REAL WORLD APPLICATIONS

World Wide Web Robots

Finding information on the World Wide Web can be challenging. A search engine is often used to find Web sites related to the topic of interest. Search engines, which can be found at a number of different World Wide Web addresses, allow you to look for Web pages that are related to a certain topic. For example, if you typed the word “chemistry” into a search engine, it would return a list of links to pages related to chemistry.

There are millions of pages on the World Wide Web, so how do search engines quickly sort through them all and show you the ones related to your topic? Search engines accomplish this by making huge indexes of millions of Web pages. The indexes contain information about the Web sites. That way, when you type a query, or keyword search, into the search engine, the search engine can simply go to that word in its index and tell you what Web pages are related to it.

Crawling the Web

How are these indexes compiled? Many of them are created using Web robots, which are also known as Web Crawlers, Web Wanderers, Web Worms, and Web Spiders. These robots are actually just computer programs that have been designed to search the Web and to compile information.

A simple Web robot might be designed to follow every link on a page, and all of the links from those pages, and so on. For every page, specific information would be recorded. The programs for most Web robots are much more complicated, allowing them to follow and record only certain links, instead of every one. The differences in the way the robots are programmed is one reason that different Internet search sites (which rely on different robots for their information) return different Web pages in answer to the same query.

Your Turn to Think

1. Name some search engines you have used on the Web to find information.
2. What is a Web robot?
3. Why would different search engines use robots with different programs?