

## Input table

List of RSS / atom feeds in → csv file

The output is written to SQLITE db

e.g.

[https://www.tijd.be/rss/ondernemen\\_durfkapitaal.xml](https://www.tijd.be/rss/ondernemen_durfkapitaal.xml)

[https://www.tijd.be/rss/ondernemen\\_auto.xml](https://www.tijd.be/rss/ondernemen_auto.xml)

<http://feeds.abcnews.com/abcnews/usheadlines>

[http://rss.cnn.com/rss/cnn\\_topstories.rss](http://rss.cnn.com/rss/cnn_topstories.rss)

<http://www.cbsnews.com/latest/rss/main>

<http://www.nytimes.com/services/xml/rss/nyt/National.xml>

[http://online.wsj.com/xml/rss/3\\_7085.xml](http://online.wsj.com/xml/rss/3_7085.xml)

[US general](http://content.usatoday.com/marketing/rss/rsstrans.aspx?feedId=news2)

<http://rss.csmonitor.com/feeds/usa>

<http://feeds.nbcnews.com/feeds/topstories>

<http://feeds.nbcnews.com/feeds/worldnews>

<http://feeds.reuters.com/Reuters/worldNews>

<http://feeds.abcnews.com/abcnews/usheadlines>

[http://rss.cnn.com/rss/cnn\\_topstories.rss](http://rss.cnn.com/rss/cnn_topstories.rss)

<http://www.cbsnews.com/latest/rss/main>

<http://www.nytimes.com/services/xml/rss/nyt/National.xml>

[http://online.wsj.com/xml/rss/3\\_7085.xml](http://online.wsj.com/xml/rss/3_7085.xml)

[US general](http://content.usatoday.com/marketing/rss/rsstrans.aspx?feedId=news2)

<http://rss.csmonitor.com/feeds/usa>

<http://feeds.nbcnews.com/feeds/topstories>

<http://feeds.nbcnews.com/feeds/worldnews>

<http://feeds.reuters.com/Reuters/worldNews>

Every time the python script is run:

- It needs to go through the rss feed list from csv
- Collect all the rss feeds and write the results to a database. (only the new results)
- insert the log events

Collecting rss items:

Every feed returns several. "news" items → each new item is written to a row in the result table: rss\_items

(new item = item where the combination of channel url + id of the news item not yet exists)

logging table needs:

every time a rss read it needs to log

- url
- datetime of run

- status of the rss query:
  - success
  - fail + fail reason
  -

the log is written to the table → rss\_log

### Output table rss\_items

The table rss\_items has 1 line for each news item

The table output rss\_items contains these fields

Table rss_items	description	example
channel_url	url of the rss	<a href="https://www.tijd.be/rss/ondernemen_durfkapitaal.xml">https://www.tijd.be/rss/ondernemen_durfkapitaal.xml</a>
channel_title	Channel name	Tijd.be Ondernemen durfkapitaal
item_id	Item id	10212851
item_title	Title of the RSS feed	Belgisch investeringsfonds 9.5 Ventures haalt nog eens 10 miljoen op
item_category	Item category	Ondernemen
item_Link	Item url	<a href="https://www.tijd.be/r/t/1/id/10212945">https://www.tijd.be/r/t/1/id/10212945</a>
item_description	Item description	17:38 9.5 Ventures, een durfkapitaalfonds dat grote bedrijven helpt innovatieve start-ups uit te bouwen, heeft in snel tempo zijn maximumkapitaal opgehaald. Elisa Vlerick gaat aan de slag bij het fonds.
item_Pubdate	Publication date of the item	Fri, 06 Mar 2020 16:38:35 GMT
item_full_json	Full JSON structure of the item	(no example at this moment)