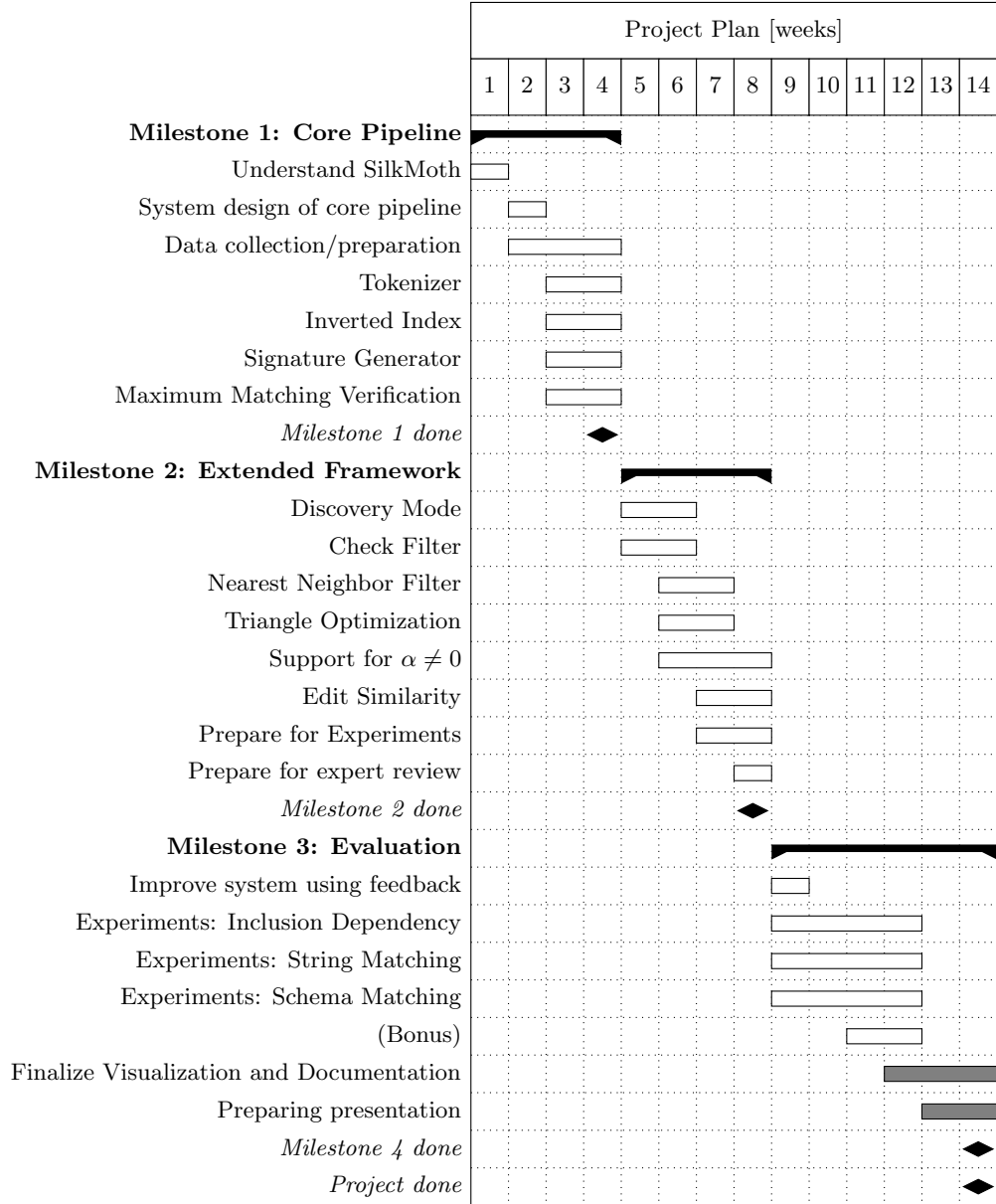


Figure 1: Implementation Plan. First week starting from 17.04.2025.