Automatic Appraisal System

Min-Ling Lai, BS, Polun Chang, PhD

Institute of Health Information and Decision Making, National Yang-Ming University, Taipei, Taiwan/ROC

Abstract

In Taiwan, we have systematically evaluated health web sites due to our national Health Web Award since 2002. Through appraisal of the field masters, recognized websites will be awarded by Label of Excellent. It's also the way to provide identified credit for general public. This study intends to further digger the possibility of automating the process mentioned above.

Keywords：text mining

Introduction

People in Taiwan are getting used to read health and medical related information on line, but they may not be capable to judge the accountability of gotten information. To lift up the quality involved in health information for people, and to encourage providers to enhance qualitative content, the government initiated annual contest for Health Information Website. Through appraisal of the field masters, recognized websites will be awarded by Label of Excellent. It's also the way to provide credit to be identified by general public. This study intends to further digger the possibility to automate the process mentioned above.

Methods

First, to collect web pages, which are to provide trained and tested function, and to utilize cutting words system, provided by Academia Sinica (http://ckipsvr.iis.sinica.edu.tw/), to cutting words on the web pages in order to calculate the weighting of feature.

Second, term selection and Extraction.

Third, design a categorizing tool by SVM, support vector machines, and by means of result, produced by first step, to classify whether health information web pages are related.

Fourth, creating a crawler to catch those web pages, which cover 'tw' in the last two words of their web addresses, automatically.

Fifth, using the categorizing tool to subgroup those web pages, filtered by the crawler, in order to get health information websites.

Sixth, initiating a phase pool system to give points to those health information websites according to scoring standard of annual contest, such as whether footnotes, creator and posting time included, are provided in the article.

Finally, comparison of grading difference between manual and the system.

Program is coded by VB 6.0, database by SQL2000.

Results

Steps one to three are almost accomplished, and steps four to six are incomplete.

Fig.1 the cutting words system to calculate the weighting of feature.

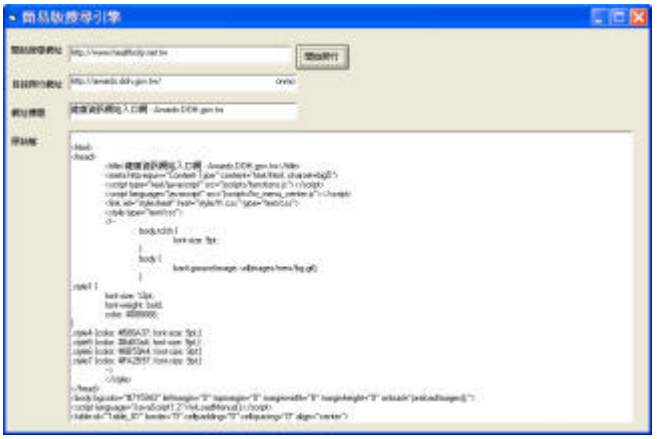

Fig.2 the crawler to catch web pages.

Discussion

Once the result of the system is accountable, it may be established broadly in public health institutions. Keep on searching and apprizing health information websites on the internet, and to be reference for people to select high profile websites. If evaluation being poor, public health institutions may review the content of web pages voluntarily and find out inappropriateness to fix it.