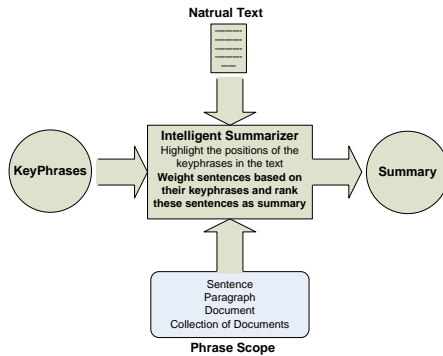


Text Analytics Engine



Description of the invention

The ever increasing growth of textual based information is driving a need for effective text information management and summarization tools. University of Waterloo researchers within the world renowned Pattern Analysis and Machine Intelligence (PAMI) lab have developed a Text Analytics Engine that extracts the most important concepts from natural text and can discriminate between non-important terms and terms which hold the concepts that represent the meaning of the text.

The methodology comprises a number of steps that utilize a natural language analysis approach whereby each sentence is broken down into phrases according to where verbs occur in the sentence. Natural language analysis is further used to identify the important and non-important terms in a sentence with respect to the meaning of the sentence. The methodology then interacts with the WordNet database to abstract the extracted key-terms into new concepts based on their synonyms or related terms. This phase provides an abstraction of the general idea or meaning of the sentence and a ranking system is used to classify the sentences in order of relevancy of meaning. Lastly the methodology generates a summary of the document based on this intelligent natural language analysis.

Advantages

The Waterloo Text Analytics Engine is differentiated from other text and concept summarizers based on its extra sentence based analysis granularity capability that enables extraction of key concepts that have a higher accuracy to the “meaning” of the text being analyzed. A more “meaningful summary” is particularly useful in trying to crack the more difficult challenge of performing sentiment analysis which is a highly coveted commercial capability for companies trying to identify consumer opinion about products/brands.

Potential applications

- Sentiment analysis.
- Enterprise feedback management.
- Enterprise content management.

Reference

8810-7261

Sectors

Sentiment Analysis
Enterprise CRM
Text/Web Mining

Patent status

Issued U.S. 9,075,879

Stage of development

Prototype concept extraction engine developed
Seeking industrial partner for text summarization and intranet/internet search engine

Contact

Scott Inwood
Director of Commercialization
Waterloo Commercialization Office
519-888-4567, ext. 33728
sinwood@uwaterloo.ca
uwaterloo.ca/research