ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

An Novel Study Multi-Document Summarization Based On Ontology.

J Refonaa*, S L Jany Shabu, A Christy, S Dhamodaran

Department of Computer Science and Engineering, Sathyabama Institute of Science & Technology, Chennai, Tamil Nadu-600119.

ABSTRACT

Domain metaphysics, as an abstract model, provides a meaty framework for linguistics illustration of matter information. During this paper, we tend to explore the practicability of victimization the ontology in finding multi-document report issues in the domain of disaster management. We offer associate empirical study of various approaches during which the metaphysics has been used for report tasks. In depth experiments on a group of press releases relevant to cyclone Wilma in 2005 demonstrate that ontology-based multi-document report methods beat alternative baselines in terms of the outline quality.

Keywords: framework, multi document,

Corresponding author

ISSN: 0975-8585

INTRODUCTION

It is standard that hurricanes, earthquakes, and other natural disasters cause physical destruction and loss of life and property round the world. So as to efficiently analyze the trend of the disasters and minimize the consequent loss for future scenario, effective info gathering strategies area unit vital. Specifically, a myriad of news and reports that area unit associated with the disaster is also recorded within the variety of text documents. The domain consultants expect to get condensed info concerning the elaborate disaster event description, e.g., the organic process tendency of the disaster, the operational standing of the general public services, and the reconstruction method of the homestead. Within the following, a representative situation is provided, within which the data frequently investigated by a disaster analyst is represented. Scenario: cyclone Wilma gone through South Florida in October, 2005. Throughout Wilma, the ability provide in Miami was influenced. The domain consultants wish to examine the standing of the ability provide throughout Wilma and when Wilma passed. Such data will offer domain analysts a preliminary summary of however the facility offer was influenced by the cyclone, and afterward, domain analysts will contact the corresponding department and establish a group of measures that will be useful once true happens again. In the domain of disaster management, over thousands of hundreds of reports square measure typically free by the regime or native emergency offices throughout the disaster, that cowl most events relevant to the disaster and also the time span are days to months, betting on however severe the disaster is. The data are given in an exceedingly format of newswire, containing plenty of routine news on multiple aspects of the disaster. In such a case, it's extraordinarily troublesome for domain specialists to quickly find either the foremost vital data overall (generic summarization) or the foremost relevant data to a given query (query/topic-focused summarization). Therefore, multi document summarization techniques are often wont to extract meaningful data from multiple reports. A domain metaphysics associated with disaster management, describing the ideas and also the corresponding relations of those concepts, is usually provided by domain specialists [1]. Such an metaphysics contains plentiful abstract data connected to the document set, which can be helpful for users to summarize the documents. A natural question is however we will utilize the metaphysics to get high-quality summaries, i.e., representing topics with non redundant sentences. In this paper, we have a tendency to explore the practicableness of using the metaphysics into multi-document report issues in disaster management domain. We have a tendency to 1st discuss a way to represent a sentence as a vector victimization the domain metaphysics. We then delve into the issues from 2 directions: generic and query-focused report. In generic report, we provide comprehensive studies of the centroid-based sentence selection approaches by victimization totally different vector area models, and explore the chance of utilizing the metaphysics to attain the goal of reducing data redundancy. In query-focused summarization, we have a tendency to optimize the ultimate outline results by employing ontology-based question enlargement ways into the summarization. We have a tendency to conduct experiments on a set of press releases associated with cyclone Wilma, and also the results show that ontology-based ways will offer promising performance for report.

RELATED WORK

GENERIC SUMMARIZATION

For generic report, a strikingness score is typically appointed to each sentence, the sentences area unit hierarchical according to the strikingness score, so the highest hierarchical sentences are hand-picked because the outline supported the ranking result. Recently, each unsupervised and supervised ways have been planned to research the data contained during a document set, and extract extremely salient sentences into the summary supported syntactical or applied math options [2]. For example, MEAD [6] is associate degree implementation of the centroid based method during which the sentence scores area unit computed primarily based on sentence-level and inter-sentence options. However, most existing ways ignore the abstract information within the sentence level. In most cases, the abstract information will offer users additional decipherable results for summaries. Some researchers utilize the express ideas among sentences to handle multi-document report [4], [5], e.g., victimization Wikipedia. However, such techniques can not be directly applied to domain-specific document report tasks, since Wikipedia contains too several ideas not relevant to a particular domain. In our previous work , we explored the chance of victimization domain-specific metaphysics for multi-document summarization; but, no elaborated linguistics relationship is taken into account.

ISSN: 0975-8585

QUERY-FOCUSED SUMMARIZATION

For given topic or question ought to be incorporated into summaries, and the sentences material the user's declared data would like should be extracted. Several ways for generic summarization can be extended to include the question data. Saggion et al. [7] bestowed a strong summarization system developed inside the GATE design that creates use of robust parts for linguistics tagging and co-reference resolution provided by GATE. Wei et al. [11] incorporated the query influence into the mutual reinforcement chain to cope with the necessity for query-oriented multi-document summarization. Wan et al. [12] used each relationships among sentences and relationships between the given question and also the sentences by manifold ranking. Likelihood models have additionally been projected with completely different assumptions on the generation method of the documents and also the queries [13].

QUERY EXPANSION

Query enlargement is that the method of augmenting the user's query with extra terms so as to boost search results. For instance, once we square measure able to search "panther" by some search engine, we are able to expand such question by adding synonyms of "panther" to the question, like "jaguar," "cougar," etc. Query enlargement has conjointly been explored within the field of document summarization, wherever the standard of the generated summary are often improved. As an example, Daume and Marcu propose an even question enlargement technique within the language modeling for IR framework. However, it fails to consider the linguistics connection between the sentences and the question string.

DISASTER MANAGEMENT DOMAIN

DOMAIN DESCRIPTION:

It is renowned that hurricanes, earthquakes, and other natural disasters cause large physical destruction, loss of life and property round the world. The aim of the disaster management program is to boost economical coordination and collaboration among public safety organizations by facultative the practical sharing of emergency alerts and incident related data between disparate systems. One among the disaster management systems aims to investigate the news and reports related to the disaster to supply taciturn and recapitulative information for domain specialists.

DOMAIN-SPECIFIC ONTOLOGY

Generally speaking, a metaphysics is usually provided by domain specialists in disaster management domain [13]. Such an metaphysics provides answers for the queries regarding what entities exist in disaster management, and the way such entities is connected among a hierarchy and divided according to similarities and variations among them. The ontology represented during this paper is said to the domain of cyclone management, involving 109 ideas and 326 concept relations. This metaphysics is obtained from the disaster management project at Florida International University (http://www.bizrecovery.org). The metaphysics is formed for the purpose of analysis enclosed during this project, and is provided by the domain specialists from the State Emergency Operations Center (EOC)1 of Florida. The metaphysics consists of the basis, a set of ideas, a collection of is-a relations, a collection of equivalent-class and a collection of people.

ISSN: 0975-8585

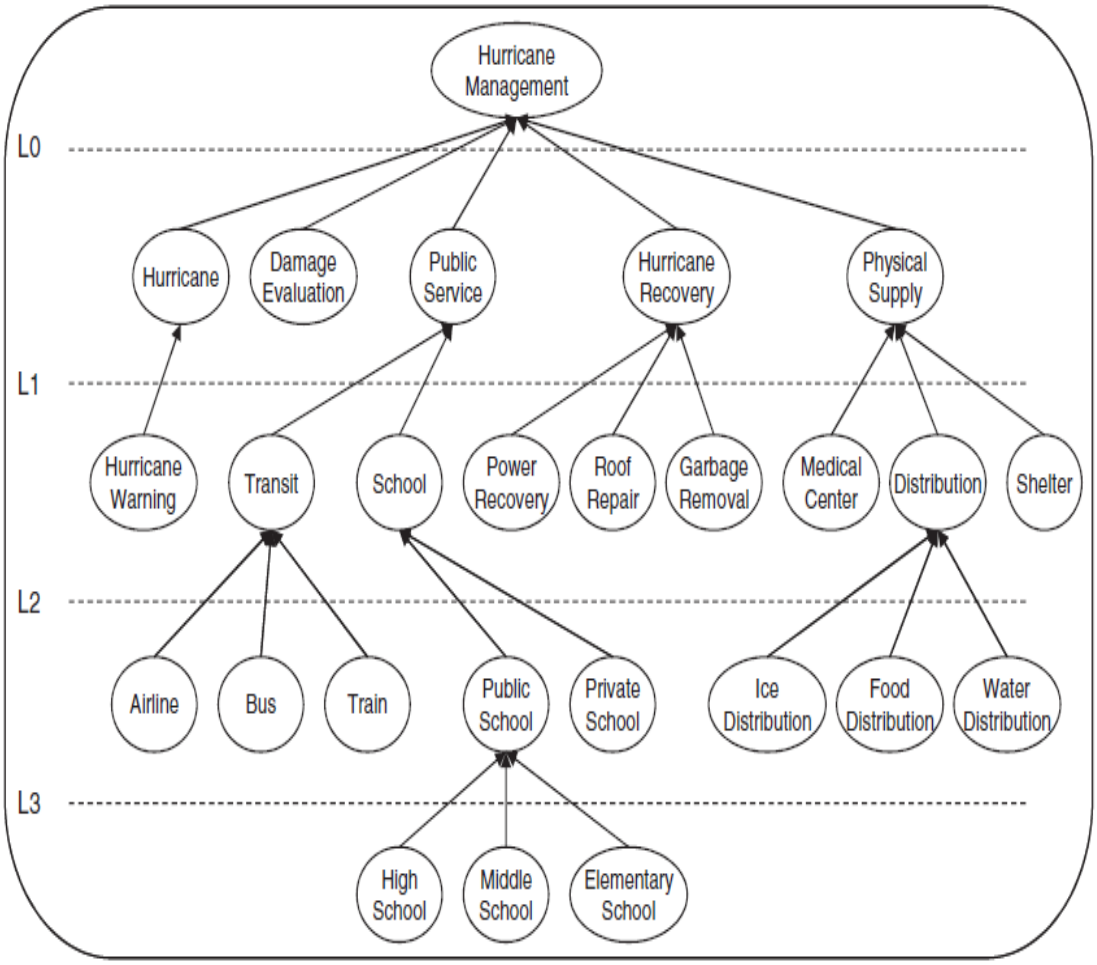

Fig 1: "Hurricane management" ontology.

SUMMARIZATION APPROACHES

To address the summarization problems within the domain of hurricane management, we tend to 1st map most sentences within the document set onto the domain metaphysics, then profit of the intrinsic properties of the metaphysics to represent every sentence. During this section, we tend to explore the result of the metaphysics in multi-document summarization tasks from 2 directions: generic summarization and query-focused summarization.

SENTENCE MAPPING

Ontology in disaster management domain provides US ample conceptual and linguistics data, which could facilitate the procedure of multi-document summarization. To utilize the metaphysics for higher understanding the documents, we at first decompose the gathering of domain-specific documents into sentences, then map every sentence to the metaphysics hierarchy. For every thought of the metaphysics hierarchy, a bunch of keywords (i.e., nouns) area unit appointed by the consultants for the sake of sentence mapping. The procedure of sentence mapping is dead supported the subsequent criteria.

1) If the sentence is said to only 1 thought, map this sentence to the corresponding thought.
2) If the sentence is said to 2 or additional ideas, map this sentence to the smallest amount common ascendant (LCA) of these ideas. If the LCA is that the most general thought of the metaphysics, then map the sentence to the first specific ideas.

In this method, we tend to calculate the word set overlapping between a sentence (only considering nouns within the sentence) and therefore the keyword set appointed to every thought because the live of

relatedness, then rank the scores to pick out the foremost connected concept. Since totally different ideas within the metaphysics have different unambiguous representative noun sets appointed by domain experts, it's unlikely that constant noun can seem in additional than one thought. Once the condition of the second criterion holds, it implies that the sentence contains totally different words that can map to totally different ideas. So as to avoid that a single sentence are going to be connected to multiple ideas and so\ make additional redundant data, we tend to introduce the LCA of concepts and link the sentence to the LCA if it contains 2 or additional ideas. Supported these criteria, we will guarantee that most sentences area unit mapped to a minimum of one thought of the ontology as a result of the metaphysics ideas area unit representative in a specific domain, and therefore the mapping is affordable since the mapped sentences will be considered instances of the corresponding ideas.

SENTENCE REPRESENTATION

For generic account in disaster management domain, the main task generally is to distill the foremost vital overall information from a group of documents associated with the disaster. To emphasize the range of topic coverage during a generic summary, we tend to use the quality K-Means methodology to cluster the sentences of a document assortment into totally different topical groups, then apply sentence coefficient models inside every topical cluster for sentence choice. Additionally, we explore the intrinsic properties of the metaphysics hierarchy to cut back the information redundancy.

CONCLUSION

In this paper, we relate empirical study on methods that utilize the Metaphysics. This sort outs Multi - Documents that are different and Summarize issues in Disaster Management Domain. In case of generic summarization, we have to use Vector House Models which are totally different. The Vector House Models constitutes sentences within the document collection. This also explores the Feasibility of various combinations of the VSMs. For Sentence Set, Centroid - Based ways are used to Cluster them. As a result the Cluster are Extracted from Vital Sentences which are closer to the Centroids of the Sentence. The final outline is generated by ranking sentences and reducing the information lay-off. For Query Focused summarization, we can develop impact on question expansion in the summarization tasks. The metaphysics is made in conceptual information associated with accurate domain. We will be working on the problem of Ontology-Based Multi - Documents Summarization. Most Importantly on Update Summarization and Comparative Summarization which are other Document Summarizations. To improve the Standard of the outline and to perform Stratified Text Categorization , we deeply explore the way to utilize the Stratified Correlations within the Ontology. This is also the most Interesting Direction. Additionally, we use info Extraction Techniques for improving Summarization results. We also have a tendency to extend our planned Methodology for Summarizing the Victimized Public Ontologies. WordNet and Wikipedia are Public Ontologies. For Add-ons General and Quantified problems has to be taken under consideration.

REFERENCES

[1] G. Erkan and D. Radev, “Lexpagerank: Prestige in multi-document text summarization,” in Proc. EMNLP, vol. 4, 2004, pp. 365–371.
[2] H. Hsu, C. Tsai, M. Chiang, and C. Yang, “Topic generation for web document summarization,” in Proc. IEEE SMC, 2008, pp. 3702–3707.
[3] E. Klien, M. Lutz, and W. Kuhn, “Ontology-based discovery of geographic information services—An application in disaster management,” Comput., Environ. Urban Syst., vol. 30, no. 1, pp. 102–123, 2006.
[4] C. Lee, Z. Jian, and L. Huang, “A fuzzy ontology and its application to news summarization,” IEEE Trans. Syst., Man, Cybern., B Cybern., vol. 35, no. 5, pp. 859–880, Oct. 2005.
[5] V. Nastase, “Topic-driven multi-document summarization with encyclopedic knowledge and spreading activation,” in Proc. EMNLP, 2008, pp. 763–772.
[6] D. Radev, H. Jing, M. Sty, and D. Tam, “Centroid-based summarization of multiple documents,” Inf. Process. Manage., vol. 40, no. 6, pp. 919–938, 2004.
[7] H. Saggion, K. Bontcheva, and H. Cunningham, “Robust generic andquery-based summarisation,” in Proc. ECAL, 2003, pp. 235–238.
[8] J. Tang, L. Yao, and D. Chen, “Multi-topic based query-oriented summarization,” in Proc. SDM, 2009.

[9] D. Wang, T. Li, S. Zhu, and C. Ding, “Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization,” in Proc. SIGIR, 2008, pp. 307–314.
[10] X. Wan and J. Yang, “Multi-document summarization using clusterbased link analysis,” in Proc. SIGIR, 2008, pp. 299–306.
[11] F. Wei, W. Li, Q. Lu, and Y. He, “Query-sensitive mutual reinforcement chain and its application in query-oriented multi-document summarization,” in Proc. SIGIR, 2008, pp. 283–290.
[12] X. Wan, J. Yang, and J. Xiao, “Manifold-ranking based topicfocused multi-document summarization,” in Proc. IJCAI, 2007, pp. 2903–2908.
[13] X. Yong-dong, W. Xiao-long, L. Tao, and X. Zhi-ming, “Multi-document summarization based on rhetorical structure: Sentence extraction and evaluation,” in Proc. IEEE SMC, 2008, pp. 3034–3039.