Content-based Text Mining Technique for Retrieval of CAD Documents

余文德,許家洋
Construction Management
Architecture
wenderyu@chu.edu.tw

Abstract

The computer aided design (CAD) document provides an effective communication medium, a legal contract document, and a reusable design case for a construction project. Due to technological advancements in the CAD industry, the volume of CAD documents has been increased dramatically in the database of construction organizations. Traditional retrieval methods relied on textual naming and indexing schemes that require the designers (engineers and architects) to memorize in details the meta-information used to characterize the drawings. Such approaches easily overwhelmed the users' memory capability and thus caused low reusability of CAD documents. In this paper, a content-based text mining technique is adopted to extract the textual content of a CAD document into a characteristic document (CD), which can be retrieved with similarity matching using a Vector Space Model (VSM), so that the automated and expedited retrievals of CAD documents from vast CAD databases become possible. A prototype system, namely Content-based CAD document Retrieval System (CCRS), is developed to implement the proposed method. After preliminary testing with a CAD database with 2,094 Chinese annotated CAD drawings collected from two real-world construction projects and a public engineering drawing database, the proposed CCRS is proven to retrieve all relevant CAD documents with relatively high precision when appropriate query is specified. Finally, three search strategies are recommended for the users to narrow down search scope while a target CAD document is desired. It is concluded that the proposed content-based text mining approach provides a promising solution to improve the current difficulty encountered in retrieval and reusability of vast CAD documents for the construction industry.

Keyword：CAD, text mining, information retrieval, characteristic document, construction engineering.