ABSTRACT

Automatic recognition of temporal expressions in natural language text is an active area of research in Artificial Intelligence (AI) and Computational Linguistics. Temporal information is the level of information expressing the time of events occurrences described in the text. It conveys the end-points, durations and intervals of events occurrences that happen in the real world. The events may be as significant as “World War II” or as mundane as “Boarding the Bus”. Though temporal information representation and reasoning have been widely discussed in AI, They have started receiving greater attention in the area of Natural Language Processing (NLP).

Extraction and representation of time expressions are necessary components for reasoning about occurrences of events in natural language text as the information derived from event times impose a chronological ordering of the events described in the text. It is also essential for any task requiring resolution and positioning of events on timeline.

The aim of the dissertation is to design a framework for extraction, representation and reasoning with temporal events in natural language text. Event times cannot always be expressed on absolute scale and are inter dependent on other events in the text. Inclusion of temporal relations between events in the model is therefore fundamental. Thus the
framework has been devised to be capable of representing all possible temporal information about events such as end-points, durations, and inter-event relations inorder to facilitate reasoning with them. The properties of a general specification language to markup temporal and event information in text are investigated. The role of the devised framework in the development of robust text summarization systems has been investigated.

In text summarization systems, extractive methods especially, identify and sequence the sentences which involve significant events. The temporal information associated with the event such as the point of occurrence and duration helps to anchor the events on a timeline. The extraction of significant events and the relations among them would therefore suffice in generating cohesive summaries not only for single document but also for multi-documents as well.

In this work, the DUC documents annotated with event time information are used for evaluation of the automatic text summarizer. The generated summaries of lengths varying from 50 to 400 are compared with ideal ones published using ROUGE software. The exact F-measure obtained for the summaries of different lengths are \( F_{50} = 0.29 \), \( F_{100} = 0.39 \), \( F_{200} = 0.44 \), and \( F_{400} = 0.52 \). The average F-measure has been found to be 0.41.