

CHAPTER 1

1.1 Introduction

Today, most of the real-world scenarios of human life can be express using literature. Literary works have become the most common medium of information. It places an emphasis on many topics that can help address the human nature and its condition, may it be a story of the history or a fictionalized account of a human-tragedy. Through literature, experiences, thoughts, and understandings of different people can be recorded and analyzed by other readers.

In the early life of every person, especially those who are in the domain of arts and literature, they are expected to inevitably read materials as they go on with their living. This contains academic books, stories and other literary works. Plain reading is pretty much very easy to do, but to comprehend with the material it will require some difficulty. In fact, some people find it hard to identify literary elements like characters and their role in the story, the settings, theme, plot and the interpretation of a certain reading. This is due to the reason that most literary works need experience with reading comprehension.

Imagine if computers can do that too, by the time a computer knows how to understand literature, it will be easier for them to understand the way people think. Analyzing literature requires any reader to think critically, and by doing it successfully, they tend to develop new knowledge and information.

The project focuses to create a new way for computers to develop knowledge by using the natural language. It aims to replicate how a human can create logic by extracting various information from a literary work. Information such as description of characters, tone of the text, point of view, settings, summary, and interpretation of the story. The system will be designed to perform Higher-Order Logic for the understanding of the semantics of the text as well as different disciplines of Natural Language Processing for lemmatization purposes and semantics and discourse analysis. The project will be using short stories as a fundamental literary sample input for the system, as it contains most of the key element that the system requires.

1.2 Objectives

1.2.1 General

- To be able to extract the major key elements (i.e. Character Description, Settings, Tone, Point of View, Plot) of a literary work, specifically short stories, as well as process and produce its own interpretation of the given text.

1.2.2 Specific

- To be able to classify the correct labeling of elements in the given literary work.
- To be able to use Higher-Order Logic as the fundamental learning algorithm.
- To be able to generate a natural language derive by the given literary work.

1.3 Conceptual Framework

1.3.1 Conceptual Framework of the Study

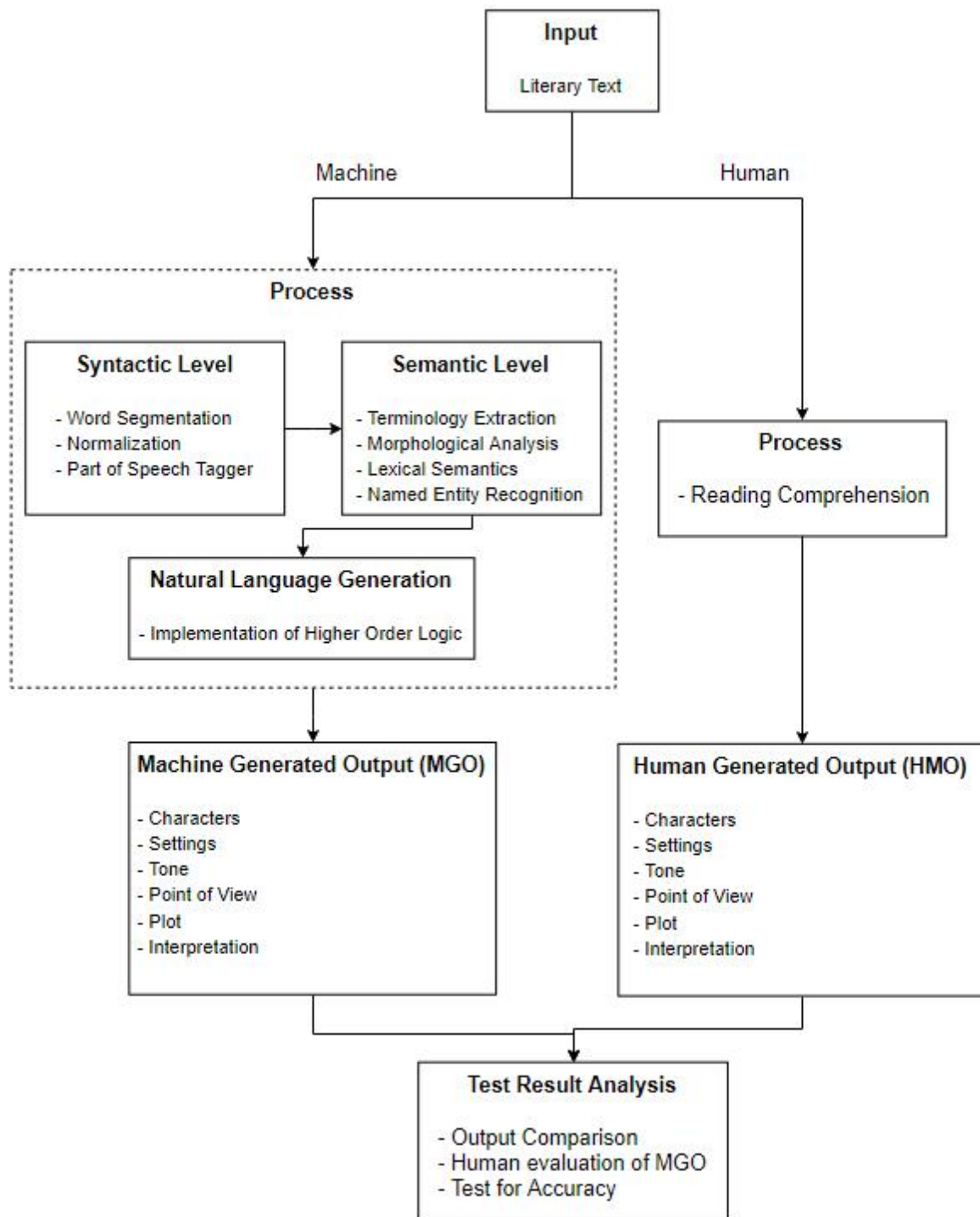


Figure 1.1 Conceptual Framework of the Study

The figure above (Figure 1.1) shows the conceptual framework of the study. Here in this study, the input will be a literary text sample that will be process and comprehend both by machine and human. In this study, the proponents will determine the Precision, Recall, and F-measure of the system through the analysis of test result by comparing the output of both parties (human and machine), furthermore, the result of the machine will be evaluated by an expert whether it achieves the desired output.

1.3.2 Conceptual Framework of the System

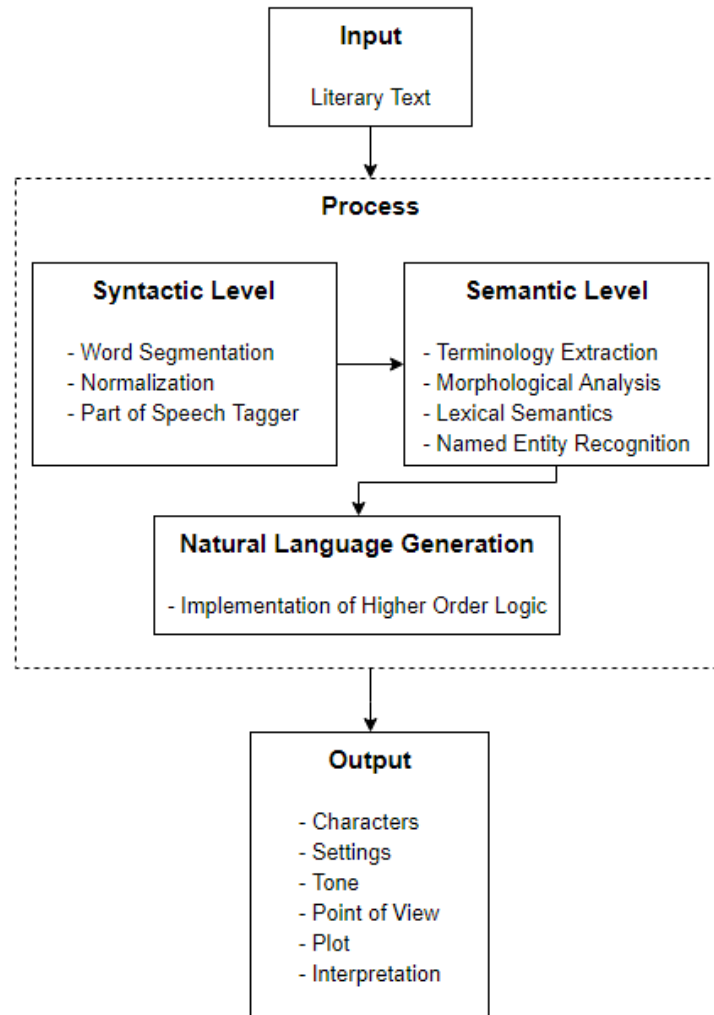


Figure 1.2 Conceptual Framework of the System

The figure above (Figure 1.2) illustrates the initial conceptual framework of the system. The input for the system is a literary text sample in a document or text file format (i.e. .docx / .txt). It will then be process in three levels, in order, Syntactic Level, Semantic Level, and the Generation of Natural Language. In which the first two levels are focus on the Natural Language Processing and the last level on the learning and semantic analysis as well as language generation.

1.4 Statement of the Problem

The study aims to identify the major elements of a short story, such as Characters, Settings, Tone, Point of View, and Plot. As well as to create its own interpretation of the given literary sample. This also aims to answer the following question:

- 1.) What is the accuracy of the identified key elements of a short story in terms of the following?
 - i) Precision
 - ii) Recall
 - iii) F-Measure
- 2.) What is the accuracy of the system's literary interpretation in comparison with the interpretation of humans?

1.5 Significance of the Study

The expected findings of this study aims to achieve a more accurate way of extracting information in a literary text using the combination of concepts of Natural Language Processing (NLP) and Higher-Order Logic.

Furthermore, the system can be an additional refinement in the application of machine learning for it will have a capability to analyze and generate its own natural language base on the given literary sample.

Generating an interpretation of a literary work has always been a tedious task for humans what more if we are expecting the same result from a machine. This study intends to replicate how humans analyze and comprehend literary texts, having a reasoning capability using Higher-Order Logic.

1.6 Scope and Limitation

The study will focus on developing a system that extracts literary information which are characters, tone of the text, point of view, settings, summary, and interpretation of the story. The study will use short story as the material in deconstruction and interpretation. Short story is a narrative that has fully developed theme than plot but significantly shorter than novel. This will be used as an input in the system because short story contains elements that the study requires. English would be the language of the short story to be able to interpret by the system. The study will use Higher Order Logic(HOL) as an algorithm to propose necessary outputs, and to interpret the short story. Higher Order Logic is a form of predicate logic with additional quantifiers and with stronger semantics than first-order logic.

The study will not cover other literary texts such as poems, songs, novels and others. Poems doesn't have the complete key elements that the system needs to deconstruct. The study will not cover short story in other languages, English is rich with information and lot of things to be experimented with. The system will not use other logic-based algorithm, wherein HOL have few predecessor like First-Order Logic(FOL) and Second-Order Logic(SOL). HOL have the power that deconstruction needs because of its stronger semantics and interpretation. In the interpretation, there could be various perspective but in this study there will be only one perspective, which would be neutral perspective.

1.7 Definition of Terms

- Characters (literature) – a person, or sometimes even an animal, who takes part in the action of a short story or other literary work.
- Settings (literature) – the time and place in which the story happens.
- Plot (literature) – a series of events and character actions that relate to the central conflict.
- Tone (literature) – an attitude of a writer towards a subject or an audience.
- Point of View (literature) – the mode of narration that an author employs to let the readers “hear” and “see”.
- Natural language processing – is a field of computer science, artificial intelligence and computational linguistics concerned with the interactions between computers and human (natural) languages (e.g. English, and other language).
- Lexicon – collection of information about the words of a language about the lexical categories to which they belong.
- Syntax – defines the rules exploited to organize the words in a sentence.
- Semantics – study of meaning in language.
- Parser – an algorithm that takes a grammar, a lexicon, and a string of words, decides whether the string of words can be derived from the grammar, and lexicon.
- Knowledge Base(KB) – collection of knowledge used by the system.
- Higher Order Logic – form of predicate logic with additional quantifiers and with stronger semantics than first-order logic.
- Word Segmentation – separation of chunk of continuous text into separate words.
- Normalization – the process of transforming text into a single canonical form that it might not have had before.
- Part of Speech Tagger – given a sentence, it determines the part of speech for each word.
- Terminology Extraction – automatically extract relevant terms from a given corpus.
- Morphological Analysis – separate words into individual morphemes and identify the class of the morphemes.
- Lexical Semantics – analysis of lexical units which include not only words but also sub-words or sub-units and even compound words and phrases.
- Named Entity Recognition – determines which items in the text map to proper names, such as people or places, and what the type of each such name is (e.g. person, location, organization).
- Precision – the fraction of relevant instances among the retrieved instances.
- Recall – the fraction of relevant instances that have been retrieved over total relevant instances.
- F-measure – a measure that combines precision and recall. Also known as harmonic mean or balanced F-score.