Homework Assignment 9

Exercise 9.1
What is a Web crawler and what is its basic mode of operation?

Exercise 9.2
What features does a good Web crawler have?

Exercise 9.3
Why do crawlers need to check for duplicate URIs and how do they do it?

Exercise 9.4
Why do crawlers need to check for duplicate content? How does shingling work and what problems does it solve?

Exercise 9.5
What do you need focused crawling for and how does it work?