Enhancing Web Based Information Retrieval With Regular Expression Based Pattern Matching

Christopher M. Frenz  
Department of Computer Engineering Technology

The present era of human history is often referred to as an information age since economies have greatly increased their focus on the generation, collection, and analysis of information. As the technologies that enable such information processing tasks continue to increase in pervasiveness and power, we are witnessing an explosion of information availability. Whereas, in the past finding information often proved to be challenging, today it is more often a problem of finding too much information and needing to separate out the information that is relevant from that which is not. One can witness this phenomenon first hand by simply conducting a search at Google or a like search engine, where searches will generally return tens to hundreds of thousands of hits. This talk will demonstrate a method of search refinement that allows searchers to specify text patterns and use those patterns to filter through potential documents of interest. Text patterns can often make the retrieval of information easier than using keywords, since certain types of information such as a generic phone number can be easily specified by a pattern similar to (XXX) XXX-XXXX, as compared to trying to express an ability to match any given phone number using keywords alone. Applications of this methodology will be demonstrated in the areas of mining biological mutation data, determination of gene/protein expression in biological tissue types, and in the extraction of C/C++ code from Web resources.