Informatics Institute of Technology

in collaboration with

University of Westminster, UK.

Natural Language Generation for Large Documents: A Text-to-Knowledge-to-Text Approach

A dissertation by

Supun Pramoda Wanniarachchi

W1534230 | 2014058

Supervised by

Mr. Sudharshana Welihindha

Submitted in partial fulfillment of the requirements for the
BEng (Hons) Software Engineering Degree

Department of Computing
May 2018

(Word Count – 12792 | Page Count – 100)

Abstract

In generally a significant amount of time and effort need to be spent for document writing and summarizing data. Most of the time people gather few points in their mind and trying to deliver these point as an article. They plan the entire document based on those few points and structure it well. Existing softwares simply get the document title from user and generate few sentences which contain a lot of limitations. They fail to gather what user exactly trying to add to the document. They blindly generate random text and most of them are not useful for the users. Therefore, this research area has not been successfully addressed yet.

The project mainly focuses on developing a Large document generation tool, using A Text-to-Knowledge-to-Text Approach in Natural Language Generation (NLG). Which would understand input of the users and plan the large document content based on user inputs. The main objective of this is to overcome the above mentioned issue of the existing content generation tools and to develop a tool which would feel like written by a human. Using this tool user can plan the content of each paragraphs and only generate the content what they expected. To reduce, the unpredictable input from users the proposed system uses Attempto Controlled English (ACE), which is a controlled natural language (CNL). The major advantage of this tool is, it could be used by anyone who willing to write. They don't have proper domain knowledge about the subject, since all the knowledge based handle by the system itself. User only need to add few points what should include in each paragraph of the large document.

Based on the experts feedback and evaluation questionnaire, it was positive feedback. Therefore, this tool is highly useful for the target audience and the project can be considered as successful.

Subject Descriptors:

- I.2.4: Knowledge Representation Formalisms and Methods
- I.2.5: Programming Languages and Software
- I.7.1: Document and Text Editing
- I.7.2: Document Preparation

Keywords. Natural Language Generation , Attempto Controlled English, Controlled Natural Language, Web Ontology Language