



# Summarizing Persian Text as an Important Application

Nasim Argha  
Department of Computer  
Science and Research Branch of IAU  
Shahrood, Iran  
Nasim.Aargha@yahoo.com

Reza Mirshahi  
Department of Computer  
Science and Research Branch of IAU  
Shahrood, Iran  
Reza\_mirshahi@yahoo.com

*Abstract*— The process of abbreviating sample text and as a result having a short text is called summarization. Input, output and target summarization are types of summarization. After an introduction part, different kinds and examples of summarizer system mentioned. After that a new method to summarize Persian text mentioned. In this new approach, shortest and also most important sentences among input text will be selected. Then a combination of graph based and LSA method considered in order to summarize sample text. In this paper prioritizing words is according to title, keywords and other important factors. At this stage, selecting the sentences which their words has more priority started. At the end of this paper, implementation results mentioned.

*Keywords*- Summarization, Key word, processing language, Abstraction, Prioritizing

## I. INTRODUCTION

Nowadays most of the texts are electronic so it is important and also essential to have summary of long texts. It is not good idea to read long text to get the main point, so trying to use a method in order to summarize text is more time consuming and better than reading the whole long text. The process of summarizing text is called summarization. Using a specific method to summarize a text by machine has some disadvantages: First, we can mark a special part of text that we desire to have summary of that. Second, we informed about the text before. According to recent researches about the function of summarizing system, they conclude that the procedure of summarization is done in three stages: Identification, interpretation, generation.

In next parts of the paper, each of the above functions are explained in detail.

There are different kinds of summarization and summarizer system that mentioned in the following part and also a new summarizing system architecture analysed. Our proposed method can identify both important and unimportant words

and then starts to omit useless words and summarizes the text completely.

## II. SUMMARIZATION TYPE

### A. Classification Method

Classification method first, calculates the word's frequency and then locates sentences. After that, it prioritizes the words and sentences [1, 2].

### B. Term Frequency-Inverse Document Frequency Method

In this method, the number of selected sentences equal with the value tha set before. In sentence vector, the most suitable sentences have important information. The process of omitting unimportant words from vector makes the result brief and also short, so we wouldn't have repetitive information.

### C. Graph Based Method

In this method there is a graph with different nodes. Each word in a text is as a node in graph. It can use graph theory to know similarities between intra document and inter document.

### D. Chain Method

Chain of word has known as a cluster of words with somehow the same meaning, like street, road, highway, etc. We should select the strongest word among this cluster. For instance highway is the strongest word among street, road and highway.

## III. RELATED WORKS

In previous section, some types of summarizing methods introduced. In this part, articles and works related to our method about summarizing Persian text are analyzed.

In our method, it combines graph based and LSA method in order to summarize text. Both methods used in different