Tiki Crawl

A crawler tool for checking links or gathering content from websites.

This is the alpha version of a crawler tool for checking links or gathering content from websites. It relies on the Crawler library from Spatie. Kudo to those guys.

Introduction

Tiki-Crawl is a powerful tool that allows you to crawl and index external websites or content from remote sources. It enables you to create a local index of content from external sites, which can be useful for various purposes such as aggregating content, building a search index, or archiving information.

With Tiki-Crawl, you can specify a list of URLs or feeds from external website or remote sources that you want to index, and TikiWiki will periodically crawl these sources to fetch and index the content. This process is often referred to as web scraping and web crawling

What is Tiki-Crawl?

Tiki-Crawl is a web scraping and web crawling feature that enables users to fetch and index content from external websites or remote data sources. It automates the process of gathering information from these sources and can stores it.

Requirements

To use Tiki-Crawl effectively, ensure that you have this version installed (but it may very well be that it works on older versions):

- PHP: Tiki-crawl requires `php version 8.2.6` or higher.
- Node: Tiki-crawl requires `node version 18.16.0` or higher.

Installation

First you need to clone the tool from https://gitlab.com/tikiwiki/tiki-crawl.

This piece of code has been tested with `php 8.2.6` and `node v18.16.0`, make sure you have those version installed (but it may very well be that it works on older versions, code is not that complicated).

```
composer update npm install
```

Usage

First you need to configure your options. You can override any options from `config/config.default.php` by creating a `config/config.php` file where you will be adding new values that will override the one in `config/config.default.php`, for example: