…There are a number of websites which list this sort of data. I'm focusing on the static sites for the moment.

I'm using are R with `{rvest}` (and a few other Tidyverse packages thrown in for good measure).

```
library(glue)
library(dplyr)
library(purrr)
library(stringr)
library(rvest)
```

The data are paginated. Fortunately the URL includes the page number as a GET parameter, so stepping through the pages is simple. I defined a global variable, `URL`, with a `{glue}` placeholder for the page number.

This is how I'm looping over the pages. The page number, `page`, is set to one initially and incremented after each page of results is scraped. When it gets to a page without at results, the loop is stopped.

```
page = 1

while (TRUE) {
  items <- read_html(glue(URL)) %>% html_nodes(SELECTOR)
  #
  if (length(items) == 0) break                  # Check if gone past last page.

  # Extract data for each item here...

  page <- page + 1                               # Advance to next page.
}
```

That's the mechanics. Within the loop I then used `map_dfr()` from `{purrr}` to iterate over `items`, delving into each item to extract its name and price.

```
map_dfr(items, function(item) {
    tibble(
      part = item %>% html_node("p") %>% html_text(),
      price = item %>% html_node(".product-item--price small") %>% html_text()
    )
  })
```

The results from each page are appended to a list and finally concatenated using `bind_rows()`.

```
> dim(parts)
[1] 987   2
```

Scraping a single category yields 987 parts. Let's take a look at the first few.

```
> head(parts)
  part
price
  <chr>
<chr>
1 R986110000 Bosch Rexroth New Replacement Hydraulic Axial Piston Motor For CAT
240-7732 $1,534.00
2 R986110001 Bosch Rexroth New Replacement Hydraulic Axial Piston Motor For CAT
213-5268 $1,854.00
3 R986110002 Bosch Rexroth New Replacement Hydraulic Axial Piston Motor For CAT
```

```
266-8034 $1,374.00
4 R986110003 Bosch Rexroth New Replacement Hydraulic Axial Piston Motor For CAT
296-6728 $1,754.00
5 R986110004 Bosch Rexroth New Replacement Hydraulic Axial Piston Motor For CAT
136-8869 $1,494.00
6 R986110005 Bosch Rexroth New Replacement Hydraulic Axial Piston Motor For CAT
255-6805 $1,534.00
```

That's looking pretty good already. There's one final niggle: the data in the `price` column are strings. Ideally we'd want those to be numeric. But to do that we have to strip off some punctuation. Not a problem thanks to functions from `{stringr}`.

```
parts <- parts %>%
  mutate(
    price = str_replace(price, "^\\$", ""),       # Strip off leading "$"
    price = str_replace_all(price, ",", ""),      # Strip out comma separators
    price = as.numeric(price)
  )
```

Success!

```
> head(parts)
  part
price
  <chr>
<dbl>
1 R986110000 Bosch Rexroth New Replacement Hydraulic Axial Piston Motor For CAT
240-7732  1534
2 R986110001 Bosch Rexroth New Replacement Hydraulic Axial Piston Motor For CAT
213-5268  1854
3 R986110002 Bosch Rexroth New Replacement Hydraulic Axial Piston Motor For CAT
266-8034  1374
4 R986110003 Bosch Rexroth New Replacement Hydraulic Axial Piston Motor For CAT
296-6728  1754
5 R986110004 Bosch Rexroth New Replacement Hydraulic Axial Piston Motor For CAT
136-8869  1494
6 R986110005 Bosch Rexroth New Replacement Hydraulic Axial Piston Motor For CAT
255-6805  1534
```

The great thing about scripting a scraper is that, provided that the website has not been dramatically restructured, the scraper can be run at any time to gather an updated set of results.