Motivation:

An imperative stride towards the development and completion of a software product is the rationale behind it. Rationale knowledge, in the context of software engineering, corresponds to the reasoning and logic behind the decision taken during all the phases of software development. Such decisive conversations are spread out across various mediums. Rationales may be discussed face-to-face or over other modes of communication such as chats, emails, or issue trackers to mention a few. Hence, it is important to extract and mine these messages that were exchanged over different modes of communication to manage the rationales. These rationales captured can thus help developers analyze decisions, improve understandability, and document the pertinent information better.

Problem Statement:

The aim of this thesis would be to analyze the different types of rationale elements present in the messages of issue trackers. The focus would be on the following entities:

Issue: the problem to be solved, or a feature to be implemented

Proposal: the feasible solutions that could address the issue

Arguments: the argument put across in favor of or against a particular suggestion

Decision: the final conclusions made to resolve the issue

Approach:

The tentative approach for this research would be as follows:

1) Formulate the research questions based on the investigation to be done

2) Collect the data from the prescribed repositories

3) Manual Content Analysis:

This phase consists of three parts:

3.1) Coding Guide: A coding guide will be developed in this phase where the rationale elements to be studied will be defined.

3.2) Annotation: Based on the coding guide, manual content analysis would be done to annotate the messages and classify them. In the second iteration of this particular phase, the disagreements and conflicts between annotators will be handled.

3.3) Ground Truth: Section 3.1 and Section 3.2 together will help in the formation of ground truth which will be used as the benchmark for measuring the accuracy of the model which will perform the automatic classification of rationale.

4) Automatic Classification:

This phase would involve training the classifier based on the ground truth formation made in the previous step. This phase consists of two parts:

4.1) Binary Classification: Classification of data into two categories. In the context of this research, this would be data containing rationale elements and those that do not contain rationale information.

4.2) Multi-Class Classification: Granular classification of data into multiple classes. In the context of this thesis, this would be classifying data into issues, proposals, arguments and decisions.

Results (Expected):

We expect to have trained a classifier with commendable accuracy and precision to carry out both kinds of classification: binary classification and multi-class classification.

1) The binary classifier will be able to distinguish between the existence and absence of rationale information in a message.

2) The multi-class classifier will be able to make granular classifications by associating a message to the right kind of rationale element such as issue, proposal, argument and decision.

Conclusion:

By the end of this thesis, would have built two kinds of models to carry out automatic extraction of rationale elements from messages and classify them into their appropriate classes based on the ground truth obtained by manual analysis.