Volume : 2 | Issue : 1 | Jan 2013 • ISSN No 2277 - 8179

Research Paper

Computer Science

KEYWORDS : Document, Matrix, Ontology, Summarization and T-SCAN.

An Ontology based Anatomy Approach to Temporal Topic Summarization

Reka Subbiyan — Computer Science Department, PARK College of tecknology, Coimbatore, India.

Sowmyadevi Kanagasabapathy — Computer Science Department, PARK College of tecknology, Coimbatore, India.

ABSTRACT

Anatomy based summarization method called Topic Summarization and Content Anatomy (TSCAN) was proposed to summarize the content of a temporal topic in existing work. A temporal similarity (TS) function is applied to generate the event dependencies and context similarity to form an evolution graph of the topic. A unique feature of TSCAN is the introduction of the event segmentation process to extract the semantic construct event before summarization. Experimental results indicate that both methods offer similar accuracy in their selections of the paragraphs.

I. INTRODUCTION

Searching the web has played an important role in human life in the past couple of years. A user either searches for specific information or just browses topics which interest him/her. Typically, a user enters a query in natural language, or as a set of keywords, and a search engine answers with a set of documents which are relevant to the query. Then, the user needs to go through the documents to find the information that interests him. However, usually just some parts of the documents contain query-relevant information.

Our approach follows what has been called a term-based strategy: find the most important information in the document(s) by identifying its main terms, and then extract from the document(s) the most important information (i.e., sentences) about these terms .Moreover, to reduce the dimensionality of the term space, we use the latent semantic analysis, which can cluster similar terms and sentences into 'topics' on the basis of their use in context. The sentences that contain the most important topics are then selected for the summary.

II. RELATED WORK

An Efficient Spectral Algorithm for Network Community Discovery and Its Applications to Biological and Social Networks [J. Ruan and W. Zhang] Modern data mining is often confronted with problems arising from complex relationships in data. Recently, datasets that can be represented as graphs, or interaction networks, have received onsiderable attention in various domains.

Web-Page Summarization Using Click through Data [J-T. Sun, D. Shen, H-J. Zeng, Q. Yang, Y. Lu, and Z. Chen] The objective of this research is to study how to use extra knowledge of the click through data to improve Web-page summarization. The click through data contains many users' knowledge on Web pages' contents. Typically, a user's query words issued before a target page is selected often react the true meaning of the target Web-page content. Therefore, it would be helpful if the knowledge contained in the click through data can be uncovered to complement the Web page contents.

III. SYSTEM ARCHITECTURE

In this paper, we have presented a topic anatomy system called TSCAN, which extracts themes, events, and event summaries from topic documents. Moreover, the summarized events are associated by their semantic and temporal relationships, and presented graphically to form an evolution graph of the topic. It shows the deficiency in content extraction. Experiments based on official TDT4 topics demonstrate that TSCAN can produce highly representative summaries that correspond well to the reference summaries composed by experts. Latent semantic summarization methods is no query or topic are provided to the summarization task, summarization outputs and performance judgments tend to lack consensus. Lex Rank and Topic anatomy methods are just use term frequency only, so the accuracy in text anatomy is reduced.

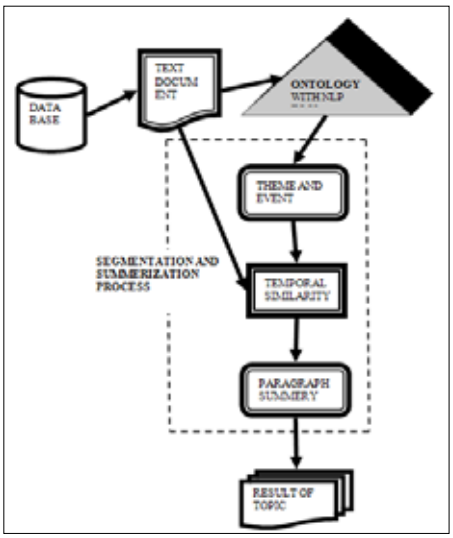

Figure 3.1 System Architecture

In proposed system first, we collect vocabularies and synonyms. Then, we put those terms by the Data model of ontology. The first step of ontology based TSCAN approach is to determine the by comparing the terms of documents with terms in the ontology. If the term does not exist in the ontology, we ignore it. Otherwise, we record the number of times the word appears in the ontology. The ontology decomposes the specific domain into several objects for describing them.

IV. EXPERIMENT RESULTS

The superior SDCS scores achieved by our method demonstrate the advantage of using event segmentation for temporal topic summarization.

Although our summaries are not as diverse as those of the K-means method, they are more coherent.

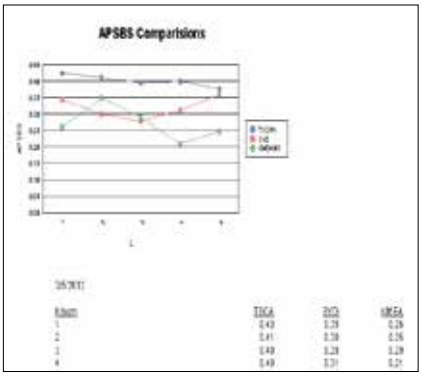

Figure 5.1 APSBS Comparisions

Volume : 2 | Issue : 1 | Jan 2013 • ISSN No 2277 - 8179

A popular measurement frequently used to judge the content coherence of a set of documents is the average pair wise document similarity TSCAN achieves superior APSBS (Average Pair wise Summary Block Similarity) scores.

The reason is that our summaries focus on events in the first few significant themes; therefore, summary blocks have similar contexts. By contrast, the summaries compiled by other approaches try to cover diverse themes, so they are less coherent than our summaries. For all the summarization methods, APSBS decreases as the size of summaries increases.

V. CONCLUSION

The system we have built is a knowledge-based summarization system with the knowledge of topics coming from ontology. The knowledge is composed of not only in recognizing important topics in the document, but also in recognizing the relationships and the relationship types that exist between them. This extracted knowledge is represented in the form of evolution graph. In the future, we will research other method to determine the relationships between concepts more accurately instead of the above simple method and improve the method of ontology construction in a large data set.

REFERENCE

1. Wenjie Li, Mingli Wu and Qin Lu (The Hong Kong Polytechnic University) and Wei Xu and Chunfa Yuan (Tsinghua University). Extractive Summarization using Inter- and Intra- Event Relevance, 2006 | 2. Ani Nenkova. Automatic Text Summarization of Newswire: Lessons Learned from the Document Understanding Conference, 2005. | 3. Maciej Janik, Krys J. Kochut. Wikipedia in action: Ontological Knowledge in Text Categorization. | 4. Christian Bizer, Jens Lehmann, Georgi Kobilarov, Soren Auer, Christian Becker, Richard Cyganiak, Sebastian Hellmann. DBpedia - A Crystallization Point for the Web of Data | 5. Leonhard Hennig, Winfried Umbrath , Robert Wetzker. An Ontology-based Approach to Text Summarization, 2008. | 6. Xing Jiang, Ah-Hwee Tan. Learning and inferencing in user ontology for personalized semantic web search, Information Sciences, 2009, 2794-2808. | 7. 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. | 8. 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 lnfl. Workshop on Concept Lattices and their Applications Ostrava, Czech Republic, Sept. 2004, 13-24. | 9. Lixin Han, Guihai Chen. A fuzzy clustering method of construction of ontology-based user profiles,Advances in Engineering Software, 2009, 535-540. | 10. 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. | 11. 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. | 12. W. Clyde, K. D. Holsapple, K. D. Joshi. A collaborative approach to ontology design. Communications of the ACM, Vol. 45, No. 2, 2002. | 13. B. Ganter, R. Wille, Formal Concept Analysis: Mathematical Foundations. Springer-Verlag Berlin,1999. | 14. Mingli Feng. Construction of User-Query Semantic Ontology (UQSO) for Personalized Topic Search Engine. Xihua University, The thesis of master degree, 2010. | 15. Tzone I. Wang, Tung Cheng Hsieh, Kun Hua Tsai, Ti Kai Chiu, Ming Che Lee. Partially constructed knowledge for semantic query. Expert Systems with Applications 36, 2009, 10168-10179. |