I. INTRODUCTION

Extraction is a process of finding the feature values in documents that are useful in the text mining process. Extraction is a crucial part of processing a document to ensure the success of the text mining process [1]. Meanwhile, text mining is a process of exploring and analyzing to get information from a set of data in the form of text and unstructured data, or at least semi-structured. The exploration process of text mining continued in doing text similarity [2].

A match is made between the steps performed and the sequence diagram in this similarity process. Cosine similarity is one of the measurement metrics used to measure the similarity between sentences related to each other and determine a score based on similar words in the sentence.

Software requirements specification (SRS) is a document that explains how to develop software. In simple terms, SRS contains documentation about the details of the application to be worked on. It not only describes the software that will be made, but it must contain information about the purpose of manufacture [3][4][5]. In the SRS document, there are functional requirements specifications, namely to describe the functionality of the system or system services. In addition, there are non-functional requirements used to determine the attributes or overall quality of a system. The SRS also includes use case diagrams, use case descriptions, and sequence diagrams.

In this study, the example of the SRS document used is the Integrated Service Application (APTU) KPKNL Bandung, an application to manage the process of submitting service tickets at the State Property and Auction Service Office. Table I, examples of interpreting the requirements in the SRS document for the APTU Application. There is a difference in interpreting the activities that exist in the Use Case Description artifact with a Sequence Diagram that provides an overview of the functionality of a process to show the involvement of an activity related to the Use Case Description

TABLE I. EXAMPLE OF INTERPERETATION BETWEEN USE CASE DESCRIPTION AND SEQUENCE DIAGRAM

Use Case Description	Sequence Diagram
The APT officer approves the ticket	1. Select the ticket approval menu 2. Press the approve button.

Therefore, to overcome this problem, a process will be carried out to find a similarity from the results of the APTU Software Requirements Specification document, which uses data processing, testing the validity, and reliability of the research results. This study uses a similarity calculation based on contextual or meaning between words using the Word Similarity for Java application tools.

This study the purpose of this study was to extract the steps performed on the Use Case description, then was compared the results of this extraction with the sequence diagram using the concept of text mining.

In this study, there are contributions and novelties, as follows:

a. Extract the steps performed on the use case description and sequence diagram.
b. Implementing text preprocessing on the extracted step performed and sequence diagrams.
c. Based on the extraction results, the similarity process can be carried out between the steps performed in the use case description and the sequence diagram.

Validate through Agreement Coefision (Gwet's AC1).