

Homework Assignment 8

Exercise 8.1

What is a Web crawler and what is its basic mode of operation?

Exercise 8.2

Why do large-scale crawlers maintain a local DNS component?

Exercise 8.3

Why do crawlers need to check for duplicate URIs and how do they do it?

Exercise 8.4

How does shingling work and what problems does it solve?