LITERATURE REVIEW:

Web scraping is a technique for automatically extracting data from a website. To develop customized scrapers, there are numerous frameworks and Application Programming Interfaces and also some configurable ready-to-use scraping tools are available.‘‘Web scraping: Applications and tools,’’[8] gives idea about different types of web scraping techniques.Glez-Pea et al. [9] and Haddaway [11] provide detailed overviews of frameworks and tools for various extraction tasks.

The World Wide Web, currently present a huge amount of information in different formats and from different origins.Moreover, in many circumstances, this is the only source of information available to public. Nonetheless, extracting data from websites is tough, with one of the most difficult difficulties being automatically recognizing ‘‘the appearance of items of interest and their features on the web pages and saving them in a database in a uniform manner" [3] on websites.

An important decision when developing a web scraping tool is, therefore, whether to develop one’s own application (see, for example, [6]) or to use an existing tool. ‘Legality and Ethics of Web Scraping’[7] describe whether web scraping is legal or illegal and the Ethics of the web scraping.

The goal of our paper is to scrape Bengali website and store the data in the database . The paper will give developers a quick overview of the challenges of scraping Bengali web pages and how they can be overcome. We used Python's Beautiful soup library to develop the scrapers for our demonstration tool.

### References

1. D. Glez-Peña, A. Lourenço, H. López Fernández, M. Reboiro Jato, and F. Fdez Riverola. Web Scraping Technologies in an API World. Briefings in Bioinformatics, 15 (5): 788-797, 2014. <http://doi.org/10.1093/bib/bbt026>
2. N. R. Haddaway. [The Use of Web-scraping Software in Searching for Grey Literature](https://www.researchgate.net/publication/282658358_The_Use_of_Web-scraping_Software_in_Searching_for_Grey_Literature). *Grey Journal (TGJ)*, 11 (3), 2015.
3. M. I. Varlamov and D. Y. Turdakov, ‘‘A survey of methods for the extraction of information from Web resources,’’ Program. Comput. Softw., vol. 42, no. 5, pp. 279–291, Sep. 2016.
4. R. Mitchell, Web Scraping with Python: Collecting Data from the Modern Web. Newton, MA, USA: O’Reilly Media, 2015.
5. <https://www.researchgate.net/publication/324907302_Legality_and_Ethics_of_Web_Scraping>
6. O. Castrillo-Fernández, ‘‘Web scraping: Applications and tools,’’ Eur. Public Sector Inf. Platform, Spain, Topic Rep. 2015/10, Dec. 2015. [Online]. Available: https://www.europeandataportal.eu/sites/default/files/ 2015\_web\_scraping\_applications\_and\_tools.pdf