CS-258

SOFTWARE ENGINEERING

Report on problem statement
Web Crawler and Scraper

Group G10

Team

cse210001070@iiti.ac.in
cse210001069@iiti.ac.in
cse210001067@iiti.ac.in
cse210001066@iiti.ac.in
cse210001079@iiti.ac.in

A web crawler is a program that automatically navigates the web by browsing links and downloading content from websites. The purpose of a web crawler can range from indexing the web for search engines to analyzing websites for specific information.

A web scraper is a program that extracts specific information from websites. It is used to extract data from websites for a variety of purposes such as data analysis, data migration, or to gather specific information for a database. Web scraping can be done manually or with the use of a web scraper program.

Both web crawlers and web scrapers are important tools for data extraction and analysis, and they are widely used in various industries. However, it's important to be mindful of ethical and legal considerations when using these tools, as some websites may restrict or prohibit automated scraping.

Types of Software Development Process Models

There are several types of software development process models, each with its own advantages and disadvantages. Some of the most commonly used models are:

Waterfall Model: This is a sequential model where each stage of development must be completed before the next stage can begin. It is best suited for projects with clear and well-defined requirements.

Agile Model: This is an iterative model that focuses on continuous delivery of working software and encourages collaboration between the development team and the customer. It is best suited for projects with rapidly changing requirements.

Spiral Model: This is a hybrid model that combines elements of the waterfall model and the agile model. It is best suited for projects with high levels of risk and uncertainty.

V-Model: This is a linear sequential model that uses a V shaped diagram to represent the different stages of development. It is best suited for projects with strict requirements and high levels of rigor and discipline.

Model for our development
We think the Agile model is the best method for developing Web Crawler and scraper.
Since we are newbies in this domain, we think this would be the best choice.This model is an iterative and flexible approach to software development that emphasizes collaboration, constant feedback, and the ability to adapt to changing requirements.

Incremental development: The team can work on small, manageable parts of the project at a time, rather than trying to complete the entire project at once. This allows for constant feedback and the ability to make adjustments as needed.

Regular sprints: Agile encourages regular sprints, which are short development cycles that allow the team to focus on a specific set of tasks and deliver tangible results.

Collaboration and communication: The Agile model places a strong emphasis on collaboration and communication, which is important for a team of newbies to ensure that everyone is on the same page and working towards a common goal.

Conclusion

Developing a Web crawler and scrawler is a complex and challenging task that requires a structured and well planned approach. The Agile Model is the best choice for this project because it makes sure we keep changing the project targets as our knowledge increases in this domain which allows for continuous improvement, encourages collaboration between our team. Overall, the Agile model can provide a supportive and flexible framework for a team of newbies to develop, helping them to deliver a high-quality product in an efficient and effective manner.