CS 8080 - Information Retrieval techniques

312217104035

S.Dhinesh

I affirm that I will not give or recieve any unauthorised help on this exam, that all work will be my own

S.Dhinesh

1.

a. Centralized crawler-indexer architecture.

   The most common type of search engines is centralized crawler-indexer architecture. Crawlers are programs (software-agents) that traverse the Web sending news or updated pages to a main server where they are indexed. Crawlers are also called robots, spiders, wanderers, walkers, and know bots. In spite of their name, a crawler does not actually have to and run on remote machines, rather the crawler runs on a local system and sends requests to remote web servers. The index is used in a centralized fashion to answer queries submitted from different places in the web. The following figure shows the software architecture of a search engine based on the users, consisting of the user interface and the query engine and another that consists of the crawler and indexer modules.