HW: Web Crawling

Please note: even though the homework assignments are optional, you are encouraged to answer them, as they will help you prepare for your final exam. In case of questions reach out to email: usman@ifis.cs.tu-bs.de

1.-What's the purpose of large scale web crawlers? And how do they differ from focused crawlers?
2.-Why do crawlers need to check for duplicate URIs and how do they do it?
3.-How does shingling's basic idea help in detecting near duplicates?