

Scraping Certified Inspectors Data From nachi.org by Chedy Chaaben

There is an estimated total of 11,100 rows of data that needs to be scraped from a directory found at <https://www.nachi.org/certified-inspectors/browse/us/>, and produced in .csv format.

The scraped data shall fill the following columns within each row produced:

FirstName, LastName, CompanyName, City, State, Website, InterNACHI, OfficePhone, MobilePhone, FaxPhone

If data for any of the above fields is not present within the directory listing and respective directory listing pages being scraped, I will put “null” in its place.

Here is an example to reference, and how to identify the data needed:

The screenshot shows the 'Certified Home Inspectors in United States' page on nachi.org. The page has a blue header with the 'International Association of Certified Home Inspectors' logo. Below the header, there's a search bar on the left and a main content area. The main content area has a title 'Certified Home Inspectors in United States' and a description. Below the description, it says 'Showing page 1 of 111 total pages.' There is a table with three columns: 'Member', 'Servicing', and 'Phone'. The first row shows 'Richard Walters, HI-4139' as the member, 'Home Inspections Resource' as the company, 'Theodore' as the servicing location, and '(251) 370-7676' as the phone number. There is a small icon of a person with a plus sign next to the phone number.

Member	Servicing	Phone
Richard Walters, HI-4139 Home Inspections Resource	Theodore	(251) 370-7676

FirstName

- The first blue word in the directory listing
- In this example, it is “Richard”

LastName

- The blue word after the FirstName
- Ignore the comma, and space immediately after the FirstName
- In this example, it is “Walters”

CompanyName

- The grey word found under FirstName and LastName
- In this example, it is “Home Inspections Resource”

City

- The word under the ‘Servicing’ column in the directory listing
- In this example, it is “Theodore”

Office Phone


- The number under the ‘Phone’ column in the directory listing
- In this example, it is “(251) 370-7676”

The remainder of the data must be scraped from the respective directory listing’s directory listing page.

Below, is an example of the directory listing page for the directory listing example used above:

nachi.org/certified-inspectors/richard-walters-hi-4139-50089

International Association of Certified Home Inspectors®





Richard Walters, HI-4139

Home Inspections Resource

Richard Walters, HI-4139 is a Theodore, AL **Certified Professional Inspector®** and certified by the International Association of Certified Home Inspectors® (InterNACHI®)

✓ **\$25,000 Honor Guarantee** backed by InterNACHI

 **Call Richard at (251) 370-7676**

 **Email Richard Walters, HI-4139**

State

- Found in the first paragraph of text, following the City (which has already been identified)
- Ignore the comma, and space immediately after the City
- In this example, State is "AL" - just the two letters identifying the respective US state

Website

- Found just to the left of the contact form
- It should be the full URL that is being linked to
- In this example, Website is "http://homeinspectionsresource.com/"



Home Inspections Resource Website
Visit [Home Inspections Resource](http://homeinspectionsresource.com/) on the web at [http://homeinspectionsresource...](http://homeinspectionsresource.com/)

at [http://homeinspectionsresource...](http://homeinspectionsresource.com/)

InterNACHI

- Found just to the left of the contact form
- The combination of letters and numbers under the InterNACHI ID section.
- In this example, InterNACHI is "NACHI16041715"

InterNACHI ID
An InterNACHI® ID can be used to verify Richard Walters, HI-4139's certification at any time.

 **NACHI16041715**

Email Richard Walters, HI-4139

Your Name

Your Phone Number

Your Email

Your Mobile Phone

Some directory pages will have a Phone Numbers section to the right of the contact form, and some will not. The example we've been using above does not have a Phone Numbers section.

If there is no Phone Numbers section, MobilePhone and Fax will be "null"

Mobile Phone

- Found when the Phone Numbers section has a phone number labeled as 'Mobile'
- In this example MobilePhone is "(941) 204-4429"
- It is possible for MobilePhone and OfficePhone to be duplicate values

at <http://www.swnil.com>

Phone Numbers

(941) 204-4429 Office

(941) 204-4429 Mobile

InterNACHI ID

Your Err

Your Me

Fax

- Found when the Phone Numbers section has a phone number labeled as 'Fax'
- In this example Fax is "(775) 853-2430"
- It is possible for Fax and OfficePhone to be duplicate values

at <http://www.wentinspectionserv...>

Phone Numbers

(775) 853-2430 Office

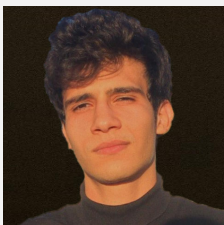
(775) 853-2430 Fax

InterNACHI ID

Your Emal

Your Mess

Additional Inspection Services



chedychaaben@gmail.com



chedychaaben



chedychaaben.com