AUTOMATIC EXTRACTION-BASED TEXT SUMMARIZATION

HESHAM AHMED SALEH **YAHYA ALAA MASSOUD**

The **Automatic Extraction-Based Text Summarization** tool ranks the sentences within the text by importance using an algorithm that we have developed. It reorganizes the summary to focus on a certain topic; by selection of the most frequent keyword. It extracts the most important sentences and removes insignificant transition phrases, unnecessary clauses, and excessive examples.

The algorithm works by these simplified steps:

Calculate the occurrence of each word in the text.

This is achieved by separating the text in to sentences, and further separating it in to words using the ***strtok()*** function which breaks the given string into a series of tokens using specified delimiters. *[Eg. '.' '?' '-' ',' '!']* The occurrence is calculated afterwards.

Associate words with their grammatical counterparts.

This is done by comparing only the first ***n*** letters of the words with each other, using the ***strncasecmp()*** function which can be found in the string.h library. *[Eg. 'create' and 'creation']*

Assign each word with points depending on their frequency.

After calculating the occurrence of each word, a scoring function assigns points to each word according to the amount of times that word appeared within the text.

Rank sentences by the sum of their words' points.

Separate sentences are ranked by their total score, which is the sum of their individual words' frequency points.

Return some of the most highly ranked sentences in chronological order.

Finally, only a few of the highest ranked sentences are returned in the form of a compact and meaningful summary, in the same order they appeared in the original text.