Internal

Scraping Requirements

Prepared by: Muneer Nawab, CTO

On behalf of: Crowdlinker Inc.

Date prepared: October 28, 2017

Date Updated:

Table of Contents

Change Log

Version 1.0	Muneer Nawab	October 28, 2017	Initial Version

Introduction

This document summarizes requirements for various scraping projects.

Project 1: Business Card to Add on LinkedIn

High Level Requirements:

1. Compile a spreadsheet list of people, with the following information:
   a. Name, Company Name, Email, Position (if available), Phone Number (if available)
   b. Usually this will come from business cards or from a written list

- To transcribe the information, use a scanning app like https://fullcontact.com or http://evernote.com
  c. If email is unavailable, look it up using http://hunter.io or https://findthat.email
  d. Keep the order of the data. Number each entry 1 to x

2. Write a script which will take the data from step 1 and look it up on LinkedIn, using Name and Company Name. Use Artoo.
   a. Return the LinkedIn profile URL of the first match
   b. Record the data to a new CSV file.
   c. User can review the data in the CSV file before proceeding to Step 3

3. For a given set of LinkedIn profiles (i.e. from Step 2), iterate through each one and add

4. Rewrite steps 3 and 4 using Casper JS or Puppeteer (evaluate which is better.)

5. the contact, and specify a note.
6. *If we are have an invite from them already, accept the invite.*
7. *If we are already connected (or we just accepted their invite) send the note as a direct message instead.*
8. *Record all activity in a log file.*

Project 1b: Get number of employees for a list of linkedin profiles.

(highest priority)

visit the linkedin profile for each of these people, click through to their company, and return the number of employees from that company on linkedin. (e.g. "See all 4,054 employees on LinkedIn" → 4000

Input:
- LIst of Linkedin profiles, eg:
    - https://www.linkedin.com/in/erinbury/
    - https://www.linkedin.com/in/mattgray1

(Could be as Json or csv)

Output:
- Linked in profile, e g. https://www.linkedin.com/in/mattgray1
- Name
- Position
- Location
- Company Name
- Company Number of Employees

Extra Feature:
- Ability to sort the output list by number of employees (which should be an optional features.)