

News Headline Building using Hybrid Headline Generation Technique for Quick Gist

Urmila Shrawankar, G. H. Raisoni College of Engineering, Nagpur, India

Kranti Wankhede, G. H. Raisoni College of Engineering, Nagpur, India

ABSTRACT

A considerable amount of time is required to interpret whole news article to get the gist of it. Therefore, in order to reduce the reading and interpretation time, headlines are necessary. The available techniques for news headline construction mainly includes extractive and abstractive headline generation techniques. In this paper, context based news headline is formed from long news article by using techniques of core Natural Language Processing (NLP) and key terms of news article. Key terms are retrieved from lengthy news article by using various approaches of keyword extraction. The keyphrases are picked out using Keyphrase Extraction Algorithm (KEA) which helps to construct headline syntax along with NLP's parsing technique. Sentence compression algorithm helps to generate compressed sentences from generated parse tree of leading sentences. Headline helps user for reducing cognitive burden of reader by reflecting important contents of news. The objective is to frame headline using key terms for reducing reading time and efforts of reader.

KEYWORDS

Headline Construction, KEA, Keyword Extraction, NLP, Parse Tree Generation, Sentence Compression

INTRODUCTION

Huge amount of information is present in a lengthy news articles. Therefore, rather than reading the detailed news article one can read only the headline and understand the key concept. Headline is an important part of news. It helps to gain vital information in less amount of time. It also comprises of a single sentence that gives an entire idea of a piece of writing. It reduces the cognitive burden of a reader by taking in only significant information.

Reading of lengthy news article is very time consuming and tedious process. Therefore, there is a need of headline for saving reader's time and quick understanding of vital content of news article. Headline construction generally includes analysis of the input text, understanding the important concepts and then formation of headline. For framing effective headline, key terms can be used which are retrieved by applying keyword extraction techniques (Habibi & Popescu-Belis, 2015). Headline generation is one of the vital techniques used to reduce information overload and key concept retrieval from the text (Kaikhah, 2004).

In literature, it is observed that there are many approaches available for automatic headline generation such as statistical versus linguistic, abstractive versus extractive etc. The techniques available are used to improve the quality of headline (Soricut & Marcu, 2007; Banko et al., 2000). In extractive headline generation approach, most relevant text is extracted and gets compressed to the

DOI: 10.4018/IJNCR.2017010103

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

proper size (Malhotra & Dixit, 2013). In abstractive headline generation approach, input text gets analyzed, significant words get selected and then glued together in order to form headline (Banko, Mittal, & Witbrock, 2000). The statistical and linguistic approaches use HMM Hedge model, Hedge Trimmer model (Dorr & Zajic, 2003), and knowledge about language structure for headline generation.

The research work presented in this paper is extended version of the work presented in (Shrawankar & Wankhede, 2016). The aim of this work is to construct headline from detailed lengthy news article by applying keyword extraction and some techniques of NLP. This work is only restricted to English news articles. The dataset used in this work consists of BBC channel news of various types including sports, politics, technical, business, education etc. Along with this, any online news article can also be given as input to the system. The input news article undergoes many pre-processing steps like sentence segmentation, tokenization etc. The key terms are retrieved by applying available keyword extraction techniques (Habibi & Popescu-Belis, 2015) which helps to construct proper headline. The keyphrases are picked out by using Keyphrase Extraction Algorithm (KEA) (Bohne et al., 2011; Witten et al., 1999; Kumar & Srinaathan, 2008; Li, He, & Yangnan, 2014), which helps to improve the quality of headline. As the structured news contains most informative sentences in the leading paragraph, some of them are selected for parse tree generation. Then the parse tree of some leading sentences is generated by using parsing technique of NLP (Li, He, & Yangnan, 2014). Natural language processing is based on the idea of designing and building a computer system that will recognise, analyse, understand and generate natural language sentences which are human language oriented. The applications of NLP are mainly divided into two parts:

- **Text-Based Applications:** Searching a specific topic or keyword in large databases, extracting information from different documents, translating one language into another making it understandable, all these topics comes under text based applications.
- **Dialogue Based Applications:** This involves topics such as answering the questions, telephone services without an operator, teaching systems, voice controlled machines.

The remaining paper is organized as follows. Comparative study is presented in Section II. The system model is presented in Section III. Section IV portrays the implementation details for headline construction. Section V describes the results, analysis and discussion of the system and section VI describes the conclusion of system.

COMPARATIVE STUDY

There are various techniques for keyphrase extraction such as Gen-Ex, Keyphrase Extraction Algorithm (KEA), KP-Miner, N-Gram Filtration etc. Gen-Ex is machine learning based Keyphrase Extraction system made up of two different components: Extraction and Genitor. The extractor is based on the set of heuristics to identify phrases. It includes twelve parameters and flags for truncation such as stem truncation, stem boosting. This way the extractor system stems each selected phrase. In order to optimize the extractor system Genitor was implemented to determine Optimize parameter from keyphrases. KP- Miner is based on heuristics features that rely on term frequency and position. It generates TF-IDF count and then refinement is applied to whole process. It re-orders the final list of keyphrases consisting of long and short keyphrases.

In KEA, the stemming of phrases is done using iterated Lovins stemmer instead of truncator. It focuses on TF-IDF weight of phrases and relative position of candidate phrases in the document.

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/news-headline-building-using-hybrid-headline-generation-technique-for-quick-gist/188781

Related Content

The Worst-Case Stabilization Time of a Self-Stabilizing Algorithm under the Weakly Fair Daemon Model

Tetz C. Huang, Ji-Cherng Lin, Chih-Yuan Chen and Cheng-Pin Wang (2010). *International Journal of Artificial Life Research* (pp. 45-52).

www.irma-international.org/article/worst-case-stabilization-time-self/46028

A Power-Efficient and Quantum-Resistant N-Bit Cryptography Algorithm

Rohini S. Hongal and Rajashekar B. Shettar (2020). *International Journal of Natural Computing Research* (pp. 18-33).

www.irma-international.org/article/a-power-efficient-and-quantum-resistant-n-bit-cryptography-algorithm/267984

Visualization of Protein 3D Structures in 'Double-Centroid' Reduced Representation: Application to Ligand Binding Site Modeling and Screening

Vicente M. Reyes and Vrunda Sheth (2011). *Handbook of Research on Computational and Systems Biology: Interdisciplinary Applications* (pp. 583-598).

www.irma-international.org/chapter/visualization-protein-structures-double-centroid/52334

PSO-CGO: A Particle Swarm Algorithm for Cluster Geometry Optimization

Nuno Lourenço and Francisco Baptista Pereira (2011). *International Journal of Natural Computing Research* (pp. 1-20).

www.irma-international.org/article/pso-cgo-particle-swarm-algorithm/55446

A Secure Data-Hiding Approach Using Particle Swarm Optimization and Pixel Value Difference

Tanmoy Halder, Sunil Karforma and Rupali Halder (2016). *Handbook of Research on Natural Computing for Optimization Problems* (pp. 363-381).

www.irma-international.org/chapter/a-secure-data-hiding-approach-using-particle-swarm-optimization-and-pixel-value-difference/153821