URN etd-0407109-183708-568
Author Kao-Pin Huang
Author's Email Address No Public.
Statistics This thesis had been viewed 76 times. Download 30 times.
Department
Year 2007
Semester 1
Degree Master
Type of Document Master's Thesis
Language zh-TW.Big5 Chinese
Title Automatic Collection System for Quintessence Articles of Blogs
Date of Defense 0000-00-00
Page Count 57
Keyword
Information Retrieval
Blog
Document Classification
TF (Term Frequency)
IDF (Inverse Document Frequency)

Abstract
Because information science and technology progress continuously, more and more precious materials such as characters, image, video-information or music, multimedia are recorded. These materials have reflected the mankind’s accumulated professional field knowledge. Quite abundant articles and image contents have already been accumulated in every blog at present. However, we still lack of the automatically collecting system for quintessence articles. As for the user's demand of searching, because of the large amount of articles of blogs, one can only rely on inquiring the key word to find relevant articles, browse the content of the articles and disregard those irrelevant ones. An ordinary searching result might include a lot of garbage articles, for example, advertising articles. By using our automatic collection system for quintessence articles of blogs instead of a ordinary searching engine, one can easily find the relevant articles of the input keywords.

Managing articles of blogs is an important issue nowadays. In this study, we use the term frequency and inverse document frequency for the article classification, to reach the effective categorizing results. In addition, this system provides the function of searching both articles and pictures. The experimental results demonstrate that the quintessence articles are distinguished from the garbage articles. It helps readers to find interesting articles easily among tons of them.
Advisory Committee - advisor

Files etd-0407109-183708-568-001.pdf

indicate access worldwide
Date of Submission 2008-08-11