Developing an Automatic Accreditation Support System of health websites
Min-Ling Lai, BS, MS, Polun Chang, PhD
Institute of BioMedical Informatics, National Yang-Ming University, Taipei, Taiwan/ROC

Abstract
The annual national quality and accreditation label awarding approach has been used to identify quality health websites since 2002 in Taiwan. Since the accreditation process has resumed great amount of professional manpower every year, this study was designed to simplify the process by using the text categorization and information extraction approach.

Background
The annual national quality and accreditation label awarding approach [1] has been used to identify quality health websites since 2002 in Taiwan. The accreditation process has been well designed, implemented, and standardized in 2004. Each enrolled website was evaluated by at least 10 healthcare IT and clinical professionals manually at two stages using up to six categories of 50 items of criterions, such as authorship, currency, objective, navigation, etc. The top 10% ranked websites would be awarded an Excellent Quality label.

Among these categories, which are (1) website structure, (2) website basic information and privacy policy, (3) website information appropriateness, (4) website content updating, (5) website management, and (6) website information accuracy/completeness/reliability, some IT-related categories and more technical criterions appear being able to be examined by the support of systems. Since the current process has resumed great amount of professional manpower every year, this study was designed to simplify the process by supporting the accreditation process using the text categorization and information extraction approach [2-4].

Methods
The system was designed by first random sampling some web pages from the health information websites, examining whether the content meets the criterions, and giving rating according to results. Two key steps are, first to judge whether candidate websites are health-related, and second to calculate proportion of fitted pages to total pages. Automate accreditation system is designed as followed:

1. Creating a crawler to catch those web pages, which cover ‘tw’ in the last two words of their web addresses, automatically, and then using Text Categorization to filter health information websites.

2. Evaluating by Information Extraction to selected pages. For instance, judge whether the pages are remarked by author and posted time first, then count proportion of fitted pages to total.

3. Comparing between original manual process and new model.

Result
Two representative screenshots are shown in Figure 1 and 2.

Figure 1 the crawler to catch web pages. Figure 2 the preprocessing of Text Categorization: the cutting words system to calculate the weighting of feature.

Discussion
The pilot system has been developed and once the testing is done, citizens may input web address into the system, the system will calculate scores by Award KPIs, and a report will be output as their reference. In addition, the system can be installed into related government institutes to collect emerging pages and give scores at all time. General public may review the end result of latest evaluation as their reference to choose their favor.

References