

# Executive summary

## Background and Overview

There has been a great waste of time whenever non-local college students in Orem/Provo needed to go online and look for a room to stay for their school year. Most of the time, they go to craigslist.org because it is simple to use and many people post their ad on the site. The process is very time-consuming because, even though it is simple to use, it still takes several clicks to get the detailed results. Even if they do get the results, additional screening is often required because sometimes it is very easy to forget whether the ad has already been looked at or not.

The project was originally initiated with a personal business problem of not being able to find timely information of housing where pets are allowed in the property (I own a cat).

The project has been proposed to cut down repetitive steps and avoid reading the posts that already has been looked at.

The problems have been solved through the following features that the solution offers:

1. No need to launch web browser and go to craigslist.org
2. The solution does not display the same posts twice (or more)
3. The solution automatically fetch all the detailed information about house listings in the result set (Title, Link, price, description, address, postdate)

This document has been prepared to demonstrate the details of the solution.

## Implementation

The solution is divided into two big parts: userForm and getData sub procedure. Detailed information of each part will be discussed in this section.

### 1. UserForm

The screenshot below is the form that the user will interact with the program. When the user presses 'Launch Program' button, which can be found in the Final Project tab, the userform will launch.

**Screenshot 1. UserForm**

To briefly explain how to use the form, the user must enter the minimum price of the rent in the textarea to the right of 'Min Price'. Then the user must enter the maximum price in the textarea to the right of 'Max Price'. Pets allowed check boxes are optional. If the user wants to rent a room that allows cats and/or dogs, the user can simply check it. If not checked, the program will assume that the user wants to get housing where no pets are allowed.

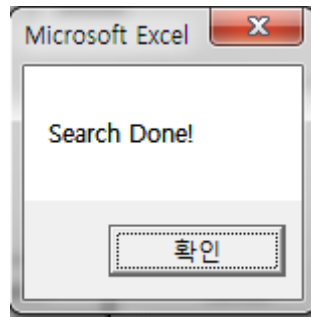
When all the information have been entered, the user needs to press 'Search!' button. This will launch `getData` sub procedure, passing min price, max price, and the pet preference to the procedure. If the user do not want to search any housing, he/she can simple click 'Cancel' button, located right to the 'Search!' button, and this will close the userForm.

## 2. `getData` sub procedure

When the search has been initiated by the user, codes found in the `getData` sub procedure will be executed. The following steps will take place during the execution:

- ① The program will insert the search terms into the first few cells of working sheet (minimum price, maximum price, cat/dog allowed). See Screenshot 3 for more details.
- ② The program will create a uri which will be used on craigslist.org based on the parameters passed in from the userForm.
- ③ The program will invoke the internet explorer and use the uri to retrieve the result set.
- ④ The program will gather information for each post that will appear on the internet explorer screen. After extracting the title and the link to detailed information of a post, it will visit each link to the posts, pulling information such as description, price, date, and address. The entry will appear right after the last record.
  - A. During this process, the program will use links to compare currently working post with all other posts to determine whether the same post exists in the worksheet.
    - i. If the same post already exists, the program will move on to the next post.
    - ii. If there is no same post, the program will drill into the detailed webpage of the posting, and it will extract necessary information.
- ⑤ Once the program reached the end of the result set, it will notify the user (See screenshot 2 below). This is used to ensure that the codes have been fully executed.
- ⑥ The user then can continue to search by entering different search terms into the

userForm and execute the sub procedure again, or click 'cancel' button to simply work with the records that have been fetched.



Screenshot 2 Search Completed

During the execution of the sub procedure, the user will not see the internet explorer window; however, the current worksheet will refresh as new records are being put it. This is intentional because sometimes it takes a lot of time to fill in the necessary information into the worksheet due to large number of records found, and it may appear that the program is not responding if worksheet does not refresh.

Minimum Price:		300000				
Maximum Price:		500000				
Cats Allowed?	Yes					
Dogs Allowed?	Yes					
Title	Link	Price	Address	Description		Date

Screenshot 3 Initial layout of the workbook

### 3. Web agent

Web agent contains a list of methods which can be utilized to on a webpage to easily navigate through. The following methods were used predominantly used throughout the getData sub procedure to complete the project.

- moveTo method: Moves the cursor to the place where a specified string appears on the current working webpage
- getText method: Grabs the text starting from current cursor until it reaches specified string value. These two methods were
- openPage method: Opens up a webpage on a web browser based on a given uri

## Learning and conceptual difficulties encountered

There were several difficulties encountered during the development of the actual solution. Ability to search records in KSL.com had to be dropped due to technical difficulties. Also working with the detailed page to obtain necessary information was very difficult and time consuming.

### 1. Web agent and KSL.com

Originally, the proposed solution was supposed to fetch housing information from KSL.com, which is also popular website to go to get information about housing. Initially, web agent worked just fine with KSL.com, however, it stopped working on KSL.com after a few days due to an unknown reason. KSL.com uses different web site to pull the housing data, rentler.com. However, rentler.com does not fully support the features that KSL.com offers—for example, there is no search option for pet allowance in the house. Personally, this is a critical feature for the solution to be able to provide, so I had to stop working with KSL.com. Professor recommended non-visual approach, but the time co

2. Inconsistency among the records

As far as formatting is concerned, I encountered inconsistencies among the posts in craigslist.org. Essentially, there were different cases which I needed to be able to handle in order to pull the right information from a webpage.

- i. Pictures of the property, followed by address with google map
- ii. Pictures of the property, followed by addresses without google map
- iii. No pictures of the property, followed by address without google map
- iv. No pictures of the property, and not address available

Since I could not visit all the postings and inspect the source code for each, it was very hard to find different patterns. Initially, I thought there were only two cases where address with the google map and one without. I was able to find different patterns after hours of investment in terms of time, and this problem was finally resolved.

3. 'Br' tag in the description

Sometimes the description portion of the entry takes a lot of space despite the fact that it does not have a lot of text that goes with it. I also had to spend a lot of time to find a solution that will get rid of trailing spaces at the end of the description due to br tag. I tried replace method and RTrim method just to name a few. However, I was not able to succeed in getting rid of unnecessary spaces at the end.