

LU13a - Crawler Aufbau

Damit alle relevanten Seiten von DokuWiki (<https://wiki.bzz.ch>) heruntergeladen werden können, nutzen wir eine Kombination aus Collector, Downloader und Orchestrator.

WikiPageCollectorService

Der WikiPageCollectorService sammelt die URLs von den Seiten, die heruntergeladen werden sollen.

Der Benutzer soll einen Unterordner beziehungsweise einen „Namespace“ angeben können und sämtliche darunterliegenden Seiten sollen rekursiv gesammelt werden.

Leider gibt es dazu keine geeignete und funktionierende XML-RPC-Funktion. Daher nutzen wir dieselbe Funktionen, die auch vom Wiki verwendet wird, um Unterseiten aufzulisten.

```
<nspages . -h1 -exclude -simpleList -textPages="">
```

Ajax <https://wiki.bzz.ch/start?idx=>

WikiPageDownloaderService

WikiCrawlerPipelineService

Orchestriert wird das Ganze Filter

From:

<https://wiki.bzz.ch/> - **BZZ - Modulwiki**

Permanent link:

<https://wiki.bzz.ch/de/modul/ffit/3-jahr/java/learningunits/lu13/a?rev=1764628760>

Last update: **2025/12/01 23:39**

