README

This script implements a simple web crawler that lists the static assets of each web-page it visits.

Usage instructions

Use the following command to run the program (from the root directory of the project):

```
ruby runner.rb <URL> <OUTPUT FILE>
```

The two arguments are optional with the following default values.

- **URL** https://gocardless.com/
- **Output file** sitemap.xml

We start by initializing a WebCrawler object, with a single url in its 'to_visit' list. Then, by calling the 'crawl' method we start processing pages we have not visited yet. When visiting a page we add all its links (ones we did not visit yet) to the 'to_visit' list. When the 'to_visit' list is empty we finish crawling and terminate the program.
My main focus was to write a script that would be as simple and readable as possible. Code that would easy to maintain and extend.
I am using Nokogiri library to parse the HTML of the web-pages and to construct the output XML (I hope that this is acceptable).
One issue I had to solve was to make sure that the program will always terminate. In order to do so I am maintaining a list of all already-visited urls and making sure that the program won't 'visit' any url from that list. In addition, I do not add any urls from a different domain then the given one. Assuming the number of web-pages of that domain is finite, the program will terminate.
Another issue I had to overcome was the formatting of the links. Some links are relative to the current domain, and some are in their full form. I am using regular expressions to identify the different cases and act accordingly.
I have added simple unit tests that cover the code.
I hope that this simple program satisfy your requirements. I could have written a more complex and efficient program, but I was hoping that you would appreciate a simple solution in short time.

The output for *https://gocardless.com/* is in the supplied *sitemap.xml* file.
The GitHub repository of this project can be found in `https://github.com/aharonidan/webcrawler`

Please let me know if you need anything else from me.

Regards,
Dan