

## Schnittstelle Crawler zu DataRefinery

### Datenaustausch

Es wurde sich für ein Dateibasiertes Protokoll entschieden. Das bedeutet das der Datenaustausch zwischen Crawler und DataRefinery über Dateien stattfindet. Die Daten liegen im JSON Format vor, da JSON valides JavaScript ist und sie somit einfacher manipulieren lässt und es native unterstützt wird. Jeder API Call generiert hierbei seine eigne Datei die spezifischen Daten enthält.

### Tauschordner

Pfad: /Projektordner/toRef/

```
const path = '../toRef/' + enhanceTarget(target)+customName + '.json';
```

### Namenskonvention

enhanceTarget(target) ist eine Funktion zur Namens Normalisierung, wobei target der Trigger für den preprocessor.js ist und die User Eingabe enthält. Jedes Leerzeichen wird durch ein underscore ersetzt mittels Regular Expression.

Bill Gates -> Billd\_Gates

apiCalls.js liefert den dementsprechende URL, um eine request an die Wikipedia API zu stellen sowie einen Tag für writer.js womit feststellbar ist aus welchen API Call die Daten stammen. writer.js schreibt den cn (für Custom Name) mit in den Dateinamen der erzeugten Datei. Ein cn besteht aus einem underscore „\_“ und dem Tag, was wiederum ein Kürzel aus dem API Call ist.

Hier beispielhaft anhand der getPageViewsURL():

getPageViewsURL() -> liefert letztendlich die Page Views -> „pv“.

Zusammen mit unserer Bildungsvorschrift:

ArtikelName + cn -> ArtikelName + „\_“ Tag -> ArtikelName + „\_pv“

```
//_pv = tag im dateinamen fuer pageviews
const cn = "_pv"
return [url, cn];
```