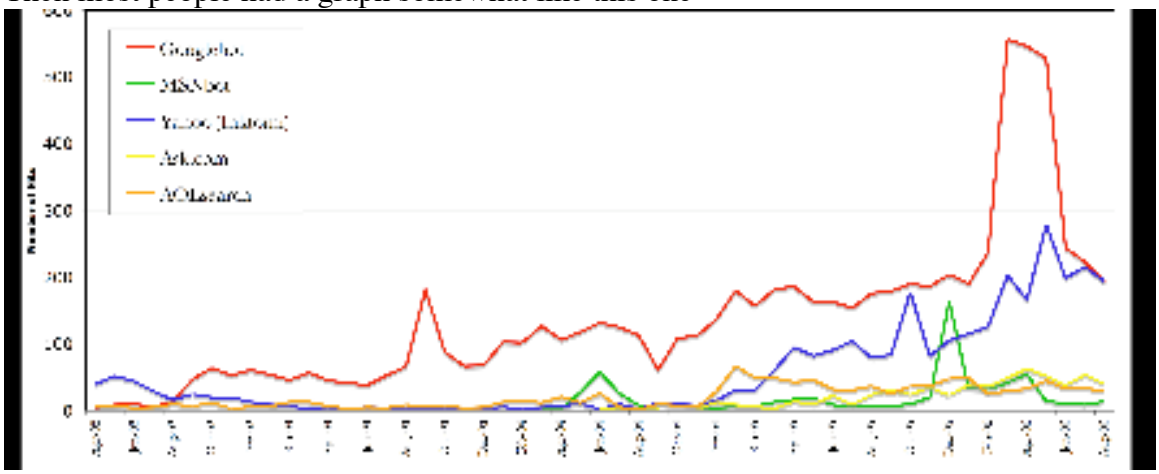
Generally, given the open-endedness of the assignment it is not surprising that people went in lots of different directions. Almost everyone answered the three questions I posed. Almost everyone reported starting by spending time figuring out just what was in the access log file. Almost everyone complained about the size of the file and how long it took to do anything.

Somewhat surprising is that only one person reported trying to use any public utilities to analyze this file. (And that person reported essentially failure) Dozens of program are available. For laughs I grabbed one and ran it.

Most people fond 8-16 crawlers by looking at robots.txt accesses and then eliminating the ones the rare ones. A typical quote was "Google, Yahoo, MSN search, Lycos, Lexis-Nexis, AlltheWeb, AOL search, and Ask.com. There were other requests for robots.txt (from, presumably, "friendly" web-spiders), but they were unfamiliar and/or not identifiable except for an IP address. For the purposes of this assignment, requests from the unidentifiable spiders were ignored."

Then most people had a graph somewhat like this one



to show the number of accesses by various crawlers as a function of time. Many noted and or tried to look for an increase in about early 2006 which they attributed to the announcement of the Microsoft grant.