Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 87 (2016) 32 – 37
Procedia Computer Science

Fourth International Conference on Recent Trends in Computer Science & Engineering

A Study on Ontology based Abstractive Summarization

Jishma Mohan M[a,*], Sunitha C [a],Amal Ganesh[a], Dr.Jaya A[b]

[a]*Department of Computer Science & Engineering ,Vidya Academy of Science & Technology,Thrissur,680501,India*
bDepartment of Computer Applications ,B S Abdur Rahman University

Abstract

With widespread use of Internet and the emergence of information aggregation on a large scale, a quality text summarization is essential to effectively condense the information. Automatic summarization systems condense the documents by extracting the most relevant facts. Summarization is commonly classified into two types, extractive and abstractive. Summarization by abstraction needs understanding of the original text and then generating the summary which is semantically related. Abstractive summarization requires the understanding of complex natural language processing tasks. There are many methods adopted for abstractive summarization. Ontology is one among the approach used for getting abstractive summary for a specific domain. In this paper, we discuss about various works carried out using ontology for abstractive text summarization.

Keywords: Abstractive summarization,Ontology

1. Introduction

Summarization is the process of extracting important information from the source text and to present that information to the user in the form of summary. When this is done by means of a computer, i.e. automatically, we call this Automatic Text Summarization [1]. The automatic summarization of text is a well-known task in the field of natural language processing. Document summaries can be abstractive or extractive. Extractive summary extracts the important sections of the text and reproduce them in exactly the same words as were used originally in the text and therefore it is inconsistent. However, abstractive summarization consists of understanding the source text by using linguistic method to interpret and examine the text. Abstractive methods need a deeper analysis of the text. These methods have the ability to generate new sentences, which improves the focus of a summary, reduce its redundancy and keeps a good compression rate [2]. A document summary can be either generic or query-dependent (user-focused). A user-focused summary presents the information that is most relevant to the initial search query, while a generic summary gives an overall sense of the document content. Abstractive summarization techniques are broadly classified into two categories: Structured based approach and Semantic based approach. Different methods that use structured based approach are as follows: tree base method, template based method, ontology based method,

*Corresponding author. Tel.: +91-9947-389-370
E-mail address: mohanjishma@gmail.com

Peer-review under responsibility of the Organizing Committee of ICRTCSE 2016
doi:10.1016/j.procs.2016.05.122

lead and body phrase method and rule based method. Methods that use semantic based approach are as follows: multimodal semantic model, information item based method and semantic graph based method.

Ontology based summarization has recently emerged as a subfield of information extraction. Motivated by the definition of text summarization in natural language processing, ontology summarization is defined as the process of distilling knowledge from ontology to produce an abridged version for a particular user (or users) and task (or tasks). According to this definition, the information content of a summary depends on either user's needs or/and task's requirements. The paper is divided into sections. Section 2 describes an overview of ontology along with the reasons for developing ontology. Section 3 specifies the implications of ontology. Section 4 describes the related works. Section 5 specifies some of the methods for evaluating ontology. Section 6 concludes the survey.

2. Ontology: An Overview

Ontology is defined as a formal and explicit specification of a shared conceptualization. Generally, ontologies are defined for particular domains. Since information extraction is essentially concerned with the task of retrieving information for a particular domain, formally and explicitly specifying the concepts of that domain through an ontology can be helpful to this process. Ontology together with a set of individual instances of classes constitutes a knowledge base [3]. Classes are the focus of most ontologies. Classes describe concepts in the domain. For example, a class of wines represents all wines. Specific wines are instances of this class. A class can have subclasses that represent concepts that are more specific than the superclass. For example, we can divide the class of all wines into red, white, and rose wines. A concept can be referenced by several terms (for example: "computer science", "computing", "information technology" are synonyms) and a term can reference several concepts (for example the term "bank" can be used to reference a "river bank" or a "commercial bank"). The roles of linguistic ontologies are twofold: The first one is to present and define the vocabulary used. This is achieved by a dictionary which list all the terms actually used in language. Secondly, linguistic ontology is the result of a terminology agreement between users' community. This agreement defines which term is used to represent a concept in order to avoid ambiguity. This process is called vocabulary normalization. When a concept could be described by two synonym terms, the normalization process selects one of those to be the preferred label of the concept.

2.1. Reasons for developing Ontology

- Sharing common understanding of the structure of information among people or software agents is one of the goals in developing ontologies. For example, suppose several different Web sites contain medical information or provide medical e-commerce services. If these Web sites share and publish the same underlying ontology of the terms they all use, then computer agents can extract and aggregate information from these different sites. The agents can use this aggregated information to answer user queries or as input data to other applications.

- Enabling reuse of domain knowledge was one of the driving forces behind recent surge in ontology research. For example, models for many different domains need to represent the notion of time. This representation includes the notions of time intervals, points in time, relative measures of time, and so on. If one group of researchers develops such an ontology in detail, others can simply reuse it for their domains. Additionally, if we need to build a large ontology, we can integrate several existing ontologies describing portions of the large domain.

- Making explicit domain assumptions underlying an implementation makes it possible to change these assumptions easily if our knowledge about the domain changes. Hard-coding assumptions about the world in programming-language code make these assumptions not only hard to find and understand but also hard to change, in particular for someone without programming expertise. In addition, explicit specifications of domain knowledge are useful for new users who must learn what terms in the domain mean

- Separating the domain knowledge from the operational knowledge is another common use of ontologies. We can describe a task of configuring a product from its components according to a required specification and implement a program that does this configuration independent of the products and components themselves. We can then develop an ontology of PC-components and characteristics and apply the algorithm to configure made-to-order PCs.

- Analyzing domain knowledge is possible once a declarative specification of the terms is available. Formal analysis of terms is extremely valuable when both attempting to reuse existing ontologies and extending them.

3. Implication of Ontology

Ontology has been used in different domains for providing metadata of concepts and their relationships. This metadata can be used for ensuring interoperability among different systems, modeling contextual information, inferencing, reasoning and efficient searching of contents and resources. Till date, ontology has been used in almost every domain such as e-governance, medical science, chemistry, social sciences and agriculture. In knowledge engineering, ontology is used to share and reuse knowledge and as standard for communication between computer and man. Ontology is applied to various fields of computer as a conceptual modeling tool, and is used to organize information and manage knowledge. Ontology extension is used to add the new concepts and relationship into the existing ontology, which is a more complex task.

Many information extraction systems enable the recognition of keywords within documents (e.g. 'Rembrandt' is a 'Person', '15 July 1606' is a 'Date').But such information is of little use without acquiring the relation between the entities (e.g. 'Rembrandt' was born on '15 July 1606'). The Artequakt project [4] aims to implement a system that searches the Web and extracts knowledge about artists, based on an ontology describing that domain, and stores this knowledge in a Knowledge Base (KB) to be used for automatically producing tailored biographies of artists. Here an ontology is developed for the domain of artists and paintings.

A set of ontologies were developed in order to facilitate the integration of a variety of combinatorial, simulation and optimization models related to agriculture [5].In practice, building an ontology-based base is a valid and feasible method. The work is primarily divided into the following steps: Firstly, to build AgriOnto hierarchy. Secondly, to formalize the text knowledge on the basis of AgriOnto. Thirdly, to compile and check the knowledge. Finally, to build the knowledge-based service systems. AgriOnto indicates formal definition of agriculture and their relation. The definition and relations form an integrated hierarchy of agriculture.

Ontology technology is considered to be a highly suitable means of supporting educational-technology systems. The tree organization of a knowledge domain is an important property that can significantly reduce the processing, but it is insufficient to describe the rich network of relations that ties the concept structures. With the advancement of artificial intelligence technologies, ontology technologies enable a linguistic infrastructure to represent conceptual relationships between course materials. Ontologies help us to make the knowledge that is represented in learning content explicit.

Another example used in ontology aims to diagnose or predict diabetes in the earlier stage by using reasoner [6]. A reasoner is a piece of software that is able to infer logical consequences from a set of asserted facts or axioms. The reasoner utilizes symptoms ontology for diagnosing the same. The symptoms ontology is a domain specific ontology which helps to provide the possible meaningful factors which leads to diabetes. The research aims to discuss the way of construction of symptoms ontology for diabetes and utilization of same for diagnosing diabetes.

4. Related Works

In recent days, there has been an explosive growth in the volume of textual information available. Hence it is very important to present the data to the user in an abstract version. Summarization will make this process easy.

Ontology based summarization methods involve reduction of sentences by compressing and reformulation. These methods use linguistic and natural language processing techniques.

Thanh Tran and Philipp Cimiano [7] presented an approach for interpreting keyword queries using background knowledge available in ontologies. Based on a few assumptions about how people describe their information needs, an approach was presented which translates a keyword query into a DL conjunctive query which can be evaluated with respect to an underlying knowledge base (KB). One major problem the approach suffers from is the fact that it does not consider that keywords can be ambiguous with respect to labels in the ontology and simply considers the first matching ontology element to start the exploration.

In paper [8] the authors address the issue of identifying the concepts in an ontology, which best summarize what the ontology is about. A number of criteria were jointly considered, and correspondingly a number of algorithms were developed and linearly combined, to identify key concepts of an ontology. The criteria include: name simplicity which favors concepts that are labeled with simple names while penalizing compounds; basic level which measures how "central" a concept is in the taxonomy of the ontology; density highlights concepts which are richly characterized with properties and taxonomic relationships; coverage aims to ensure that no important part of the ontology is neglected; and popularity identifies concepts that are commonly used. The summarization results, i.e. key concepts, were evaluated against human assessors' summaries, referred to as ground truth.

Nesrine Ben Mustapha [9] introduced a comprehensive framework for building a domain-specific ontology. Two methods for ontology acquisition were applied in order to create the domain ontology. The first was to create a small domain-specific core ontology from scratch and then apply a focused web crawler to this ontology in order to retrieve domain related web pages and interesting domain terms for extending the knowledge base. The second acquisition approach takes a well-established thesaurus as a basic vocabulary reference set and converts it to an ontology representation. Then a domain specific and a general corpus of texts were used in order to remove concepts that are not descriptive for the domain.

Xiang Zhang [10] proposed a novel approach to automatic ontology summarization based on RDF Sentence Graph. Summaries are customizable: users can specify the length of summaries and their navigational preferences. The authors compared five different centrality measurements in assessing the salience of RDF sentence and defined a reward-penalty re-ranking algorithm to make the summaries comprehensive. The evaluation showed that weighted in-degree centrality measures and several eigenvector centralities all have good performance in producing qualified summaries after re-ranking. Shown by the experiments, the approach of ontology summarization was feasible and promising.

Many researchers have made effort to use ontology to improve the process of summarization. Most documents on the web are domain related because they discuss the same topic or event. Each domain has its own knowledge structure and that can be better represented by ontology. The fuzzy ontology with fuzzy concepts is introduced for Chinese news summarization [11] to model uncertain information and hence can better describe the domain knowledge. In this approach, first the domain experts define the domain ontology for news events. Next, the document preprocessing phase produces the meaningful terms from the news corpus and the Chinese news dictionary. Then, term classifier classifies the meaningful terms on the basis of events of news. For each fuzzy concept of the fuzzy ontology, the fuzzy inference phase generates the membership degrees. A set of membership degrees of each fuzzy concept is associated with various events of the domain ontology. News summarization is done by news agent based on fuzzy ontology. The benefit of this approach is that it exploits fuzzy ontology to handle uncertain data that simple domain ontology cannot. This approach has several limitations. First, domain ontology, Chinese dictionary and news corpus has to be defined by a domain expert which is time consuming. Secondly, this approach is limited to Chinese news, and might not be applicable to English news

Mithun and Munirathnam [12] presented a semi-automatic development of an ontology library for the topics defined in the National Intelligence Priorities Framework (NIPF). They use Jaguar-KAT, a state-of-the-art tool for knowledge acquisition and domain understanding, with minimized manual intervention to create NIPF ontologies loaded with rich semantic content. Jaguar automatically builds domain-specific ontologies from text. The text input to Jaguar can come from a variety of document sources, including Text, MS Word, PDF and HTML web pages, etc. The ontology/knowledge base created by Jaguar includes ontological concepts, hierarchy and contextual knowledge base.

5. Methods for Evaluating Ontology

Ontology evaluation approaches [18] are unevenly distributed in two major categories. On one hand, a few principled approaches exist that define a set of well-studied, high level ontology criteria to be manually assessed (e.g., OntoClean ,Ontometric). On the other hand, the use of ontology evaluation in the context of ontology learning has led to the development of automatic approaches that cover different evaluation perspectives and levels. Evaluation levels refer to the aspects of the ontology that are evaluated (e.g., labels, conceptual structure). Perspectives are defined by what is considered to be good "quality" ontology. Table 1 describes the goal and description of the various methods for evaluating ontology.

Table 1.Methods for evaluating ontology

Method	Goal	Description
OntoMetric	Helps to choose the appropriate ontology for a new project	Compares the importance of the project objectives and study the characteristics of the ontologies Gets for every candidate ontology a quantitative measure of its suitability
Natural Language Application metrics	Helps evaluate the content of ontologies with respect to various metrics:	
	1. Precision and Recall metrics	Measures for each ontology (a) how many identified items are correct and (b) how many items that would have been identified are effectively identified
	2. Cost-based evaluation metrics	Characterizes the performance in terms of the cost of errors or the value of correct things
	3. Tennis measure	Gives a measure of the "fit" between an ontology and a corpus (domain knowledge) by using a vector space model of instances (terms)
	4. Lexical Comparison level measure	Compares the contents of two ontologies without considering their conceptual structure
OntoClean	Helps evaluate a formal ontology	Cleans the taxonomical structure of ontologies Compares the ontology vs. a predefined ideal taxonomical structure to detect inconsistencies
EvaLexon	Helps evaluate ontologies created by ontology	The method stays at the linguistic level. Compares the vocabulary of the triples mined with the input text as such and with a set of words considered to be relevant for that text.

6. Conclusion

Abstractive summarization methods produce highly consistent information and less redundant summary. Generally, abstractive summarization is a challenging area because of the complexity of natural language processing. Many works are being carried out in the field of abstractive summarization especially by making use of ontology in various domains. Now we should consider the possibility of accommodating all these explicit domains in a single platform to build a robust, extensible summarization system which will enable us to get abstractive summary from different domains. This study examines a review on ontology based abstractive summarization methods and its importance in different domains. Some of the methods for evaluating ontology are also specified. Certainly, this study can be adapted in a way that new researchers to the area of text summarization can get a better understanding on ontology based approaches.

References

1. Atif Khan, Naomie Salim," A review on abstractive summarization methods," in Journal of Theoretical and Applied Information Technology,2014 Vol. 59 No.1
2. Kallimani, J. S., K. G. Srinivasa, R. B. Eswara. "Information Extraction by an Abstractive Text Summarization for an Indian Regional Language." In: Proceedings of IEEE NLP-KE, Tokushima, Japan, 27-29 November 2011.
3. Chris Nowak. "On ontologies for high-level information fusion". In Proceedings of the Sixth International Conference on Information Fusion.Cairns,2003
4. Harith Alani, Sanghee Kim, David E. Millard, Mark J. Weal ,Wendy Hall, Paul H. Lewis, Nigel R. Shadbolt," Automatic Ontology-based Knowledge Extraction and Tailored Biography Generation from the Web", IEEE Intelligent Systems, 18(1), pp. 14–21.
5. Aqeel-ur-Rehman, Shaikh ZA," ONTAgri: scalable service oriented agriculture ontology for precision farming," In International Conference on Agricultural and Biosystems Engineering p: 1-3.
6. A.Jaya, "A Standard Methodology for the Construction of Symptoms Ontology for Diabetes Diagnosis" International Journal of Computer Applications (0975 – 8887) Volume 14– No.1, January 2011
7. Thanh Tran, Philipp Cimiano, Sebastian Rudolph and Rudi Studer," Ontology-based Interpretation of Keywords for Semantic Search" Institute AIFB, Universität Karlsruhe, Germany
8. Peroni, S., Motta, E., d'Aquin, M.: "Identifying Key Concepts in an Ontology Through the Integration of Cognitive Principles with Statistical and Topological Measures". In: 3rd AsianSemantic Web Conference, Bangkok, Thailand (2008)
9. Nesrine Ben Mustapha, Hajer Baazaoui Zghal, Marie-Aude Aufaure, and Henda Ben Ghezala," Combining Semantic Search and Ontology Learning for Incremental Web Ontology Engineering" National School of Computer Sciences, University of Manouba, 2010 la Manouba, Tunisia
10. Zhang, X., Cheng, G., Qu, Y.: "Ontology Summarization Based on RDF Sentence Graph". In: 16th Inter. World Wide Web Conference Banff, Alberta, Canada, May 8-12 (2007)
11. C.-S. Lee, et al., "A fuzzy ontology and its application to news summarization, "Systems, Man, and Cybernetics, Part B:Cybernetics, IEEE Transactions on, vol.35, pp. 859-880, 2005.
12. Mithun Balakrishna, Munirathnam Srikanth," Automatic Ontology Creation from Text for National Intelligence Priorities Framework (NIPF)", Lymba Corporation Richardson, TX, 75080, USA
13. Fernandez M, Gomez-Perez A, Juristo N, "METHONTOLOGY: From Ontological Art to Towards Ontological Engineering" In Proceedings of AAAI97 Spring Symposium Series,Workshop on Ontological Engineering, 1997; 33-40
14. P. Kalaivani A. Anandaraj K. Raja "An ontology construction approach for the Domain of poultry science using protégé" International Journal of Information Technology and Management Sciences / Volume 1, Issue 2, 2011
15. Lei, Y., Uren, V., Motta, E.,"Semsearch: A search engine for the semantic web," In Proceedings of the 15th International Conference on Knowledge Engineering and Knowledge Management (EKAW). (2006) 238–45 Chris Nowak. "On ontologies for high-level information fusion". In Proceedings of the Sixth International Conference on Information Fusion.Cairns,2003
16. H. Alani, S. Dasmahapatra, N. Gibbins, H. Glaser, S. Harris, Y. Kalfoglou, K. O'Hara and N. Shadbolt,"Managing Reference: Ensuring Referential Integrity of Ontologies for the Semantic Web," Proc.13th Int'l Conf. Knowledge Eng. and Knowledge Management (EKAW'02)
17. M. Vargas-Vera, E. Motta, J. Domingue, S. Buckingham Shum, and M. Lanzoni, "Knowledge Extraction by using an Ontology-based Annotation Tool," Proc. First Int'l. Conf. on Knowledge Capture, (K-CAP'01), Workshop on Knowledge Markup & Semantic Annotation, Victoria, B.C.,Canada, 2001.
18. Brank, J., Grobelnik, M., Mladenic., D.: A Survey of Ontology Evaluation Techniques. In:Conference on Data Mining and Data Warehouses (SiKDD), Ljubljana, Slovenia (2005)
19. D. R. Radev, et al., "Introduction to the special issue on summarization," Computational Linguistics, vol. 28, pp.399-408, 2002.
20. D. Das and A. F. Martins, "A survey on automatic text summarization," Literature Survey for the Language and Statistics II course at CMU, vol. 4, pp. 192-195, 2007.
21. P.E. Genest and G. Lapalme, "Framework for abstractive summarization using text to-text generation," in Proceedings of the Workshop on Monolingual Text-To-Text Generation, 2011, pp. 64-73
22. N. Guarino," Formal Ontology in Information Systems," In N. Guarino, editor, Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, June 6-8, 1998,