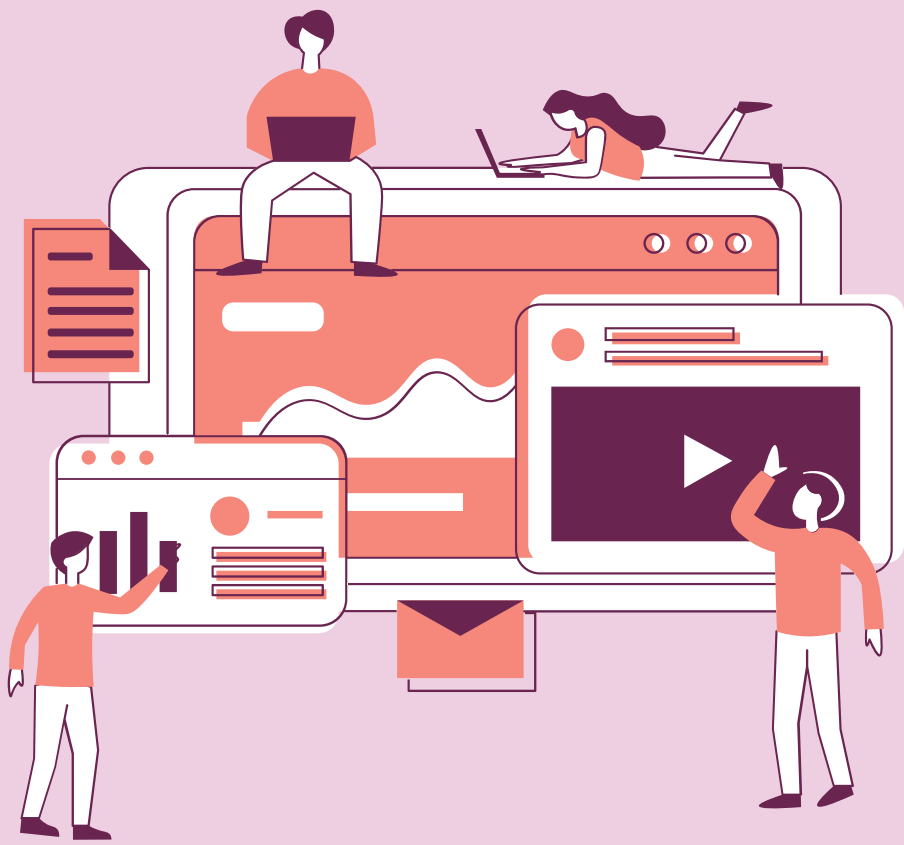# SpiderX

The primary function is to gather information from websites to facilitate various applications such as search engines, data mining, and web archiving.

Politeness Policy

- It adhere to politeness policies to avoid overloading web servers.

Traversal Logic

- This typically involves retrieving the HTML content of a web page, extracting hyperlinks, and recursively visiting these links while adhering to specified constraints

Data Storage

- As the it visits web pages, it collects and stores various types of data such as HTML content, metadata, and links

Multithreading

- To improve efficiency and speed, it often employ multithreading, allowing multiple pages to be crawled simultaneously. Each thread operates independently, following the same traversal logic.

BINUS UNIVERSITY

STEVEN GM
2702398283