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