**Users Guideline**

There are 3 tools used for search URLs for submission guidelines or homepages.

The first one search tools are made to search home page or any useful URL links from search engine it can help client to find the small publishers (usually it has 1 or 2 per publishers) and totally in list for 30% of journals.

The second one is sorting the result we need after web scraper.

The third tool are not done yet, can be update easily, change the way to run the program, aims to help client or people have no computer science knowledge.

**Environment Settings**

All code can run on any Platform with JRE, other packages can use idea to import Gradle, and open the project.

I am highly recommend running these programs on a server. It will take couple days to get all the result.

1. **Search Tools**
2. google search
3. Yahoo search
4. Bing Search

The search tools are based on copy manually search method on different search engine.

The code explanation is in source code, the things needs to change can be easily found on the top.

For all Search Tools the input should change by following (\*means required)

1. Source excel Path\*
2. Sheet name\*
3. Page numbers (result from search) \*
4. Start row number\*
5. Generate file path and sheet\*
6. Thread.sleep time if get block with error code 503
7. Key words combination

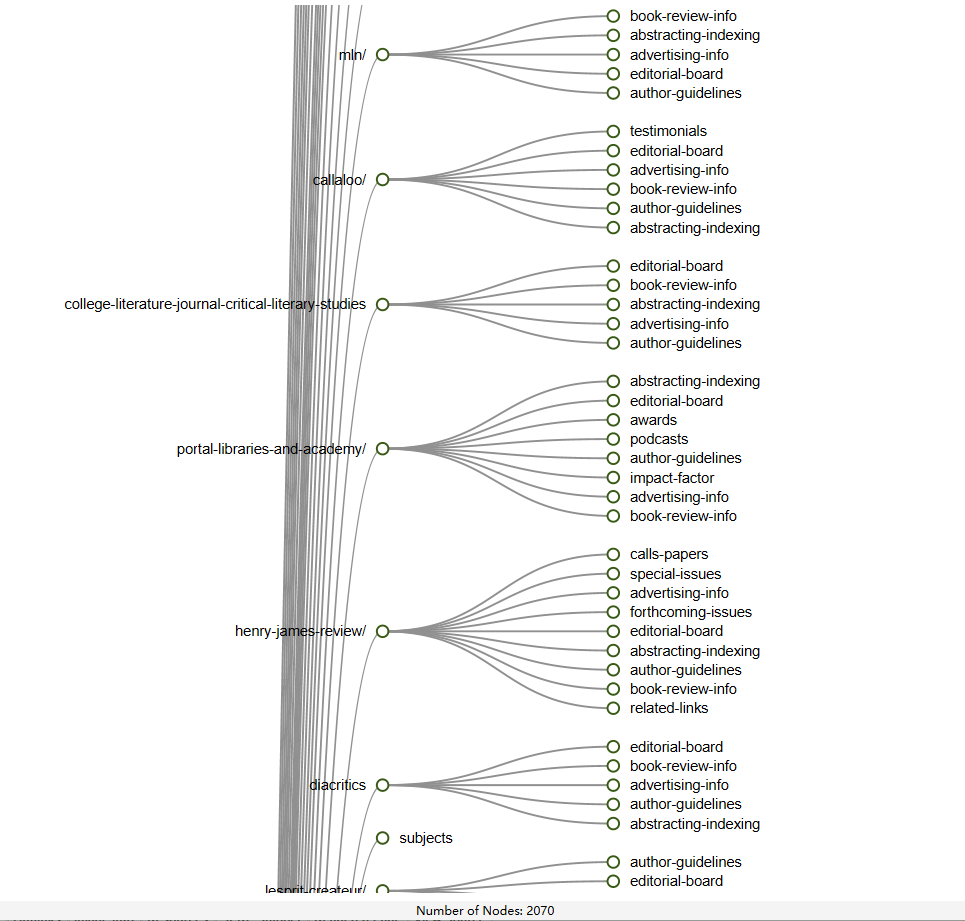
1. **result Sorting**

Fist we have get result from web scraper.

The web scraper tools can be easily find online. Here will give 2 examples

1. <https://www.webscraper.io/>
2. ScreamingFrogSEOSpider

These all can get the URLs easily from specific websites. Recommended to using scrapers can get sitemap of a website, cause client requests to get URLs



Here is a example after we have a list of URL under one publishers, it returns a huge list of URLs but we only need apart from the list. There is one already exist excel shows the homepage or submission guidelines URL formatting, by using regular expression to sort out the URLs we need.

After this step, the list of all journals homepage or submission guidelines are founded. My program are use to sort the result from the URLs, under one publisher they may have lots of journals not under SSCI or A&HCI list.

The reason I used third party software is because of some publishers have a huge websites have more than 400K inlinks.

1. **Tool GUI**

Not done yet but I can give my idea here, due to client or other non-technical people using Search tool still will be hard, under this tool I wish to visualize the input for Search engine tool. The can decide the Excel file or output result by click.