Az elmúlt fél év alatt rájöttem, hogy kegyetlen undorító dolog adatokat kinyerni az internetről, és megfogadtam, hogy messziről kerülni fogom a témát, ha lehet.

Nem lehetett sajnos.

Egyik szaktársam készített egy szuper [videót](https://miau.my-x.hu/miau/281/f12.mp4), melyben elmagyarázza, hogy hogyan lehet kevesebb mint 5 perc alatt adatokat kinyerni az arukereso.hu weboldalról egy bizonyos termékről.

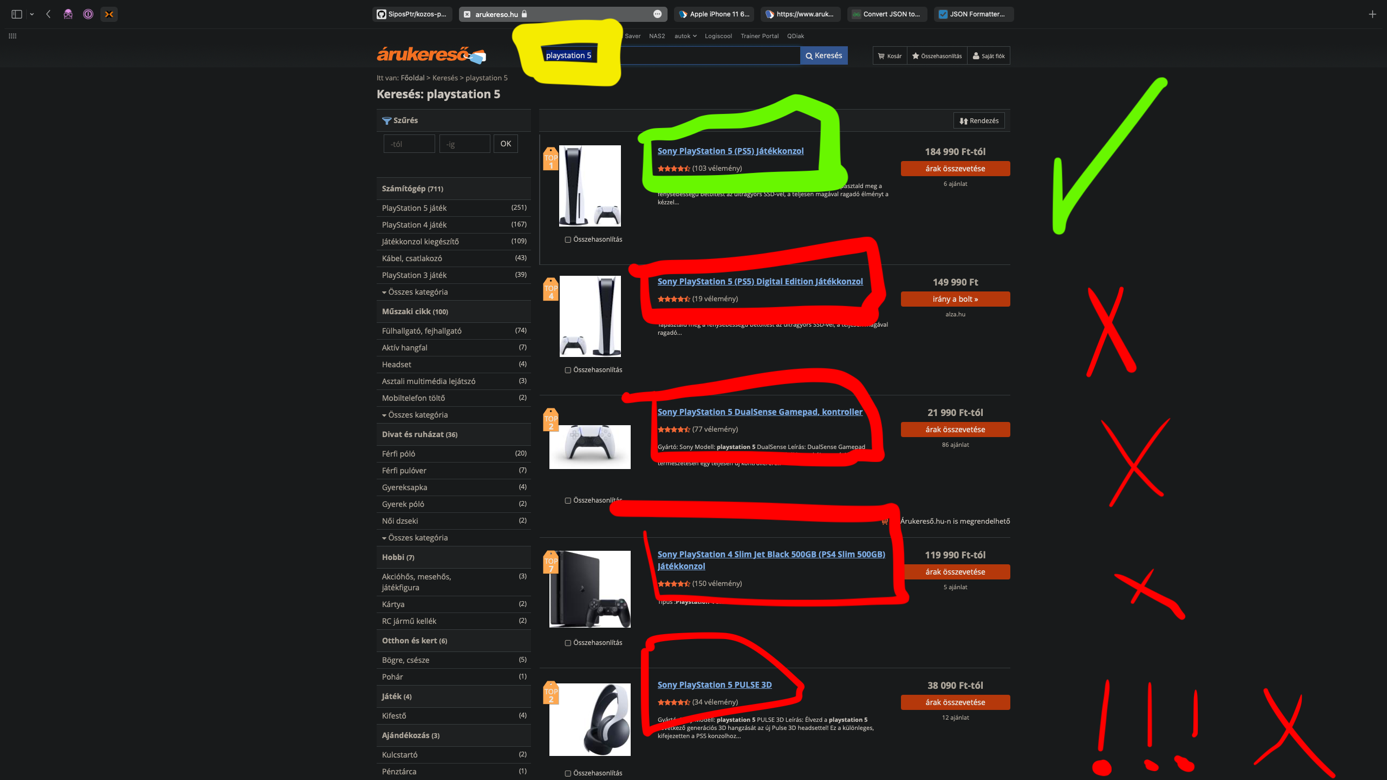
Ezt az ötletet, folyamatot szerettem volna automatizálni, hátha a következő ilyen adatkinyerős projektemben meg tudom könnyíteni magam, vagy esetleg mások munkáját.

(mint utólag észrevettem, ezt az ötletet már Pitlik tanár úr a „potenciális input-output konstelláció” email-ben leírta ☺ )

Elég buta a programom, nem is vagyok büszke rá, cserébe működik.

A program inputként elfogadja azokat a keresési szavakat, amiket mi is beírnánk az árukereső oldal keresőjébe.

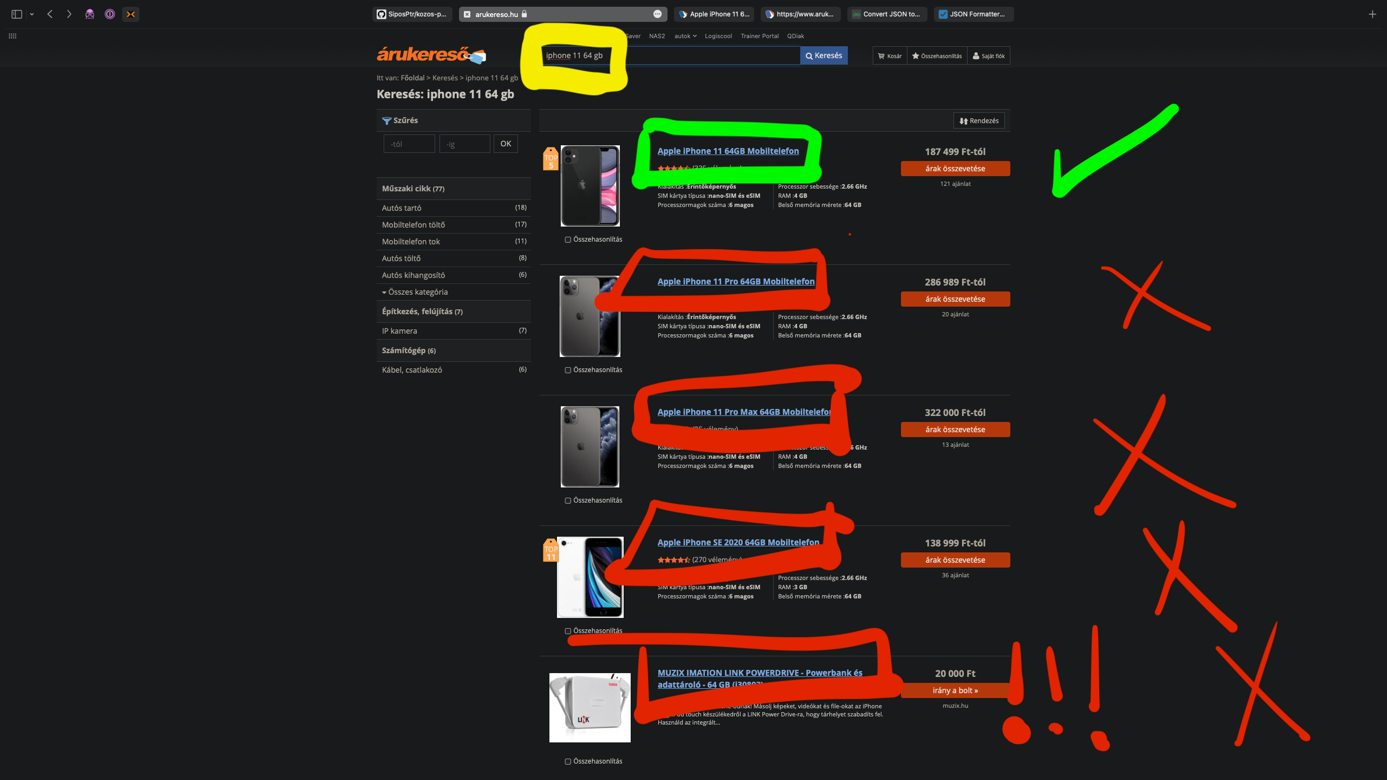
A keresés eredményei közül a legelsőt veszem figyelembe, mert a tesztelés során kiderült, hogy mindig a legelső találat a legpontosabb, az utána következők pontatlanok.



Itt a keresett kifejezés a „playstation 5” volt.

A legelső találat a megfelelő. Második találat már a Digital Edition-t mutatja, és ahogy haladunk lejjebb, egyre pontatlanabb lesz a találat.

A legutolsó abszolút rossz.



Itt a keresett kifejezés a „iphone 11 64 gb” volt.

Szintúgy, mint az előző keresésnél, itt is a legelső találat volt, ami pontosan az, amire mi rákerestünk.

A következő mar nem is a megfelelő modellt jeleníti meg.

A legutolsó szintén abszolút rossz.

Lehetett volna több keresést ugyan ilyen módszerrel ellenőrizni, de nem volt hozzá kedvem, és úgy éreztem, hogy ez a két eset bőven elég arra, hogy bizonyíthassam, hogy miért a legelső találat lesz az általam választott.

program használata:

import api

api.main("iphone 11 64 gb")

A main függvény paramétere az a keresés, amit a weboldal keresőjébe is beírnánk.

Az output még rendezetlen, hiszen nem tudom, hogy miként fog nekünk kelleni, de ezt később igény szerint lehet majd személyre szabni.

FONTOS: minden futtatáskor a program létrehoz egy ki.json fájlt. Ezt a következő futtatás előtt töröljük mindig, mert a program csak beleír a fájlba, nem ellenőrzi, hogy abba volt előzőleg valami, emiatt a json szintaxis megsérül.