# Croudf Funding Aggregator

**http://simplehtmldom.sourceforge.net/manual\_api.htm**

**Req Story 1)**  
A proof of concept for a financial services aggregator.

The aggregator is required to automatically obtain data from 20-40

websites and consolidate it into a single database which can be browsed

or searched from a simple website. Some of the websites are password

protected (passwords and usernames will be provided).

The data fields from the websites will need to be grouped and equivalent

(but differently named) fields identified and paired.

The project will include the development of a simple but functional

website which is suitable for a proof of concept demonstration.

The project involves scraping information about projects that are listed

on UK base crowd funding websites and presenting in on a webpage.

I have the urls for 40 websites, a minimum of 30 of these would need to

be included. As a POC, I am happy for the easiest 80% to be completed

with scoping only for the hardest 20%.

Some example URLs are:

<https://www.crowdbnk.com/projects/projects>

[www.crowdcube.com](http://www.crowdcube.com/)

[www.crowdfunder.co.uk](http://www.crowdfunder.co.uk/)

<https://www.investingzone.com/>

<http://www.icedragons.co.uk/>

It is important to note that the fields on the websites are boradly

equivalent but not exactly the same. So the application will need to pair

similar fields.

If I am provided by the field names and some example data from each of

the website, I will be able to conduct the pairing task.

You will note that most of the websites present the projects in a similar

manner. The output from the application will need to be presented on a

website in a similar manner that can be browsed and searched.

The crawl can be initiated manually. The POC will be used to support

discussions with partners, I would be happy to initiate the day before

the meeting to ensure that the most up to date information was included.  
  
**Req Story 2)**

At this stage, we do not have the permission of the website owners. Here in the UK, it is permissible as long as you comply with the copyright laws. So, one of the key elements is that the POC site is not available to the public.

I would like to establish the first milestone as having scraped the first three website with the data inputted into a database. Are you happy with this?

I am viewing the project in three pieces:

1. Scraping three-five websites and creating database;

2. Re-arranging the database and groups fields,

3. Building the non-public website for demonstration

I will also provide you the details of the initial websites for the scraping.

**Req Story 3**  
First we have to aggrgate these 5 site.

1. Seedrs

2. Crowdcube

3. Crowdbnk

4 Crowdfunder

5. [crowd2fund.com](http://crowd2fund.com/)