Q. Write a design document on how to crawl a website site and save the images and text from the website. Write your thoughts on how you will approach the problem along with components you need.

**Crawl a website**

* Crawling a website is an automated process of getting webpages and indexing the URL of a site on web. The crawl searches the content of a page looking for links to next page to get and index the URL.

**Approaches to Crawl a website**

* **Ready made search:** There are some ready to use software to crawl a website which is cost – effective, no huge upfront charges and it may not suites all the time it is based on the domain of site we choose

**Example:** Mozenda, ScrapeBox, etc.

* **Using website APIs:** As the name implies it is an interface and communicates between two software programmes. I permit or deny requesting information from the main program. It is speed of exchange requests and responses and two – way communication. It is also high cost during development and unfriendly for programmers.
* **Use your own WebCrawler:** It is suggested when you can put time and money specially for the web crawl. So, most of times it is not suggested, but it will customize where other two approaches cannot give and friendly to coding skilled people.

**Saving images and text from the website**

* Firstly, open your friendly browser.
* Search for the data or information you want.
* Click the images of that information.
* Right click on the image and select save as
* Select the location the image to be saved and click on save which is at right bottom.
* Finally, drag the cursor on the text you and paste in the blank sheet on your desktop.