

Python: how to parse an RSS feed

In this article we will see how to parse a remote RSS feed using a specific Python module.

The feedparser module is able to parse different types of feeds. After activating our virtual environment, we can install it as follows:

```
pip install feedparser
```

This module transforms an RSS feed into an object whose attributes (or properties) correspond to the various XML elements of the feed itself.

We can use it as follows:

```
import feedparser
from datetime import datetime

def read_rss_feed(feed_url=None):
    if feed_url is None:
        return None
    if not feed_url.startswith('https://'):
        return None
    output = []
    rss_feed = feedparser.parse(feed_url)
    entries = rss_feed.entries
    date_fmt = '%a, %d %b %Y %H:%M:%S %Z'
    for entry in entries:
        title = entry.title
        pubdate = entry.published
        date = datetime.strptime(pubdate, date_fmt)
        entry_date = date.strftime('%d/%m/%Y')
```

```
        link = entry.link
        output.append({
            'title': title,
            'link': link,
            'date': entry_date
        })
    return output
```

The function we just defined uses the URL of the feed to invoke the `parse()` method of the `feedparser` module. `rss_feed` is therefore an object whose attributes, as we have said, represent the elements of the XML document. So `entries` is an object list containing all the elements that mark up RSS feed posts. Each of these objects, in turn, has specific attributes that correspond to as many XML tags, such as `title` and `link`.

Basically we loop through the posts and return a list of dictionaries in which we store the post title, its URL and the publication date formatted in the local format. The RSS format has standard dates, so it's easy to specify the starting format in the `date_fmt` variable for later use in the `datetime.strptime()` method.