**Web Crawler**

Your task is to write a simple web crawler in a language of your choice.

The crawler should be limited to one domain. Given a starting URL – say [wiprodigital.com](http://wiprodigital.com/) - it should visit **all** pages within the domain, but not follow the links to external sites such as Google or Twitter.

**Input:** a starting url. Viz. [wiprodigital.com](http://wiprodigital.com/), wipro.com etc.

**Output:** A simple site map (xml OR json OR plain text) file. The site map should show other pages under the **same** **domain, links to static content such as images and to external URLs**.

Example:

Input: wiprodigital.com

Output:

Plain text:

http://www.example.com/

http://www.example.com/subpage/abc.htm

http://www.someotherdomain.com/xyz.htm http://www.example.com/images/myimg.gif

XML:

<?xml version="1.0" encoding="UTF-8"?>

<sitemap>

<url>

<loc>http://www.example.com/</loc>

</url>

<url>

<loc>http://www.example.com/subpage/abc.htm</loc>

</url>

<url>

<loc>http://www.someotherdomain.com/xyz.htm</loc>

</url>

<image>

<loc>http://www.example.com/images/myimg.gif</loc>

</image>

</sitemap>

We would like to see what you can produce in a couple of hours – please don’t spend much more than that. In addition, please

* **ensure that what you implement is production quality code with unit tests.**
* **briefly describe any tradeoffs/ assumptions you make through comments and / or in a README file.**
* **include the steps needed to build and run your solution with steps to run unit tests.**
* **make your solution available on Github and forward the link.**