

Summary:

The program is a web crawler implemented in Go to fetch links from a specified root URL up to a given depth.

1. It uses the `UrlData` structure to store URL and depth information and a stack (`urlStack`) to manage URLs during crawling.
2. HTML parsing is done using the `golang.org/x/net/html` package, and links are extracted from anchor (`<a>`) tags.
3. The program handles different types of URLs, including absolute, relative, and parent directory references, ensuring valid and formatted URLs.
4. The crawling process is managed by the `CrawlWebpage` function, which initiates crawling, processes HTML content, and extracts links recursively.

Overview of architecture:

The architecture of the program involves the following components:

1. **UrlData Structure:** Stores URL and depth information for each crawled URL.
2. **parseLinks Function:** Recursively parses HTML content to extract and process anchor (`<a>`) tags, extracting URLs and pushing them onto the stack for further processing.

3. **processUrl Function:** Takes a current URL and a link, processes the link based on its type (relative, absolute, parent directory reference) to generate a final URL.
4. **parseBaseUrl Function:** Parses a given URL and returns the base URL by removing the path, fragment, and query components.

Breakdown of phases of project (e.g. building backend e.g.):

1. **Initialization:**
 - Defines the `UrlData` structure.
 - Imports necessary packages.
2. **URL Processing Functions:**
 - `processUrl`: Processes different types of URLs.
 - `parseBaseUrl`: Extracts the base URL from a given URL.
3. **HTML Parsing Function:**
 - `parseLinks`: Recursively processes HTML content to extract and process anchor tags.
4. **Crawling Logic:**
 - `CrawlWebpage`: Initiates the crawling process, manages a stack of URLs, fetches HTML content, and processes links.
5. **Error Handling:**
 - Logs errors during URL processing, HTML parsing, and HTTP requests.
6. **Output:**
 - Prints the crawled links with their corresponding index.