

# A Semantic Text Summarization Method using ontology based Knowledge

Dr.A.Mekala

Department of BCA, Sacred Heart College

## To Cite this Article

Dr.A.Mekala, "A Semantic Text Summarization Method using ontology based Knowledge", *International Journal for Modern Trends in Science and Technology*, 6(11): 122-125, 2020.

## Article Info

Received on 18-October-2020, Revised on 10-November-2020, Accepted on 19-November-2020, Published on 22-November-2020.

## ABSTRACT

Data mining is a method which finds useful patterns from large amount of data. As vast amounts of information are created quickly, effective information access becomes an important matter. Particularly for important domains, such as health check and monetary areas, well-organized recovery of succinct and related information is highly desired. In this paper we propose a new user query based text summarization technique that makes use of WordNet, a common information source from Princeton University. Our summarization structure is expressly tuned to recapitulate health care documents.

**KEYWORDS:** WordNet, Text Summarization, Information Retrieval

## I. INTRODUCTION

Text summarization is the crisis of creating a short, precise, and flowing outline of a longer text file. Automatic text summarization methods are really essential to speak to the ever-growing quantity of text data accessible online to together enhanced help determine relevant in sequence and to consume relevant in sequence earlier. in sequence plays a key role in our humanity. As enormous amounts of knowledge are created and accessible through WWW, how to powerfully and effectively distribute and admittance these valuable data becomes critical. A common Web search engine tries to give out as an information access agent. It retrieves and position in order according to a user's query, and it previously makes a giant impact on how we search and organize information. But present search engines only achieve petty string processing owing to the lack of deep understanding of usual languages and human intelligence, and users usually have to go through pages before they find something useful or provide up. It may not

matter much if user wants in order about a couple of shoes, but it will be a serious difficulty for crucial tasks, such as in medical or else.

## II. METHODOLOGY

Information recovery systems consist of many intricate workings. Investigate and progress of such systems is often hindered by the intricacy in evaluating how each particular section would behave across many systems. We present a work of fiction integrated information retrieval system—the Query, Cluster, Summarize (QCS) system which is portable, modular, and permits carrying out tests with dissimilar instantiations of each of the constituent text analysis components. Most prominently, the grouping of the three types of methods in the QCS design improves retrievals by given that users more focused information well thought-out by topic.

### HelpfulMed

HelpfulMed provides admission to health check information on the Internet and in medical-related

databases for expert and higher users. Users can locate health check information by extracting noun phrases and formative relationships with other medical terminology through concept-based explore support

### **THE NEED FOR ONTOLOGY**

We notice that all the above mentioned work assumes that all information provided by different sources to be integrated is covered by a domain model. On the other hand, information is not necessarily presented in the same way. Due to this fact, information switch over is not an easy task if different actors (producers or consumers of information) have not agreed on the semantic of data. It is obligatory then to define an "alphabet" to ensure a good interpretation and accepting of exchanged data.

Enormous numbers of stipulations are used in medical domain. To interpret a health check document, understanding of these term and

The role of the ontology is to provide a common model that ensures the smallest requirements for this purpose. In fact, such a model allows one to assemble a common analysis of different sources. Ontology is a technique to decompose humankind into objects, and a way to describe these objects. This is a partial account of the world, depending on the objectives of the designer and the requirements of the submission or system. Designed for each domain, there may be a numeral of ontologies. The use of ontology differs from an application to one more, so are its design and its formalism of representation.

### **WordNet**

WordNet is a machine-readable lexical record for English commonly used in computational linguistics group of people developed at Princeton University. The record consists of linked words, first and foremost nouns, verbs, adjectives and adverbs. These words are organized into synonym sets called synsets, and associated by three lexicon-semantic relations – hypernym, meronym and pertainym.

Their association is incredibly important. Ontology is a description of the concepts and associations. High-quality ontology information is the key to improve the quality of medical information re-possession and management. In this paper we use Unified Medical Language System (UMLS) from National Library of Medicine (NLM) as our main medical ontology understanding base. The Metathesaurus is a multi-lingual terms database

that contains definitions of biomedical terms, their various names, and the relationships among them. The Semantic set of connections categorizes all concepts controlled in the Metathesaurus into organisms, anatomical structures, biological function, chemicals, proceedings, corporal objects, and concepts or ideas. The Semantic Network also defines a set of relationships stuck between these concepts. These relationships supply the structure for the network. The primary relationship is the "isa" link, which establishes the chain of command of types surrounded by the Network. There is also a set of non-hierarchical relationships, such as, "physically related to", "spatially associated to", "temporally associated to", "functionally correlated to" and "conceptually correlated to"

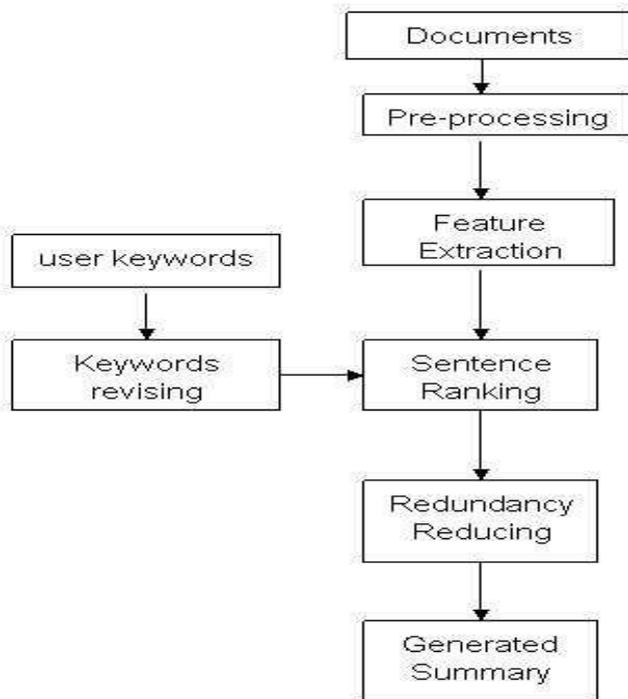
### **III. SUMMARIZATION STRUCTURAL DESIGN**

Text summarization has turn out to be an important region in text mining and generated a lot of research interest recently. There are two types of approaches the knowledge-based approaches construct a semantic representation intended for the summarization task, such as a set of logical forms using ontology knowledge or a template describing a number of key concept etc.

The exterior features-based approaches select summary substance from the source based position in order.

#### **Summarization Procedure:**

Improve the query with WordNet or UMLS ontology information. We will add applicable keywords, delete redundant keywords. We return the revised query and let the user finalize it. Estimate distance of each sentence in the manuscript to the finalized question. Distance function used will be metrics. If the detachment is less important than a threshold, the sentence will be a candidate to be included in the summary.



Calculate pair-wise distances among the applicant sentences. Then, divide contender sentences into groups based on a threshold and select highest-ranked one beginning every group.

#### IV. RESULTS

Even though we have performed a redundancy decrease step, it is not sufficient since the repetitive reporting of the same information from numerous documents.

Be deficient in of syntax examination. Due to the lack of syntax analysis, we rely solely on the quality of the original documents. Grammatically erroneous sentences in the original documents hurt the quality of our summing up.

Query examination. We require a more stylish way to analyze the innovative query.

#### V. DISCUSSION

In this paper we accessible our on-going work on customer query-based summarization scheme and our understanding of participating in the DUC 2007. Ontology knowledge is established to be an efficient way to go away from the mere keyword-based information repositioning methods. With our experiment, we feel that ontology knowledge can be further utilized in other fields of extensive information administration and knowledge finding process. Our future work includes:

- Index and systematize generated summaries for outlook access and use again.

- Integrate our summarization constituent into a broad medical information repositioning system, which may include file clustering, ranking and additional components.
- Use thresholds for selecting sentences in the summing up using statistical data of sentences in the conceptual when it is available.
- make use of some natural language processing techniques in our method, such as parsing and syntax examination.

#### REFERENCES

- [1] Q. Mei and C.X. Zhai, "Discovering Evolutionary Theme Patterns from Text—An Exploration of Temporal Text Mining," Proc. 11<sup>th</sup> ACM SIGKDD Int'l Conf. Knowledge Discovery in Data Mining, pp. 198-207, 2005.
- [2] J. Allan, editor. Topic Detection and Tracking: event-based information organization. Kluwer Academic Publishers, 2002.
- [3] Christian Bizer, Jens Lehmann, Georgi Kobilarov, Soren Auer, Christian Becker, Richard Cyganiak, Sebastian Hellmann. DBpedia - A Crystallization Point for the Web of Data
- [4] Leonhard Hennig, Winfried Umbrath, Robert Wetzker. An Ontology-based Approach to Text Summarization, 2008.
- [5] Xing Jiang, Ah-Hwee Tan. Learning and inferring in user ontology for personalized semantic web search, Information Sciences, 2009, 2794-2808.
- [6] Marek Obitko, Vaclav Snasel, Jan Smid. Ontology design with formal concept analysis, Edited by Vaclav snasel, Radim Belohlavek. In: Proc of the CLA 2004 Intl. workshop on Concept Lattices and their Applications Ostrava, Czech Republic, Sept. 2004, 111-119.
- [7] Hele-Mai Haav. A semi-automatic method to ontology design by using FCA, Edited by Vaclav Snasel, Radim Belohlavek, In: Proc. of the CLA 2004 Intl. Workshop on Concept Lattices and their Applications Ostrava, Czech Republic, Sept. 2004, 13-24.
- [8] Lixin Han, Guihai Chen. A fuzzy clustering method of construction of ontology-based user profiles, Advances in Engineering Software, 2009, 535-540.
- [9] Deryle Lonsdale, David W. Embley, Yihong Ding, Li Xu, Martin Hepp. Reusing ontologies and language components for ontology generation, Data & Knowledge Engineering, 2010, 318-330.
- [10] Rung-Ching Chen, Cho-Tscan Bau, Chun-Ju Yeh. Merging domain ontologies based on the Word-Net system and Fuzzy Formal Concept Analysis techniques, Applied Soft Computing, 2011, 1908-1923.
- [11] T. R. Gruber, "Toward Principles for the design of Ontologies used for Knowledge Sharing", in Proc. of *International Workshop on Formal Ontology*, March 1993.
- [12] Wang, Y., Bai, H., Stanton, M., Chen, W.Y., Chang, E.: Plda: Parallel latent dirichlet allocation for large-scale applications. In: AAIM 2009. 301-314
- [13] Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. *Journal of Machine Learning Research* 3(3) (2003) 993-1022.
- [14] Sontag, D., Roy, D.M.: Complexity of inference in latent dirichlet allocation. In: *Advances in Neural Information Processing Systems (NIPS)*. (2011)

- [15] Chien Chin Chen ,Meng Chang Chen,” TSCAN: A Content Anatomy Approach to Temporal Topic Summarization”, Knowledge and Data Engineering, IEEE Transactions on (Volume:24 , Issue: 1 ),pp 170-183
- [16] T. Hofmann. (1999) Probabilistic latent semantic indexing.
- [17] David M. Blei, Andrew Y. Ng, Michael I. Jordan,” Latent Dirichlet Allocation”, Journal of Machine Learning Research 3 (2003) 993-1022
- [18] D. M. Blei and J. D. Lafferty, “Dynamic topic models,” in Proceedings of the 23rd international conference on Machine learning, ser. ICML '06. New York, NY, USA: ACM, 2006, pp. 113–120.

