12.G: NETWORKED PROGRAMS (GLOSSARY)

Contributed by Chuck Severance
Clinical Associate Professor (School of Information) at University of Michigan

BeautifulSoup

A Python library for parsing HTML documents and extracting data from HTML documents that compensates for most of the imperfections in the HTML that browsers generally ignore. You can download the BeautifulSoup code from www.crummy.com.

port

A number that generally indicates which application you are contacting when you make a socket connection to a server. As an example, web traffic usually uses port 80 while email traffic uses port 25.

scrape

When a program pretends to be a web browser and retrieves a web page, then looks at the web page content. Often programs are following the links in one page to find the next page so they can traverse a network of pages or a social network.

socket

A network connection between two applications where the applications can send and receive data in either direction.

spider

The act of a web search engine retrieving a page and then all the pages linked from a page and so on until they have nearly all of the pages on the Internet which they use to build their search index.