Automatic Organization and Generation of Presentation Slides for E-Learning

K. Sathiyamurthy, Anna University Chennai, India
T. V. Geetha, Anna University Chennai, India

ABSTRACT

The effectiveness of an e-learning system for distance education to a large extent depends on the relevancy and presentation of learning content to the learner. The ability to gather documents on a particular topic from the web and adapt the contents of the document to suit the learner is an important task from the content creation perspective of e-learning. For the developer of e-learning material the provision to automatically extract, organize, and present content material would improve its effectiveness. This paper proposes to extract information from documents using language processing techniques and organizing the content into appropriate presentation slides for learning purposes using domain ontology and learning oriented pedagogy ontology.

Keywords: Articulation, Language Processing, Ontology, Pedagogy, Segmentation

1. INTRODUCTION

The rapid growth of web based educational materials demands e-learning systems for distance education to generate suitable learning content pertaining to certain topics automatically. The quality of content creation is proportional to the relevant content produced for a topic in a manner suitable for instructional mode of E-learning. Large amount of information in the web exists in natural language form. This information has to be organized into a suitable form for effective use by the learner. Instructors and Learners of E-learning can be provided with crisp and point wise summarization of text documents if relevant information is grouped under separate topics and the document can be divided into appropriate presentation slides. This technique has a wide scope in e-learning scenario for the tutor, learner and also for the administrator since presentation slides can be created for certain topics dynamically. In this work each generated slide is typically a segmented and summarized output of a particular section of the document and the procedure consists of layout structure identification of a given document (portable document format) then performing segmentation, assigning weight to sentences for summarization and title determination. The proposed work utilizes ontology and graph based
approaches for content creation for distance education in e-learning.

Ontology formally describes shared meaning of the used vocabulary. Ontology plays a very important role in text summarization and slide generation as it brings in a semantic flavor to document processing. A work on usage of three dimensional semantically structured spaces for describing e-learning materials using content, context and structure ontology was proposed by Stojanovic, Staab, and Studer (2001). In their work content or domain ontology is used in the process of defining the content of a learning material or in the process of searching a particular material. Context or pedagogy ontology describes domain independent components of learning such as example, introduction, figure, etc., that discriminate between the different parts of the learning material and help in context-relevant searching of learning materials. They have also described a structure ontology that is used to tailor the scattered chunks of e-learning contents to suit the learner needs. The proposed work in this paper uses domain ontology based on ACM classification (1998) to generate presentation slides restricted to computer science domain and pedagogical ontology having different pedagogical roles.

This paper is organized as follows. The next section explains related work done for content generation and graph based approaches. Section 3 explores the essential Language processing techniques required for slide generation. Section 4 describes the domain and pedagogical ontology and Section 5 illustrates the graph technique used in the work. The procedure for bullet point's identification for slide generation is described in Section 6, and Section 7 deals with the evaluation measures followed by results and discussion. Section 8 deals with conclusion and future direction of this work.

2. RELATED WORK

In this section, we briefly review the work on content generation and the techniques involved in document processing and graph based approaches. A work on document summarization and information extraction for generation of power point slides was proposed with focus on generation of slides but which however did not consider domain knowledge (Prasad et al., 2009). Slide generation with semantically annotated documents having a predefined GDA (Global Document Annotation) tag set was attempted by Maso and Koti (1999). In their work slides were generated from predefined annotation of documents with XML tag set for the machines to infer the semantic structure of the document. A work on separation of content from its presentation logic and its storage in an XML based content object format was proposed by Bhutoria and Mishra (2005).

Generated slides are typically a reformation of the particular document. The reformation procedure generally consists of performing document preprocessing tasks such as tokenization, segmentation and summarization. Text segmentation is the process of dividing text into meaningful blocks. TextTiling algorithm developed by Hearst (1997) describes a paragraph-level model of discourse structure based on the notion of subtopic shift, and an algorithm for subdividing expository texts into multi-paragraph “passages” or subtopic segments. Text segmentation using lexical cohesion profile (LCP), based on similarity between words to locate segment boundaries was proposed by Hideki Kozima (1993). Text summarization is a product of electronic document explosion, and can be seen as the condensation of the document collection (Teng, Liu, Ren, & Tsuchiya, 2008). In slide generation, summarization is an important procedure as it is used to prepare condensed text required for bullet point identification.

Relation of co-occurrence of words matched with concepts in domain ontology enables identification of the topics of particular segment. Graphs based techniques enables capture of the semantic structure of the document by having words connected by edges representing the co-occurrence relation. Graphs are also more robust than typical vector representations.
Asynchronous Online Foreign Language Courses
[www.igi-global.com/chapter/asynchronous-online-foreign-language-courses/11745?camid=4v1a](www.igi-global.com/chapter/asynchronous-online-foreign-language-courses/11745?camid=4v1a)