1. Use a container to store the already crawled url, when begin to crawl, compare the url with the urls in the container, if it already crawled, jump to next one
2. We can use multithreading to improve the efficient of the web crawler. We can import java.util.concurrent package, build a thread pool to manage thread, and each thread starts its own task