# Program Functionality Overview: Scraper + FAQ

## Purpose

The 'Scraper + FAQ' program is designed to:  
1. Scrape website content using a browser automation framework.  
2. Convert the scraped content into Markdown format.  
3. Generate a structured FAQ document using Google’s Gemini (Generative AI).

## Main Components

### 1. Web Scraping (using Crawlee & Playwright)

- Utilizes the PlaywrightCrawler from the crawlee library to scrape dynamic websites.  
- Likely configured to collect relevant textual content from webpages.  
- The crawling logic is defined asynchronously for performance and concurrency.

### 2. Markdown Conversion

- Converts HTML content to Markdown format using the markdownify library.

### 3. FAQ Generation

- The function `generate\_faq\_from\_markdown()` takes a Markdown file and uses Google Gemini (via the google.genai API) to generate a concise FAQ.  
- Output format:  
 ### Question 1  
 Answer 1  
  
 ### Question 2  
 Answer 2  
  
- The generated FAQ is saved to `storage/datasets/faqs/`. This will be accessible by agents.

### 4. Environment Configuration

- Sensitive credentials are managed using a `.env` file.

## Technologies Used

- Python 3  
- Playwright (via Crawlee): Web scraping  
- Google Generative AI (Gemini): FAQ generation  
- Markdownify: HTML-to-Markdown conversion  
- dotenv: Environment variable management