Yuanpei Zhang

EE322 Homework 1

Automatic Keywords Extraction Program

A potential project that I am interested in is an Automatic Keywords Extraction program, which is a program being able to recognize representative words and phrases in electronic articles automatically. By using keyword extraction algorithms such as TF-IDF, the program can efficiently determine the keywords in text documents.

Currently there are more than 2500 million Internet users worldwide, and Internet is becoming the major source for people to obtain information. Facing a huge amount of information provided in the form of text on Internet, the users always try to assimilate information on a topic of interest from multiple information sources. By using keyword search engine in database, Internet users are able to quickly find the documents that are relevant to the topics they need. However, most current database search engines are working on the keywords that are manually extracted from documents. In order to quickly and accurately find out the keywords in documents and map them to existing database content, the Automatic Keyword Extraction Program would be greatly favorable.

The Automatic Keyword Extraction Program can also be used in searching local documents stored on user’s own computers. By extracting the keywords from files, especially from the ones in the format of text, word, and PDF, it enables users to search their local documents by keywords and their relevant words instead of the traditional exact-match searching.

The potential market of this product includes all PC users. It can be easily installed on any operating system, and overcome the shortage of the pre-installed searching function. Besides PC users, the customers can also be Internet database and news companies such as ProQuest Database, Yahoo News. The Automatic Keyword Extraction Program provides such companies advantages in organizing their online data.

I pledge my honor that I have abided by the Stevens Honor System.