

Dataa verkosta -muistilista

Ohjelmia

- Verkkoselain: esim. [Google Chrome](#)
- Tekstieditori: esim. [Sublime Text](#), [Atom](#), [Notepad++](#) (Windows)
- Taulukkolaskenta: esim. [Google Sheets](#), [Calc](#), [Excel](#)
- Datansiivous: [OpenRefine](#)
- Muuta: [Klaxon](#), [Nearley](#)) jne.

Verkkoselaimen kehittäjätyökalut

- Cmd + Alt + i (OSX)
- Ctrl + Shift + i (Linux, Win)
- Safarissa kytkettävä päälle asetuksissa

Toisenlaista kopiointia

- [Top Rated Movies](#)
- Tietyn luokan sisällä olevat linkkielementit:

```
$$('.titleColumn a')
```

- Tee myös lista linkkielementtien osoitteista:

```
Array.from($$('.titleColumn a')).map(x => x.textContent)
```

- Tallenna myös leikepöydälle:

```
copy(Array.from($$('.titleColumn a')).map(x => x.textContent))
```

Lähestymistapoja raapimiseen

- Palvelut ja lisäosat: esim. [Web Scraper](#), [SelectorGadget](#)
- Funktiot: esim. [IMPORTXML](#), [IMPORTHTML](#)
- Koodi: esim. [Puppeteer](#) (Node), [rvest](#) (R), [Beautiful Soup](#) (Python), [Nokogiri](#) (Ruby)

Koodia: R

```
library(rvest)
topmovies <- read_html("https://www.imdb.com/chart/top")
movienames <- topmovies %>%
  html_nodes(".titleColumn a") %>%
  html_text()
movienames
```

Koodia: Node.js

```
const fs = require('fs');
const puppeteer = require('puppeteer');
(async() => {
  const browser = await puppeteer.launch();
  const page = await browser.newPage();
  await page.goto('http://www.imdb.com/chart/top');
  const movies = await page.$$eval('.titleColumn a', elems => {
    return Array.from(elems).map(elem => elem.textContent);
  });
  await browser.close();
  console.log(movies);
})();
```

“If you have ghosts...”

- Päätön verkkoselain: Nopeampi, mutta rajoittunut
- Oikea verkkoselain: Hitaampi, kuten oikea selain

Päätte

- Tee kyselyjä: `curl https://www.google.com`
- Lataa asioita: `xargs -n 1 curl -O < urls.txt` (olettaen, että `urls.txt` on rivitetty lista osoitteita)

Rajapinnat inspiraationa

- [Wikidata](#)
- [Maalaukset, joissa on apinoita](#)
- [Why Are There So Many 17th Century Paintings of Monkeys Getting Drunk?](#)