**Automated Cricket News Generation in Sri Lankan Style using Natural Language Generation**

**Main Input Converter:** As mentioned above score card for the relevant cricket match would be the main input to the system. But this score card is in the plain text format. The plain text is converted to JSON tree by a parser while an algorithm will be used to make this task fulfilled. This JSON tree acts as the main input to the text generator module and this JSON tree must be properly structured so that it would be easier to be utilized by the generator. The subsections of the JSON tree would be the game overview with the game results and teams played, scores by the two teams, background information such as location of the game etc. Corresponding to each player there will be a separate node and each of these nodes consists of leaf nodes which contains the performance during the match. Special annotations will also be attached with each player node if there is anything special to be included. The special annotation tags will contain the specific background information on each player. For an example with this player node system is able to deduce that the player got out by playing a poor shot. These custom tags would make the final output text more meaningful.

**Content Selector**: The content selector module would get the input and would manipulate according to some criteria given. There are several approaches for content selection. Some of them are rule based content selection and trainable content selection [28]. In trainable approach to content selection, learning based method is used to choose the content as an independent task or jointly with the surface realization NLG task. In order to use learning method there should be a significant amount of input data. Therefore, in this proposed system, the rule-based approach would be used and trainable approach can be implemented in future and compare it with the current system.

**Text Structuring:** After the content selection is performed, text structuring module is responsible to order these items that were chosen. According to sub sections of the final output such as introduction, result, overview etc., the text selected will be organized. Similar concept will be grouped together according to a rule based method. Content selection and Text structuring will be performed together.

**Aggregator**: Aggregator would get the output of the text structuring module and determine which of the items can be output together from the list. Here duplicate items will be discarded and similar items in each list after text structuring will be aggregated together based on the concepts. The purpose of this module is to form a single sentence with the consecutive items which are potentially candidates for aggregation.

**Surface Realization:** The Surface Realization Module would take the output of the aggregator module together with the linguistic functions, grammatical functions and templates. The linguistic functions would change the numeric or date format to text format. The linguistic functions will convert the data with different types formats such as numeric, date to text format. The grammatical functions would keep Automated Cricket News Generation in Sri Lankan Style using Natural Language Generation 29 the coherence and concordance of the text. Singular and plural words and grammatical gender (male and female) words would be handled here. Sentence or phrase text will be used in this proposed system as it is mentioned above. For an example, for the introduction which contains the match summary with background information, paragraph templates will be used. Most of the time grammatical functions will be helpful in sentence templates than paragraph templates because it contains the most common contents and the templates themselves are linguistically motivates with few gaps to be filled.

**Final Output Generation:** In this final output generation, it would get the output of the surface realization module and some minor validation checks will be performed. If these minor validations return false, it would make the relevant changes. Capitalization of the first letter of a sentence is one such validation check. Output of this module will be final summary generated on the given domain.