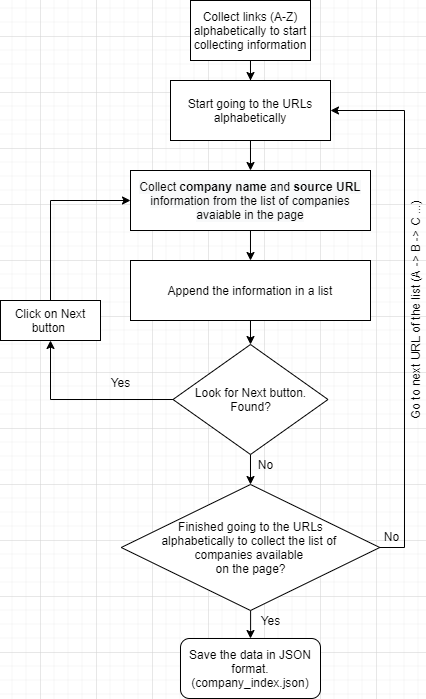
**Leadbook Assessment**

**Q. Briefly describe the architecture of your application?**

****

**Fig**: Flowchart for collecting information from the list of companies present in a page.

I have used **selenium** for crawling information from the websites.

**Steps to collect information of the list of companies from the URL.**

(https://www.adapt.io/directory/industry/telecommunications/A-1)

**Step 1**: Collect the links to go to the list of companies alphabetically.

**Step 2:** Start loading each link alphabetically and store information in a dictionary and append the dictionary into a list.

**Step 3:** Look for **Next** button

**Step 4:** If found, then Click on the button and collect the information. (List of companies)

**Step 5:** If the **Next** button is not found, check If the crawler Completed visiting all the links?

**Step 6:** If **No**, repeat from **Step 2** to **Step 4**

**Step 7:** If **yes**, Store the information in a JSON file (company\_index.json)

**Steps to collect details of a company from the URL.**

**Step 1**: Load the data from json file into a list.

**Step 2:** Start loading each page serially and collected information.

**Step 3:** Check if the page is loaded properly or not

**Step 4:** If not loaded, repeat **Step 2** again.

**Step 5:** If the page is loaded successfully, collect the details of the company and the details of each contact, save them into a dictionary and append to a list.

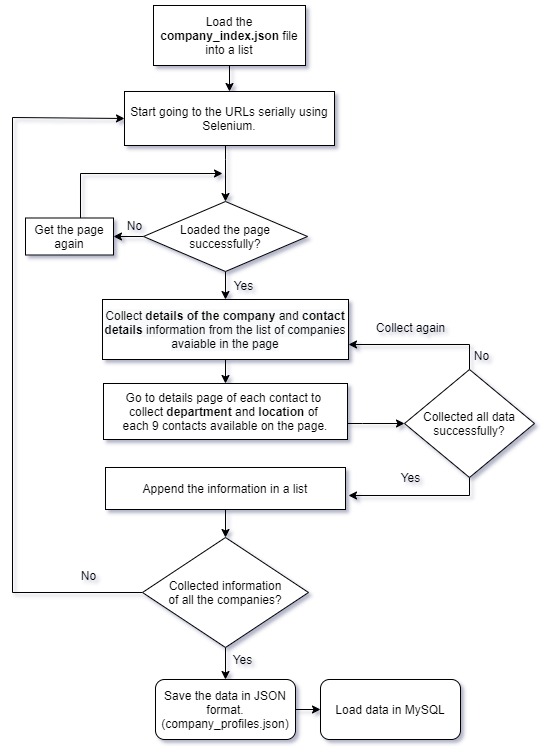
**Step 6:** Check if all required information is collected successfully or not.

**Step 7:** If **No**, refresh the page and repeat **Step 5**.

**Step 8:** If **Yes**, Store the information in a JSON file (company\_profiles.json)

**Step 9:** Load all these info into database

\*\* flowchart for collecting details of each company is presented in the next page.



**Fig**: Flowchart for collecting details of each company and the contact of that company.

**Q. Which database engine you choose and why?**

I chose mongo dB to store company index data and company profiles data in 2 different collections (company\_index, company\_profiles).

Benefit of using Mongo dB:

* Flexible data models
* Fast queries

Data can be inserted in a Mongo Db database without any predefined schema. So, the format or data model can be changed any time, without application disruption.

There were some fields which are not present for some company. So, as Mongo Db offer schema less structure, so I was able to insert the data easily.

Though, my data wasn’t huge but it could have been huge if I was given large task to complete. As I didn’t know about how much data my task will generate, without taking the risk I’ve chosen Mongo Db to store my data. And as it provides high speed for querying data from the database, I can have an advantage over for searching any data.