Assignment – Web Scrapping

Web Scrapping

      Web scraping, web harvesting, or web data extraction is data scraping used for extracting data from websites. Web scraping software may access the World Wide Web directly using the Hypertext Transfer Protocol, or through a web browser. While web scraping can be done manually by a software user, the term typically refers to automated processes implemented using a bot or web crawler. It is a form of copying, in which specific data is gathered and copied from the web, typically into a central local database or spreadsheet, for later retrieval or analysis.

Perform following steps:

- Visit website: **bbc.com/urdu**

- You will find many categories such as:

<div dir="rtl">

سائنس   فن فنکار   کھیل   ورلڈ   پاکستان

</div>

- Visit each category page and extract 100 stories for each category. (story appears when you click any news or "mazeed parhye")

- Each page contains 10 news headlines so iterate over pages to get 100 extract stories.

- Store all stories category-wise in **".csv"** file as follows:

Story	Label
...وزیر اعظم پاکستان عمران خان نے اتوار کے روز فی	pakistan
...افغانستان کے حکام کے مطابق سنیچر کو دارالحکومت	pakistan
...پاکستان پیپلز پارٹی کے چیئرمین بلاول بھٹو زردا	pakistan
...خوش قسمت ہے وہ سرکار جسے نئے خیالات سے عاری پھر	pakistan
...کیپٹن صفدر کی گرفتاری اور رہائی سے لے کر کراچی	pakistan

Submit ONLY .zip file containing all java files and dataset (.csv).