

## Chapter No. 19 Web Search Basic

## &lt;Food for Thoughts&gt;

1. Explain why this estimator would be very difficult to use in practice.
2. What are the some of the general type of queries on the web? Give an example of each type of a query.
3. Define what is Shingling? What is it usage in web crawler?
4. Define duplicate and near duplicate web pages?

## Chapter No. 20 Web Crawler

## &lt;Food for Thoughts&gt;

1. What are some of the should have features in a web crawler? Illustrate them.
2. Why a crawler must be robust and polite in extracting information from the web?
3. Two web search engines A and B each generate a large number of pages uniformly at random from their indexes. 30% of A's pages are present in B's index, while 50% of B's pages are present in A's index. What is the number of pages in A's index relative to B's?
4. What are some of the critical components of a web crawler? Explain