Crawler:

* (a)  Uploading files can be tedious. You want to automate this process somehow. One way of doing this is to make a web crawler. A web crawler is a crawler that traverses all web pages that it can find starting from a certain set of seed values (another set of hyperlinks). For example, if I were to provide my seed value as http://wwww.utsc.utoronto.ca, then your web crawler starting from this seed value must extract all data that it finds on the seed value and branch of to other links starting from the seed value and continue extracting data. Of course, there must be some stopping criteria such as the depth level? that tells the crawler to stop crawling. The data extracted from the crawler will be used for building the index.
* (b)  Again, the crawler is something that must be initiated by a very special user (admin?) of your software. Also, you must think about how frequently this must be done?