University *of Ljubljana*
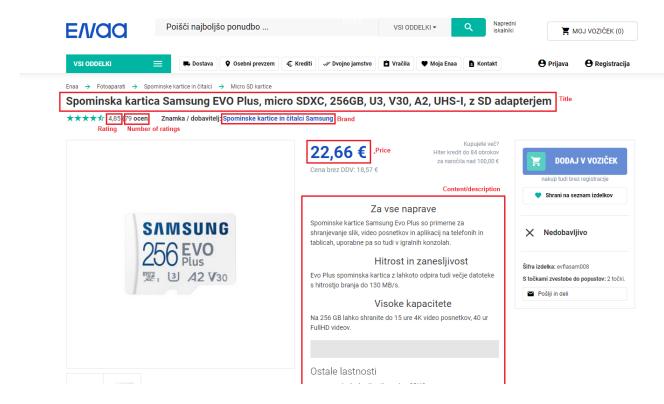Faculty *of Computer and*
Information Science

Naloga 2: Ekstrakcija strukturiranih podatkov s spleta

Marija Klemenčič, Timotej Šalamon, Krištof Zupan

1. Izbrane spletne strani

Poleg rtvslo.si in overstock smo se odločili da izberemo spletnos stran enaa.com. Na zgornji sliki je prikazana razdelitev podatkov, ki jih ekstrahiramo. Med njimi so Title, Rating, Number of ratings, Price, Brand and Content/Description.

2. Implementacija

Projekt smo implementirali v programskem jeziku python. Vsako implementacijo bomo opisali v svojem poglavju s svojimi izrazi za ekstrakcijo podatkov. Pred ekstrakcijo smo najprej z uporabo knjižnice selenium shranili celotno stran, ko je naložena za to, da ni potrebno vsakič na novo nalagati stran, ko se izvaja ekstrakcija podatkov. Tako, da smo za vsako stran shranili nov html (rendered), ki je shranjen v mapi s predpono __ (npr. __Audi.html).

3. RegEx

Ta del smo implementirali z vgrajenimi funkcijami za regex v pythonu.

Rtvslo:

- Author:

```
<div class="author-name">(.*?)</div>
```

- Time:

```
<div class="publish-meta">\s*(.*?)\s*<br>
```

- Title:

```
<h1>(.*?)<\/h1>
```

- Subtitle:

```
<div class="subtitle">(.*?)</div>
```

- Lead:

```
<p class="lead">(.*?)</p>
```

- Content: First extract article and than extract paragraphs.

```
<article class="article">(.*?)<\/article>
```

```
<p.*?>(.*?)<\/p>
```

Overstock:

- Item: Najprej ekstrakcije vseh izdelkov nato ekstrakcija posameznih elementov.

```
(?<=<tbody>)(<tr\b[^>]*>[\s\S]*?<\/tr>)
(?=<\/tbody>)
```

- Title:

```
<td valign="top">\s*<a\s[^>]*><b>(.*?)<\/
b><\/a>
```

- List price:

```
<b>List Price:</b></td><td align="left"
nowrap="nowrap"><s>(.*?)</s>
```

- Price:

```
<b>Price:</b></td><td align="left" nowrap
="nowrap"><span class="bigred"><b>
(.*?)</b></span>
```

- Saving:

```
<b>You Save:</b></td><td align="left"
nowrap="nowrap"><span class="
littleorange">(.*?) \(.*?\)</span>
```

- Saving percentage:

```
<b>You Save:</b></td><td align="left"
nowrap="nowrap"><span class="
littleorange">.*? \((.*?)\)</span>
```

- Content:

```
<span class="normal">(.*?)<br>
```

Enaa:

- Title:

```
<h1 itemprop="name" class="text-break">
(.*?)</h1>
```

- Price:

```
<div class="single-product-price">(.*?)<
div
```

- Rating:

```
<span itemprop="ratingValue">(.*?)</span>
```

- Review count:

```
<span itemprop="reviewCount">(.*?)</span>
```

- Brand:

```
<div class="product-brand">\s*Znamka /
dobavitelj: <a.*?>(.*?)</a>\s*</div>
```

- Description:

```
<div class="single-product-description">
.*?<p>(.*?)</p>\s*<br><a href="#opis"
>Celoten opis
```

4. Xpath

Ekstrakcijo z Xpath smo implementirali z uporabo knjižnice lxml, ki je izvajala poizvedbe z xPath.
Rtvslo:

- Author:

```
//div[@class='author-timestamp']/strong/
text()
```

- Time:

```
//div[@class="author-timestamp"]/text()
```

- Title:

```
//h1/text()
```

- Subtitle:

```
//div[@class='subtitle']/text()
```

- Lead:

```
//p[@class='lead']/text()
```

- Content:

```
//article[@class='article']//p/text()
```

Overstock:

- Title:

```
//tbody//tr[@bgcolor]//td[@valign]//a/b/
    text()
```

- List price:

```
//tbody//tr[@bgcolor]//td//table//tbody//
    tr//td//table/tbody//tr//td//s/text
    ()
```

- Price:

```
//tbody//tr[@bgcolor]//td//table//tbody//
    tr//td//table/tbody//tr//td//span/b/
    text()
```

- Saving:

```
//tbody//tr[@bgcolor]//td//table//tbody//
    tr//td//table/tbody//tr//td//span[
    @class='littleorange']/text()
```

- Saving percentage: Parse the string gotten from saving.

- Content:

```
//tbody//tr[@bgcolor]//td//table//tbody//
    tr//td/span[@class='normal']/text()
```

Enaa:

- Title:

```
//div[@class='section-title']/h1[
    @itemprop='name']/text()
```

- Price:

```
//div[@class='single-product-price']/text
    ()
```

- Rating:

```
//span[@itemprop='ratingValue']/text()
```

- Review count:

```
//span[@itemprop='reviewCount']/text()
```

- Brand:

```
//div[@class='product-brand']/a/text()
```

- Description:

```
//div[@class='single-product-description
    ']//text()[not(parent::a)]
```

5. RoadRunner

Pri implementaciji algoritma RoadRunner smo si pomagali s predlaganim člankom (RoadRunner: Towards Automatic Data Extraction from Large Web Sites). Pred zagonom algoritma smo iz vhodnih HTML dokumentov odstranili odvečne značke, skripte in komentarje. Z uporabo BeautifulSoap smo razčlenili HTML, da smo določene oznake uporabili kot potencialni pokazatelj za strukturirane podatke. Primerjali smo posamezne strukture za različne strani, pri čemer smo mejo za podobnost nastavili pri 0,5. Strukture, skupne obema stranema, smo uporabili za generiranje ovojnice. Generirane ovojnice vseh treh strani so smiselne.

5.1 Psevdokoda algoritma

```
Počisti in predprocesiraj obe HTML strani.

Poišči strukturne elemente na obeh straneh.

Primerjaj strukturne elemente na obeh straneh
in identificiraj skupne elemente obeh strani.
Pri tem upoštevaj določeno mero razlikovanja
elementov.

Generiraj wrapper iz zaznanih skupnih
strukturnih elementov.
```

6. Izhodi

Regex rtv Audi

```
1  {
2      "author":"Miha Merljak",
3      "time":"28. december 2018 ob 08:51",
4      "title":"Audi A6 50 TDI quattro: nemir v premijskem razredu",
5      "subTitle":"Test nove generacije",
6      "lead":"To je novi audi A6. V razred najdra\u017ejih in najbolj premijskih \
           u017erebcev je vnesel nemir, \u0161e preden je sploh zapeljal na parkirni
           prostor, rezerviran za izvr\u0161nega direktorja. ",
7      "content":"Samo poglejte njegovo masko \u2013 to ogromno satovje z radarji na
           takem pol o\u017eaju, da se ti na avtocesti tudi pri 120 km/h vsi spo\
           u0161tljivo umikajo, saj so prepri\u010dani, da gre za Pahorjev ali \u0160ar\
           u010dev avto. Seveda, novi A6 lahko cesto in promet skenira s kar petimi
           radarji, petimi ka merami, infrarde\u010do kamero za no\u010dni vid,
           dvanajstimi ultrazvo\u010dnimi senzorji in laserskim \u010ditalnikom \u2013
           lidarjem. V glavnem voja\u0161ka tehnologija v slu\u017ebi varnosti za fante,
            ki smo radi gledali Top Gun, Bonda in druge mo\u017eakarja s finimi igra\
           u010dami.\nNovo poglavjeVozni\u0161ki delovni prostor je novo poglavje
           digitalne dobe, z dvema ogromnima zaslonoma, ki tako kot naprednej\u0161i
           telefoni dregnejo blazinice va\\u01 61ih prstov, kot se sprehajate po steklu.
            A \u0161e bolj se nam zdi pomembno, da so osnovna stikala tam, kjer jih pri\
           u010dakujete. Najprej so torej zagotovili enostavno osnovo, tisti bolj \"
           advanced\" vozniki pa si lahko nato v se skupaj \u0161e veliko bolj
           prilagodijo. Velik korak naprej pri kabinskem udobju zaznavajo tudi na zadnji
            klopi, tam je prostora v vseh smereh precej ve\u010d.\n\u010ce vam pogled na
            Audijev spisek dodatne opreme ne odvzame vo lje do \u017eivljenja, potem
           vsekakor toplo priporo\u010damo nakup zra\u010dnega vzmetenja, saj dobi z
           njim A6 ve\u010d razli\u010dnih in vozni\u0161ko zelo uporabnih karakterjev.\
           nEnako velja za seksi lu\u010di z inteligentno m atri\u010dno osvetlitvijo,
           pa za \u0161portno podvozje in vsekakor za \u0161tirikolesno krmiljenje. S
           tem postane A6 med ovinki v ob\u010dutku na volanu \u0161e veliko kraj\u0161i
            in bolj agilen. Vse na\u0161teto smo presku\\u016 1ali v dru\u017ebi
           agregata 50 TDI, ki je v resnici klasi\u010dni trilitrski dizel, podkrepljen
           z elektromotorjem. Ja, ta audi je mehki hibrid z izjemnim navorom in dovolj
           mo\u010di kadar koli in kjer koli. Si pa mislimo, da bo najve\u010dji del
           trga zadovoljil \u017ee u\u010dinkovit dvolitrski mehki hibrid z mo\u010djo
           150 kilovatov.\nKlju\u010dni tehni\u010dni podatki:\n- na testu Audi A6 50
           TDI quattro tiptronic\nMere:- dol\u017eina: 4,9 m- medosna razdalja: 2,9 m-
           obra\u010dalni krog: 12,1 m- prtlja\u017enik: 530 l- masa: 1.900 kg\nPogon:-
           trilitrski \u0161estvaljni dizelski motor- mo\u010d: 210 kW- navor: 620 Nm-
           8-stopenjski samodejni menjalnik- pogon na vsa \u0161tiri kolesa- pnevmatike:
            225/60 R17- poraba: 6,6 l/100 km = 8,9 EUR/100 km- posoda za gorivo: 73 l-
           doseg: 1.106 km- izpusti CO2: 147 g/km\nStro\u0161ki pri 15.000 km in 5-letni
            uporabi:- nakupna cena: 69.080 EUR- stro\u0161ki finan\\ u010dnega lizinga:
           4.463 EUR/5 let- stro\u0161ki registracije: 10.829 EUR/5 let- stro\u0161ki
           vzdr\u017eevanja: 1.926 EUR/5 let- stro\u0161ki goriva: 6.702 EUR/75.000 km-
           stro\u0161ek 1 kompleta pnevmatik: 716 EUR- vrednosti po 5 letih po Eurotaxu:
            33.964 EUR\n- stro\u0161ki skupaj: 1.001 EUR/mesec\n"
8  }
```

Regex rtv Volvo

```
1   {
2       "author":"Miha Merljak",
3       "time":"25. januar 2019 ob 15:23",
4       "title":"Volvo XC 40 D4 AWD momentum: suvereno med najbolj\u0161e v razredu",
5       "subTitle":"Test novega modela",
6       "lead":"XC 40 je najmanj\u0161i Volvov SUV, ki se o blikovno skoraj v celotni
            naslanja na oba ve\u010dja predhodnika. \u017de samo s tem so mu vrata do
            denarnic tistih kupcev, ki i\u0161\u010dejo izstopajo\u010do, a hkrati visoko
             kultivirano in pre\u010di\u0161\u010deno dizajners ko govorico, na pol
            odprta.",
7       "content":"Volvo se je ni\u017eejih srednjih razredov v preteklosti izogibal ali
            pa je vanje vstopil z zelo ni\u0161nimi produkti, ki niso pustili ve\
            u010djega tr\u017enega pe\u010data. V primeru XC 40 ni te\u017eko napovedati,
             da bo ta tradicija prekinjena. Ponuja namre\u010d visoko kakovost kon\
            u010dne izdelave in v kabini odli\u010dno premi\u0161ljeno funkcionalnost ter
             na dotik prijetne materiale. \n\u0160e posebej hva limo \u0161tevilo,
            iznajdljivost in velikost razli\u010dnih odlagalnih prostorov ter \u0161iroke
            , \u010dvrste in zelo udobne sede\u017ee. Intuitivno in enostavno logi\
            u010dno je upravljanje z velikim vmesnikom, ki z ve\u010dfunk cijskim
            zaslonom na dotik kraljuje na z roko lahko dostopnem mestu na sredinski
            armaturi. Razo\u010daranj ne bo niti v velikosti in uporabnosti prtlja\
            u017enega prostora, ki s 460 litri prostornine sicer ni med ve\u010djimi v ra
             zredu, a se v uporabni\u0161kem smislu odkupi z dobro urejenostjo ter
            domiselnimi re\u0161itvami pregrajevanja.\nXC 40 je od tal odmaknjen
            konkretnih 21 cm, a sta vzmetenje in krmilni mehanizem tako nastavljena, da
            ponuja tudi v hitro odpeljanih ovinkih zelo dolgo nevtralno in predvidljivo
            lego. V premeru presku\u0161anega modela, ki je imel v paketu R design
            vzmetenje \u0161e nekoliko bolj trdo, se je to samo \u0161e bolj potrdilo, a
            je v tem primeru treba ra\u010dunati na manj udobno vo\u017enjo \u010dez
            razli\u010dne asfaltne grbine. Podoben razmislek velja opraviti tudi pri
            izbiri motorja.\nPresku\u0161ani 2-litrski dizel s 190 KM predstavlja vrh
            ponudbe, ki z mo\u010djo, udobjem in tudi povpre\u010dno porabo navdu\u0161i
            predvsem pri avtocestnih dolgoproga\u0161kih izzivih, v po\u010dasni mestni
            vo\u017enji ter pri pogostih postankih in speljevanjih pa deluje preve\u010d
            robusten.\nXC 40 je s \\ u010dvrsto gradnjo, funkcionalno in udobno kabino
            ter \u0161tevilnimi asisten\u010dnimi sistemi in izstopajo\u010dim
            skandinavskim dizajnom v premi\u0161ljenem trenutku vstopil na trg modnih
            mestnih terencev, v katerem se brez e ne same sence dvoma suvereno postavi
            med najdra\u017eje in najbolj premijske v mestu.\nKlju\u010dni tehni\u010dni
            podatki:- na testu Volvo XC40 2.0 TD avt awd momentum\nMere:- dol\u017eina:
            4,4 m- medosna razdalja: 2,7 m- obra\\u 010dalni krog: 11,4 m- oddaljenost od
             tal: 21 cm- prtlja\u017enik: 432 l- masa: 2.250 kgPogon:- 2-litrski 4-valjni
             bencinski motor- mo\u010d: 140 kW- navor: 400 Nm- 8-stopenjski samodejni
            menjalnik- pogon na vsa \u0161tiri koles a- pnevmatike: 235/50 R19 - poraba:
            6,3 l/100 km = 8,2 EUR/100km- posoda za gorivo: 54 l- doseg: 857 km- izpusti
            CO2: 133 g/km\nStro\u0161ki pri 15.000 km in 5-letni uporabi:- nakupna cena:
            43.619 EUR- stro\u0161ki finan\u010dne ga leasinga: 3.268 EUR/5 let- stro\
            u0161ki registracije: 8.701 EUR/5 let- stro\u0161ki vzdr\u017eevanja: 2.320
            EUR/5 let- stro\u0161ki goriva: 6.190 EUR/75.000 km- stro\u0161ek 1 kompleta
            pnevmatik: 923 EUR- vrednosti po 5 letih po Eurotaxu: 18.886 EUR- stro\
            u0161ki skupaj: 774 EUR/mesec\n"
8   }
```

Regex Overstock jewelry 1

```
1   {
2       "products":[
3           {
4               "title":"10-kt. Seven Diamond Ladies Heart Ring (0.08 TW)",
5               "listPrice":"$149.00",
6               "price":"$69.99",
7               "saving":"$79.01",
8               "saving_percentage":"53%",
9               "content":"This ladies fashion ring dazzles\nwith hearts and diamonds. The
                   gold band is crafted into delicate, open hearts.\nSeven brilliant-cut
                   diamonds add a bit of sparkle. "
10          },
11          {
12              "title":"10-Kt. Diamond Ring (.25 TW)",
13              "listPrice":"$250.00",
14              "price":"$74.90",
15              "saving":"$175.10",
16              "s aving_percentage":"70%",
17              "content":"Nineteen round diamonds accent this 10-karat yellow gold ring
                   with filigree accents."
18          },
19          {
20              "title":"10-kt. Pearl and Diamond Butterfly Earrings",
21              "listPrice":"$149.00",
22              "price":"$42.99",
23              "s aving":"$106.01",
24              "saving_percentage":"71%",
25              "content":"Perfectly proportioned 5.5- to\n6-mm cultured pearls on 10-
                   karat yellow gold settings highlight these petite\nearrings. A dainty
                   rhodium-plated gold butterfly studded wi th a diamond (0.02\ntotal
                   carat weight, J-K color, I-2 clarity) rests atop each pearl."
26          },
27          {
28              "title":"14-kt. Diamond 'S' Tennis Bracelet (2.00 TW)",
29              "listPrice":"$1,539.99",
30              "price":"$499.99",
31              "saving":"$1,040.00",
32              "saving_per centage":"67%",
33              "content":"Invest in a swirl of light with\nthis diamond 'S' tennis
                   bracelet. Crafted in 14-karat gold, the piece features\n49 diamonds for
                    two full carats. The 7.25-inch bracelet closes with a pressure\nclasp.
                    "
34          },
35          {
36              "title":"10-kt. Diamond Band Fashion Ring (.11 TW)",
37              "listPrice":"$179.99",
38              "price":"$79.99",
39              "saving":"$100.00",
40              "saving_percentage":"55%",
41              "content":"Crafted in white and yellow gold,\nthis ring displays a band of
                    s even round diamonds. Order your new gold and\ndiamond fashion ring
                   today at our low, online price."
42          },
43          {
```

```
44        "title":"14-kt. White Gold, Pearl and Diamond Ring",
45        "listPrice":"$419.99",
46        "price":"$149.99",
47        "saving":"$270.00",
48        "savin g_percentage":"64%",
49        "content":"Show your romantic side with this\n14-karat diamond and pearl
              ring. Set in a domed band of 14-karat white gold,\nthe ring features a
              7-mm cultured pearl. Curved rows of diamonds flank the\npearl ."
50      },
51      {
52        "title":"14-kt. Gold Diamond Present Future Pendant (.25TW)",
53        "listPrice":"$299.00",
54        "price":"$149.99",
55        "saving":"$149.01",
56        "saving_percentage":"49%",
57        "content":"Designed with three large, sparkling\ndiamonds to repr esent
              past, present, and future, this stunning pendant is\nset in gleaming
              14-karat gold. It incorporates a total of nine diamonds (0.25\ntotal
              carat weight, K color, I-2 to I-3 clarity).  "
58      },
59      {
60        "title":"14-kt. Diamond Solitaire Pendant (.33 TW)",
61        "listPrice":"$1,019.99",
62        "price":"$319.99",
63        "saving":"$700.00",
64        "saving_percentage":"68%",
65        "content":"In this simple, yet elegant pendant,\na round brilliant diamond
              (0.33 total carat weight, H-J color, I-1 to I-2\nclarity) is prong-set
              in 14-karat white gold."
66      },
67      {
68        "title":"14-kt. Diamond Solitaire Earrings (0.33 TW)",
69        "listPrice":"$639.99",
70        "price":"$199.99",
71        "saving":"$440.00",
72        "saving_percentage":"68%",
73        "content":"Dazzl e your way into her heart,\nwith these classic diamond
              solitaire earrings. Two brilliant-cut diamonds\n(0.33 total carat
              weight, G-H color, I-1 to I-2 clarity) are set in four\nprongs of 14-
              karat white gold."
74      },
75      {
76        "title":"14-kt. Diamond Cross Pendant (.06 TW)",
77        "listPrice":"$305.00",
78        "price":"$119.99",
79        "saving":"$185.01",
80        "saving_percentage":"60%",
81        "content":"Over a cleanly sculpted Roman\ncross of 14-karat white gold
              drapes a slender banner conta ining three bright\nprong-set round
              diamonds (0.06 total carat weight, H-I color, I clarity)."
82      },
83      {
84        "title":"14-kt. Diamond Solitaire Stud Earrings (.50 TW)",
85        "listPrice":"$999.99",
86        "price":"$359.99",
```

```
 87            "saving":"$640.00",
 88            "savi ng_percentage":"64%",
 89            "content":"Every jewelry collection needs\na classic pair of diamond
                  solitaire earrings. Set in 14-karat gold, these\ndiamond stud earrings
                  (0.50 total carat weight) have post backs with butterfly\nclasps ."
 90        },
 91        {
 92            "title":"14-kt. Cultured Pearl Diamond Earrings",
 93            "listPrice":"$508.99",
 94            "price":"$179.99",
 95            "saving":"$329.00",
 96            "saving_percentage":"64%",
 97            "content":"Create an elegant appearance with\nthese pearl and diamond stud
                  ea rrings. Set in 14-karat yellow gold, each\nearring features an 8 to
                  8.5-mm cultured white pearl. Prong-set round diamonds\naccent the
                  pearls. Posts with butterfly clasps secure the earrings."
 98        },
 99        {
100            "title":"14-kt. Diamond 7.5-8 mm Pearl Pendant",
101            "listPrice":"$196.99",
102            "price":"$69.99",
103            "saving":"$127.00",
104            "saving_percentage":"64%",
105            "content":"Add a classic to your jewelry\ncollection with this 14-karat
                  gold, diamond, and pearl necklace. The 7.5-8\n mm cultured white pearl
                  creates the focal point of the pendant, while a diamond\n(0.10 TW) adds
                   sparkle."
106        },
107        {
108            "title":"14-kt. Diamond Solitaire Earrings (.50 TW)",
109            "listPrice":"$1,369.99",
110            "price":"$409.99",
111            "saving":"$960.00 ",
112            "saving_percentage":"70%",
113            "content":"This earring set has two brilliant-cut\ndiamonds (0.50 total
                  carat weight, G-H color, I-1 to I-2 clarity) set in\nfour prongs of 14-
                  karat white gold."
114        },
115        {
116            "title":"14-kt White Gold Diam ond Band (0.50 TW)",
117            "listPrice":"$1,635.00",
118            "price":"$609.99",
119            "saving":"$1,025.01",
120            "saving_percentage":"62%",
121            "content":"Crafted of 14-karat white gold,\nthis stylish ring features a
                  bright row of 20 channel-set, prince ss-cut baguette\ndiamonds. Treat
                  her like royalty and save when you buy jewelry treasures\nat Overstock.
                  com."
122        }
123    ]
124 }
```

Regex Overstock jewelry 2

```
 1  {
 2      "products":[
 3          {
 4              "title":"14-kt. Green Jade Hoops",
 5              "listPrice":"$90.00",
 6              "price":"$46.99",
 7              "saving":"$43.01",
 8              "saving_percentage":"47%",
 9              "content":"Hoops of cool green jade rest\nbetween 14-karat yellow gold
                  endpieces. The hoops graduate in thickness from\n3 mm at the ends to 6
                  mm in the center, with approximately 29 mm overall\ndiameter."
10          },
11          {
12              "title":"14-kt. Jade Doughnut Pendant",
13              "listPrice":"$150.00",
14              "price":"$48.99",
15              "saving":"$101.01",
16              " saving_percentage":"67%",
17              "content":"The 25-mm disk hangs delicately\nfrom a 14-karat gold chain.
                  The disk features a dramatic gold Chinese character\nin the center,
                  accompanied by four stylized gold bees."
18          },
19          {
20              "title":"14-kt. Charcoal Jade and Ruby Elephant Pendant",
21              "listPrice":"$100.00",
22              "price":"$28.99",
23              "saving":"$71.01",
24              "saving_percentage":"71%",
25              "content":"Carved of rich dark grey jade,\nthis elephant pendant has 14-
                  karat yellow gold app lied to mark the feet,\ntusk, tail, and blanket.
                  A 2-mm round faceted ruby in a gold bezel setting\nforms the eye. The
                  pendant hangs from an 18-inch chain."
26          },
27          {
28              "title":"14-kt. Carved Lavender Jade Earrings",
29              "listPrice":"$80.0 0",
30              "price":"$39.99",
31              "saving":"$40.01",
32              "saving_percentage":"50%",
33              "content":"Luscious 8-mm lavender jade balls, carved with intricate Asian
                  style, dangle from a 14-karat yellow gold French hook."
34          },
35          {
36              "title":"14-kt. Jade C ross Pendant",
37              "listPrice":"$150.00",
38              "price":"$49.99",
39              "saving":"$100.01",
40              "saving_percentage":"66%",
41              "content":"Green jade and gold create this\nbeautiful cross pendant.
                  Cylindrical bars of green jade feature caps and\nce nter of 14-karat
                  yellow gold."
42          },
43          {
44              "title":"14-kt. Multicolored Jade Earrings",
```

```
45          "listPrice":"$375.00",
46          "price":"$99.99",
47          "saving":"$275.01",
48          "saving_percentage":"73%",
49          "content":"A delicate wrapping of 14-karat\nyellow gold wire holds six 6 x
                4 pear shapes of jade in various shades: brilliant\ngreen, orange,
                lavender, black, pale yellow, and white. The post earrings\nhave
                butterfly backs."
50      },
51      {
52          "title":"14-kt. Multicolored Jade Ring",
53          "listPrice":"$250.00",
54          "price":"$56.99",
55          "saving":"$193.01",
56          "saving_percentage":"77%",
57          "content":"A delicate wrapping of 14-karat\nyellow gold wire holds six 6 x
                4 ovals of jade in various shades: brilliant\ngreen, orange, lavender,
                bl ack, pale yellow, and white. A narrow gold band\ndivides to support
                the setting."
58      },
59      {
60          "title":"14-kt. Onyx and Ruby Elephant Pendant",
61          "listPrice":"$100.00",
62          "price":"$35.99",
63          "saving":"$64.01",
64          "saving_percentage":"64%",
65          "content":"Carved of rich black onyx, this\nelephant pendant has 14-karat
                yellow gold applied to mark the feet, tusk,\ntail, and blanket. A 2-mm
                round faceted ruby in a gold bezel setting forms\nthe eye. The pendant
                hangs from an 18-inch chain."
66      }
67   ]
68 }
```

Regex Enaa sandisk

```
1  {
2      "title":"SanDisk Ultra microSDXC 128GB 100MB/s Class 10 UHS-I – SDSQUNR-128G-
           GN3MN",
3      "price":"11,88 \u20ac",
4      "rating":"4,83",
5      "reviewCount":"96",
6      "brand":"Spominske kartice in \u010ditalci  SanDisk",
7      "description":"Ustvar jen za prihranek trenutkaKartica SanDisk Ultra\u00ae
           microSD\u2122 UHS-I vam daje svobodo snemanja, shranjevanja in deljenja ve\
           u010d kot kadar koli prej. S kapaciteto do 512 GB1 ima na\u0161a kartica
           SanDisk Ultra microSD\u2122 prostor za \u0161e ve\u010d ur videa v polni
           visoki lo\u010dljivosti3 in zagotavlja hitrosti prenosa do 100 MB/s2, kar vam
            pomaga hitro premikati to vsebino.Deluje s pametnimi telefoni in tabli\
           u010dnimi ra\u010dunalniki Androi d\u2122Kartice SanDisk\u00ae so zdru\
           u017eljive s pametnimi telefoni in tablicami Android.Zajemite in shranite \
           u0161e ve\u010d ur videa v polni visoki lo\u010dljivostiS kapaciteto do 512
           GB1 lahko na kartico shranite &amp;sc"
8  }
```

Regex Enaa samsung

```
1  {
2      "title":"Spominska kartica Samsung EVO Plus, micro SDXC, 256GB, U3, V30, A2, UHS
           -I, z SD adapterjem",
3      "price":"22,66 \u20ac",
4      "rating":"4,85",
5      "reviewCount":"79",
6      "brand":"Spominske kartice in \u010ditalci  Samsung",
7      "description":"Spominske kartice Samsung Evo Plus so primerne za shranjevanje
           slik, video posnetkov in aplikacij na telefonih in tablicah, uporabne pa so
           tudi v igralnih konzolah.  Hitrost in zanesljivost Evo Plus spominska
           kartica z lahkoto odpira tudi ve\u010dje datoteke s hitrostjo branja do 130
           MB/s.  Visoke kapacitete Na 256 GB lahko shranite do 15 ure 4K video
           posnetkov, 40 ur FullHD videov.  Ostale lastnosti  spominska kartica tipa
           microSDXC prim erna za uporabo v pametnih telefonih, tabli\u010dnih ra\
           u010dunalnikih, ro\u010dnih igralnih konzolah, kamerah in fotoaparatih. UHS-I
            vodilo kapaciteta 256 GB masa: pribli\u017eno 0,25 g hitrostni razredi:
            U3"
8  }
```

xPath rtv Audi

```json
1  {
2      "author":"Miha Merljak",
3      "time":"28. december 2018 ob 08:51",
4      "title":"Audi A6 50 TDI quattro: nemir v premijskem razredu",
5      "subTitle":"Test nove generacije",
6      "lead":"To je novi audi A6. V razred najdra\u017ejih in najbolj premijskih \
          u017erebcev je vnesel nemir, \u0161e preden je sploh zapeljal na parkirni
          prostor, rezerviran za izvr\u0161nega direktorja. ",
7      "content":"Samo poglejte njegovo masko \u2013 to ogromno satovje z radarji na
          takem pol o\u017eaju, da se ti na avtocesti tudi pri 120 km/h vsi spo\
          u0161tljivo umikajo, saj so prepri\u010dani, da gre za Pahorjev ali \u0160ar\
          u010dev avto. Seveda, novi A6 lahko cesto in promet skenira s kar petimi
          radarji, petimi ka merami, infrarde\u010do kamero za no\u010dni vid,
          dvanajstimi ultrazvo\u010dnimi senzorji in laserskim \u010ditalnikom \u2013
          lidarjem. V glavnem voja\u0161ka tehnologija v slu\u017ebi varnosti za fante,
           ki smo radi gledali Top Gun, Bonda in druge mo\u017eakarja s finimi igra\
          u010dami. Vozni\u0161ki delovni prostor je novo poglavje digitalne dobe, z
          dvema ogromnima zaslonoma, ki tako kot naprednej\u0161i telefoni dregnejo
          blazinice va\u0161ih prstov, k ot se sprehajate po steklu. A \u0161e bolj se
          nam zdi pomembno, da so osnovna stikala tam, kjer jih pri\u010dakujete.
          Najprej so torej zagotovili enostavno osnovo, tisti bolj \"advanced\" vozniki
           pa si lahko nato vse skupaj \\u01 61e veliko bolj prilagodijo. Velik korak
          naprej pri kabinskem udobju zaznavajo tudi na zadnji klopi, tam je prostora v
           vseh smereh precej ve\u010d. \u010ce vam pogled na Audijev spisek dodatne
          opreme ne odvzame volje do \u017eiv ljenja, potem vsekakor toplo priporo\
          u010damo nakup zra\u010dnega vzmetenja, saj dobi z njim A6 ve\u010d razli\
          u010dnih in vozni\u0161ko zelo uporabnih karakterjev. Enako velja za seksi lu
          \u010di z inteligentno matri\u010dno osv etlitvijo, pa za \u0161portno
          podvozje in vsekakor za \u0161tirikolesno krmiljenje. S tem postane A6 med
          ovinki v ob\u010dutku na volanu \u0161e veliko kraj\u0161i in bolj agilen.
          Vse na\u0161teto smo presku\u0161ali v dru\u017e bi agregata 50 TDI, ki je v
          resnici klasi\u010dni trilitrski dizel, podkrepljen z elektromotorjem. Ja, ta
           audi je mehki hibrid z izjemnim navorom in dovolj mo\u010di kadar koli in
          kjer koli. Si pa mislimo, da bo najve\u010dji de l trga zadovoljil \u017ee u\
          u010dinkovit dvolitrski mehki hibrid z mo\u010djo 150 kilovatov. – na testu
          Audi A6 50 TDI quattro tiptronic – dol\u017eina: 4,9 m – medosna razdalja:
          2,9 m – obra\u010dalni krog: 12,1 m – prtlja\\u017 enik: 530 l – masa: 1.900
          kg – trilitrski \u0161estvaljni dizelski motor – mo\u010d: 210 kW – navor:
          620 Nm – 8-stopenjski samodejni menjalnik – pogon na vsa \u0161tiri kolesa –
          pnevmatike: 225/60 R17 – poraba: 6,6 l/100 km = 8, 9 EUR/100 km – posoda za
          gorivo: 73 l – doseg: 1.106 km – izpusti CO : 147 g/km – nakupna cena: 69.080
           EUR – stro\u0161ki finan\u010dnega lizinga: 4.463 EUR/5 let – stro\u0161ki
          registracije: 10.829 EUR/5 let – stro\u0161ki vzdr \u017eevanja: 1.926 EUR/5
          let – stro\u0161ki goriva: 6.702 EUR/75.000 km – stro\u0161ek 1 kompleta
          pnevmatik: 716 EUR – vrednosti po 5 letih po Eurotaxu: 33.964 EUR – stro\
          u0161ki skupaj: 1.001 EUR/mesec"
8  }
```

xPath rtv Volvo

```
1  {
2     "author":"Miha Merljak",
3     "time":"25. januar 2019 ob 15:23",
4     "title":"Volvo XC 40 D4 AWD momentum: suvereno med najbolj\u0161e v razredu",
5     "subTitle":"Test novega modela",
6     "lead":"XC 40 je najmanj\u0161i Volvov SUV, ki se o blikovno skoraj v celotni
           naslanja na oba ve\u010dja predhodnika. \u017de samo s tem so mu vrata do
           denarnic tistih kupcev, ki i\u0161\u010dejo izstopajo\u010do, a hkrati visoko
            kultivirano in pre\u010di\u0161\u010deno dizajners ko govorico, na pol
           odprta.",
7     "content":"Volvo se je ni\u017eejih srednjih razredov v preteklosti izogibal ali
           pa je vanje vstopal z zelo ni\u0161nimi produkti, ki niso pustili ve\
           u010djega tr\u017enega pe\u010data. V primeru XC 40 ni te\u017eko napovedati,
            da bo ta tradicija prekinjena. Ponuja namre\u010d visoko kakovost kon\
           u010dne izdelave in v kabini odli\u010dno premi\u0161ljeno funkcionalnost ter
            na dotik prijetne materiale.  \u0160e posebej hval imo \u0161tevilo,
           iznajdljivost in velikost razli\u010dnih odlagalnih prostorov ter \u0161iroke
           , \u010dvrste in zelo udobne sede\u017ee. Intuitivno in enostavno logi\
           u010dno je upravljanje z velikim vmesnikom, ki z ve\u010dfunkc ijskim
           zaslonom na dotik kraljuje na z roko lahko dostopnem mestu na sredinski
           armaturi. Razo\u010daranj ne bo niti v velikosti in uporabnosti prtlja\
           u017enega prostora, ki s 460 litri prostornine sicer ni med ve\u010djimi v
           raz redu, a se v uporabni\u0161kem smislu odkupi z dobro urejenostjo ter
           domiselnimi re\u0161itvami pregrajevanja. XC 40 je od tal odmaknjen
           konkretnih 21 cm, a sta vzmetenje in krmilni mehanizem tako nastavljena, da
           ponuja tudi v h itro odpeljanih ovinkih zelo dolgo nevtralno in predvidljivo
           lego. V premeru presku\u0161anega modela, ki je imel v paketu R design
           vzmetenje \u0161e nekoliko bolj trdo, se je to samo \u0161e bolj potrdilo, a
           je v tem primeru tr eba ra\u010dunati na manj udobno vo\u017enjo \u010dez
           razli\u010dne asfaltne grbine. Podoben razmislek velja opraviti tudi pri
           izbiri motorja. Presku\u0161ani 2-litrski dizel s 190 KM predstavlja vrh
           ponudbe, ki z mo\u010djo, ud objem in tudi povpre\u010dno porabo navdu\u0161i
            predvsem pri avtocestnih dolgoproga\u0161kih izzivih, v po\u010dasni mestni
           vo\u017enji ter pri pogostih postankih in speljevanjih pa deluje preve\u010d
           robusten. XC 40 je s \\u010 dvrsto gradnjo, funkcionalno in udobno kabino ter
            \u0161tevilnimi asisten\u010dnimi sistemi in izstopajo\u010dim skandinavskim
            dizajnom v premi\u0161ljenem trenutku vstopil na trg modnih mestnih terencev
           , v katerem se brez ene s ame sence dvoma suvereno postavi med najdra\u017eje
            in najbolj premijske v mestu. – na testu Volvo XC40 2.0 TD avt awd momentum
           Mere: – dol\u017eina: 4,4 m – medosna razdalja: 2,7 m – obra\u010dalni krog:
           11,4 m – oddaljenost od tal: 21 cm – prtlja\u017enik: 432 l – masa: 2.250 kg
           Pogon: – 2-litrski 4-valjni bencinski motor – mo\u010d: 140 kW – navor: 400
           Nm – 8-stopenjski samodejni menjalnik – pogon na vsa \u0161tiri kolesa –
           pnevmatike: 235/50 R19  – poraba: 6,3 l/100 km = 8,2 EUR/100km – posoda za
           gorivo: 54 l – doseg: 857 km – izpusti CO2: 133 g/km Stro\u0161ki pri 15.000
           km in 5-letni uporabi: – nakupna cena: 43.619 EUR – stro\u0161ki finan\
           u010dnega leasinga: 3.268 EUR/ 5 let – stro\u0161ki registracije: 8.701 EUR/5
            let – stro\u0161ki vzdr\u017eevanja: 2.320 EUR/5 let – stro\u0161ki goriva:
           6.190 EUR/75.000 km – stro\u0161ek 1 kompleta pnevmatik: 923 EUR – vrednosti
           po 5 letih po Eurotaxu: 18.8 86 EUR – stro\u0161ki skupaj: 774 EUR/mesec"
8  }
```

xPath Overstock jewelry 1

```
 1  {
 2      "products":[
 3          {
 4              "title":"10-kt. Seven Diamond Ladies Heart Ring (0.08 TW)",
 5              "listPrice":"$149.00",
 6              "price":"$69.99",
 7              "saving":"$79.01",
 8              "saving_percentage":"53%",
 9              "content":"This ladies fashion ring dazzles\nwith hearts and diamonds. The
                  gold band is crafted into delicate, open hearts.\nSeven brilliant-cut
                  diamonds add a bit of sparkle. "
10          },
11          {
12              "title":"10-Kt. Diamond Ring (.25 TW)",
13              "listPrice":"$250.00",
14              "price":"$74.90",
15              "saving":"$175.10",
16              "s aving_percentage":"70%",
17              "content":"Nineteen round diamonds accent this 10-karat yellow gold ring
                  with filigree accents."
18          },
19          {
20              "title":"10-kt. Pearl and Diamond Butterfly Earrings",
21              "listPrice":"$149.00",
22              "price":"$42.99",
23              "s aving":"$106.01",
24              "saving_percentage":"71%",
25              "content":"Perfectly proportioned 5.5- to\n6-mm cultured pearls on 10-
                  karat yellow gold settings highlight these petite\nearrings. A dainty
                  rhodium-plated gold butterfly studded wi th a diamond (0.02\ntotal
                  carat weight, J-K color, I-2 clarity) rests atop each pearl."
26          },
27          {
28              "title":"14-kt. Diamond 'S' Tennis Bracelet (2.00 TW)",
29              "listPrice":"$1,539.99",
30              "price":"$499.99",
31              "saving":"$1,040.00",
32              "saving_per centage":"67%",
33              "content":"Invest in a swirl of light with\nthis diamond 'S' tennis
                  bracelet. Crafted in 14-karat gold, the piece features\n49 diamonds for
                   two full carats. The 7.25-inch bracelet closes with a pressure\nclasp.
                   "
34          },
35          {
36              "title":"10-kt. Diamond Band Fashion Ring (.11 TW)",
37              "listPrice":"$179.99",
38              "price":"$79.99",
39              "saving":"$100.00",
40              "saving_percentage":"55%",
41              "content":"Crafted in white and yellow gold,\nthis ring displays a band of
                   s even round diamonds. Order your new gold and\ndiamond fashion ring
                  today at our low, online price."
42          },
43          {
```

```
44          "title":"14-kt. White Gold, Pearl and Diamond Ring",
45          "listPrice":"$419.99",
46          "price":"$149.99",
47          "saving":"$270.00",
48          "savin g_percentage":"64%",
49          "content":"Show your romantic side with this\n14-karat diamond and pearl
                ring. Set in a domed band of 14-karat white gold,\nthe ring features a
                7-mm cultured pearl. Curved rows of diamonds flank the\npearl ."
50      },
51      {
52          "title":"14-kt. Gold Diamond Present Future Pendant (.25TW)",
53          "listPrice":"$299.00",
54          "price":"$149.99",
55          "saving":"$149.01",
56          "saving_percentage":"49%",
57          "content":"Designed with three large, sparkling\ndiamonds to repr esent
                past, present, and future, this stunning pendant is\nset in gleaming
                14-karat gold. It incorporates a total of nine diamonds (0.25\ntotal
                carat weight, K color, I-2 to I-3 clarity).  "
58      },
59      {
60          "title":"14-kt. Diamond Solitaire Pendant (.33 TW)",
61          "listPrice":"$1,019.99",
62          "price":"$319.99",
63          "saving":"$700.00",
64          "saving_percentage":"68%",
65          "content":"In this simple, yet elegant pendant,\na round brilliant diamond
                (0.33 total carat weight, H-J color, I-1 to I-2\nclarity) is prong-set
                in 14-karat white gold."
66      },
67      {
68          "title":"14-kt. Diamond Solitaire Earrings (0.33 TW)",
69          "listPrice":"$639.99",
70          "price":"$199.99",
71          "saving":"$440.00",
72          "saving_percentage":"68%",
73          "content":"Dazzl e your way into her heart,\nwith these classic diamond
                solitaire earrings. Two brilliant-cut diamonds\n(0.33 total carat
                weight, G-H color, I-1 to I-2 clarity) are set in four\nprongs of 14-
                karat white gold."
74      },
75      {
76          "title":"14-kt. Diamond Cross Pendant (.06 TW)",
77          "listPrice":"$305.00",
78          "price":"$119.99",
79          "saving":"$185.01",
80          "saving_percentage":"60%",
81          "content":"Over a cleanly sculpted Roman\ncross of 14-karat white gold
                drapes a slender banner conta ining three bright\nprong-set round
                diamonds (0.06 total carat weight, H-I color, I clarity)."
82      },
83      {
84          "title":"14-kt. Diamond Solitaire Stud Earrings (.50 TW)",
85          "listPrice":"$999.99",
86          "price":"$359.99",
```

```
 87            "saving":"$640.00",
 88            "savi ng_percentage":"64%",
 89            "content":"Every jewelry collection needs\na classic pair of diamond
                   solitaire earrings. Set in 14-karat gold, these\ndiamond stud earrings
                   (0.50 total carat weight) have post backs with butterfly\nclasps ."
 90        },
 91        {
 92            "title":"14-kt. Cultured Pearl Diamond Earrings",
 93            "listPrice":"$508.99",
 94            "price":"$179.99",
 95            "saving":"$329.00",
 96            "saving_percentage":"64%",
 97            "content":"Create an elegant appearance with\nthese pearl and diamond stud
                   ea rrings. Set in 14-karat yellow gold, each\nearring features an 8 to
                   8.5-mm cultured white pearl. Prong-set round diamonds\naccent the
                   pearls. Posts with butterfly clasps secure the earrings."
 98        },
 99        {
100            "title":"14-kt. Diamond 7.5-8 mm Pearl Pendant",
101            "listPrice":"$196.99",
102            "price":"$69.99",
103            "saving":"$127.00",
104            "saving_percentage":"64%",
105            "content":"Add a classic to your jewelry\ncollection with this 14-karat
                   gold, diamond, and pearl necklace. The 7.5-8\n mm cultured white pearl
                   creates the focal point of the pendant, while a diamond\n(0.10 TW) adds
                    sparkle."
106        },
107        {
108            "title":"14-kt. Diamond Solitaire Earrings (.50 TW)",
109            "listPrice":"$1,369.99",
110            "price":"$409.99",
111            "saving":"$960.00 ",
112            "saving_percentage":"70%",
113            "content":"This earring set has two brilliant-cut\ndiamonds (0.50 total
                   carat weight, G-H color, I-1 to I-2 clarity) set in\nfour prongs of 14-
                   karat white gold."
114        },
115        {
116            "title":"14-kt White Gold Diam ond Band (0.50 TW)",
117            "listPrice":"$1,635.00",
118            "price":"$609.99",
119            "saving":"$1,025.01",
120            "saving_percentage":"62%",
121            "content":"Crafted of 14-karat white gold,\nthis stylish ring features a
                   bright row of 20 channel-set, prince ss-cut baguette\ndiamonds. Treat
                   her like royalty and save when you buy jewelry treasures\nat Overstock.
                   com."
122        }
123    ]
124 }
```

xPath Overstock jewelry 2

```
1   {
2       "products":[
3           {
4               "title":"14-kt. Green Jade Hoops",
5               "listPrice":"$90.00",
6               "price":"$46.99",
7               "saving":"$43.01",
8               "saving_percentage":"47%",
9               "content":"Hoops of cool green jade rest\nbetween 14-karat yellow gold
                   endpieces. The hoops graduate in thickness from\n3 mm at the ends to 6
                   mm in the center, with approximately 29 mm overall\ndiameter."
10          },
11          {
12              "title":"14-kt. Jade Doughnut Pendant",
13              "listPrice":"$150.00",
14              "price":"$48.99",
15              "saving":"$101.01",
16              " saving_percentage":"67%",
17              "content":"The 25-mm disk hangs delicately\nfrom a 14-karat gold chain.
                   The disk features a dramatic gold Chinese character\nin the center,
                   accompanied by four stylized gold bees."
18          },
19          {
20              "title":"14-kt. Charcoal Jade and Ruby Elephant Pendant",
21              "listPrice":"$100.00",
22              "price":"$28.99",
23              "saving":"$71.01",
24              "saving_percentage":"71%",
25              "content":"Carved of rich dark grey jade,\nthis elephant pendant has 14-
                   karat yellow gold app lied to mark the feet,\ntusk, tail, and blanket.
                   A 2-mm round faceted ruby in a gold bezel setting\nforms the eye. The
                   pendant hangs from an 18-inch chain."
26          },
27          {
28              "title":"14-kt. Carved Lavender Jade Earrings",
29              "listPrice":"$80.0 0",
30              "price":"$39.99",
31              "saving":"$40.01",
32              "saving_percentage":"50%",
33              "content":"Luscious 8-mm lavender jade balls, carved with intricate Asian
                   style, dangle from a 14-karat yellow gold French hook."
34          },
35          {
36              "title":"14-kt. Jade C ross Pendant",
37              "listPrice":"$150.00",
38              "price":"$49.99",
39              "saving":"$100.01",
40              "saving_percentage":"66%",
41              "content":"Green jade and gold create this\nbeautiful cross pendant.
                   Cylindrical bars of green jade feature caps and\nce nter of 14-karat
                   yellow gold."
42          },
43          {
44              "title":"14-kt. Multicolored Jade Earrings",
```

```
45        "listPrice":"$375.00",
46        "price":"$99.99",
47        "saving":"$275.01",
48        "saving_percentage":"73%",
49        "content":"A delicate wrapping of 14-karat\nyellow gold wire holds six 6 x
              4 pear shapes of jade in various shades: brilliant\ngreen, orange,
              lavender, black, pale yellow, and white. The post earrings\nhave
              butterfly backs."
50      },
51      {
52        "title":"14-kt. Multicolored Jade Ring",
53        "listPrice":"$250.00",
54        "price":"$56.99",
55        "saving":"$193.01",
56        "saving_percentage":"77%",
57        "content":"A delicate wrapping of 14-karat\nyellow gold wire holds six 6 x
              4 ovals of jade in various shades: brilliant\ngreen, orange, lavender,
              bl ack, pale yellow, and white. A narrow gold band\ndivides to support
              the setting."
58      },
59      {
60        "title":"14-kt. Onyx and Ruby Elephant Pendant",
61        "listPrice":"$100.00",
62        "price":"$35.99",
63        "saving":"$64.01",
64        "saving_percentage":"64%",
65        "c ontent":"Carved of rich black onyx, this\nelephant pendant has 14-karat
              yellow gold applied to mark the feet, tusk,\ntail, and blanket. A 2-mm
              round faceted ruby in a gold bezel setting forms\nthe eye. The pendant
              hangs from an 18-inch chain."
66      }
67    ]
68 }
```

xPath Enaa sandisk

```
1  {
2    "title":"SanDisk Ultra microSDXC 128GB 100MB/s Class 10 UHS-I – SDSQUNR-128G-
         GN3MN",
3    "price":"11,88 \u20ac",
4    "rating":"4,83",
5    "reviewCount":"96",
6    "brand":"Spominske kartice in \u010ditalci  SanDisk",
7    "description":"Ustvar jen za prihranek trenutka Kartica SanDisk Ultra\u00ae
         microSD\u2122 UHS-I vam daje svobodo snemanja, shranjevanja in deljenja ve\
         u010d kot kadar koli prej. S kapaciteto do 512 GB1 ima na\u0161a kartica
         SanDisk Ultra microSD\\u212 2 prostor za \u0161e ve\u010d ur videa v polni
         visoki lo\u010dljivosti3 in zagotavlja hitrosti prenosa do 100 MB/s2, kar vam
          pomaga hitro premikati to vsebino. Deluje s pametnimi telefoni in tabli\
         u010dnimi ra\u010dunalniki Andr oid\u2122 Kartice SanDisk\u00ae so zdru\
         u017eljive s pametnimi telefoni in tablicami Android. Zajemite in shranite \
         u0161e ve\u010d ur videa v polni visoki lo\u010dljivosti S kapaciteto do 512
         GB1 lahko na kartico shranite &sc"
8  }
```

xPath Enaa samsung

```
1  {
2    "title":"Spominska kartica Samsung EVO Plus, micro SDXC, 256GB, U3, V30, A2, UHS
         -I, z SD adapterjem",
3    "price":"22,66 \u20ac",
4    "rating":"4,85",
5    "reviewCount":"79",
6    "brand":"Spominske kartice in \u010ditalci  Samsung",
7    "desc ription":"Za vse naprave Spominske kartice Samsung Evo Plus so primerne za
          shranjevanje slik, video posnetkov in aplikacij na telefonih in tablicah,
         uporabne pa so tudi v igralnih konzolah. Hitrost in zanesljivost Evo Plus
         spom inska kartica z lahkoto odpira tudi ve\u010dje datoteke s hitrostjo
         branja do 130 MB/s. Visoke kapacitete Na 256 GB lahko shranite do 15 ure 4K
         video posnetkov, 40 ur FullHD videov. Ostale lastnosti spominska kartica tipa
          microS DXC primerna za uporabo v pametnih telefonih, tabli\u010dnih ra\
         u010dunalnikih, ro\u010dnih igralnih konzolah, kamerah in fotoaparatih. UHS-I
          vodilo kapaciteta 256 GB masa: pribli\u017eno\u00a00,25 g hitrostni razredi:
          U3"
8  }
```

RoadRunner rtv

```
 1  <a> Registracija Pozabljeno geslo?</a>
 2  <form> <div class="form-group"> <label for="user">Uporabni ko ime:</label> </div> <div class="form-group"> <label for=
        "pass">Geslo:</label> </div> <div class="form-group"> <button class="btn btn-primary" type="submit">Prijava</
        button> <a class="register" href="https://moj.rtvslo.si/registracija">Registracija</a> </div></form>
 3  <a> Rekorderska steza na slani ravnici Bonneville potrebuje re itev</a>
 4  <a> Otro ki portal Moja generacija Moj splet Teletext</a>
 5  <a> Skupine Iskalnik uporabnikov Pravice uporabnikov Moderirani uporabniki</a>
 6  C:\Users\Marija\Desktop\faks\mag1\2. semester\ieps\pa2\IEPS-Assignments\Programming_assignment_2\run-extraction.py:434:
        MarkupResemblesLocatorWarning: The input looks more like a filename than markup. You may want to open this file
        and pass the filehandle into Beautiful Soup.
 7    soup = BeautifulSoup(match, 'html.parser')
 8  <a> Preusmeritve / aliasi Meniji Uporabniki Skupine Iskalnik uporabnikov Pravice uporabnikov Moderirani uporabniki
        Vsebine Administracija novic Administracija sekcij Administracija slike & video dneva Administracija anket
        Administracija nujnih informacij Administracija mailerja File upload</a>
 9  <a> V ivo  Arhiv Spored Oddaje</a>
10  <a> Rekorderska steza na slani ravnici Bonneville potrebuje re itev</a>
11  <a> Odjava Uporabni ki ra un </a>
12  <a> Ford kuga je z zamenjavo generacije postala tudi priklju ni hibrid</a>
13  <a> Gospodarstvo Lokalne novice  rna  kronika Zdravje Okolje Znanost in tehnologija Slovenci v sosednjih dr avah </a>
14  <a> Registracija</a>
15  <a> Komentarji <span class="numComments">(13)</span></a>
16  <a> RTVSLO.si Slovenija Svet  port  Kultura  ivljenjski  slog Svet zabave</a>
17  <a> <span class="bottom-tooltip">Prijava</span> Registracija Pozabljeno geslo?</a>
18  <form> <div class="tbformgroup"> <label for="user">Uporabni ko ime:</label> </div> <div class="tbformgroup"> <label
        for="pass">Geslo:</label> </div> <div class="tbformgroup"> <button class="tbbtn" type="submit">Prijavi</button> <a
        class="register" href="https://moj.rtvslo.si/registracija">Registracija</a> </div></form>
19  <a> Registracija Pozabljeno geslo?</a>
20  <form> <div class="form-group"> <label for="user">Uporabni ko ime:</label> </div> <div class="form-group"> <label for=
        "pass">Geslo:</label> </div> <div class="form-group"> <button class="btn btn-primary" type="submit">Prijava</
        button> <a class="register" href="https://moj.rtvslo.si/registracija">Registracija</a> </div></form>
21  <a> Pripravite se na ofenzivo priklju nih hibridov</a>
22  <a> Film in TV Glasba Knjige Oder Vizualna umetnost Arhitektura in oblikovanje D edi  ina Recenzije Drugo Jezikovni
        spletovalec 1. svetovna vojna</a>
23  <a> Pripravite se na ofenzivo priklju nih hibridov</a>
24  <a> Bentley se je ob stoletnici spomnil blowerja, enega najbolj ikonskih avtomobilov iz dirka ke preteklosti</a>
25  <a>  enske  so po kodbam v prometnih nesre ah izpostavljene bolj kot mo ki </a>
26  <a> Testi</a>
27  <a> Odjava Uporabni ki ra un </a>
28  <a>  enske  so po kodbam v prometnih nesre ah izpostavljene bolj kot mo ki </a>
29  <a> Evrovizija Glasba Film in TV Svet znanih Dru abno Zanimivosti</a>
30  <a> Info Kultura Filmi/serije Razvedrilo Otroci Izobra evanje Dokumentarci Duhovnost</a>
31  <a> Ve   Nastavitve <span>Cookie Consent plugin for the EU cookie law</span> Allow for all sites Strinjam se</a>
32  <a> Bentley se je ob stoletnici spomnil blowerja, enega najbolj ikonskih avtomobilov iz dirka ke preteklosti</a>
33  <a> Evropska unija Evropa S. in J. Amerika Bli nji vzhod Afrika Azija z Oceanijo</a>
34  <a> Komentarji <span class="numComments">(13)</span> Uredi Prika i komentarje</a>
35  <a> Odjava Uporabni ki ra un </a>
36  <a> Prika i komentarje</a>
37  <a> Rekorderska steza na slani ravnici Bonneville potrebuje re itev</a>
38  <a> Prijavi napako</a>
39  <a> Ve   Nastavitve <span>Cookie Consent plugin for the EU cookie law</span> Allow for all sites Strinjam se</a>
40  <a> Odjava Uporabni ki ra un </a>
41  <a> Slovenija Gospodarstvo Lokalne novice  rna  kronika Zdravje Okolje Znanost in tehnologija Slovenci v sosednjih
        dr avah Svet Evropska unija Evropa S. in J. Amerika Bli nji vzhod Afrika Azija z Oceanijo  port  Nogomet
        Ko arka Rokomet Odbojka Hokej Tenis Atletika Zimski  porti  Moto porti F1 Kolesarstvo Preostali  porti
        Rekreacija  portni  SOS Kultura Film in TV Glasba Knjige Oder Vizualna umetnost Arhitektura in oblikovanje
        D edi  ina Recenzije Drugo Jezikovni spletovalec 1. svetovna vojna  ivljenjski  slog Ture avanture Kulinarika
        Lepota bivanja Avtomobilnost Moda 196x ljubezen Svet zabave Evrovizija Glasba Film in TV Svet znanih Dru abno
        Zanimivosti Posebna izdaja MMC Podrobno Klepet Podkast  tevilke  1968 Kolumne MMC Teden MMC Analiza P razni no
        leto RTV Slovenija Info Vreme S ne ne razmere Stanje na cestah TV Slovenija Info Kultura Filmi/serije Razvedrilo
        Otroci Izobra evanje Dokumentarci Duhovnost Radio Slovenija Prvi Val 202 ARS Radio Koper Radio Capodistria Radio
        Maribor Radio Si MMR RTV 4D V ivo  Arhiv Spored Oddaje Tematski portali Otro ki portal Moja generacija Moj splet
        Teletext O RTV Slovenija Kdo smo RTV-prispevek Za medije Kontakti in informacije</a>
42  <a> Ford kuga je z zamenjavo generacije postala tudi priklju ni hibrid</a>
43  <a> Slovenija Svet  port  Kultura  ivljenjski  slog Svet zabave</a>
44  <a> 4D Spored V ivo  Arhiv</a>
45  <a> Bentley se je ob stoletnici spomnil blowerja, enega najbolj ikonskih avtomobilov iz dirka ke preteklosti</a>
46  <a> Vreme S ne ne razmere Stanje na cestah</a>
47  <a> Registracija Pozabljeno geslo?</a>
48  <form> <div class="form-group"> <label for="user">Uporabni ko ime:</label> </div> <div class="form-group"> <label for=
        "pass">Geslo:</label> </div> <div class="form-group"> <button class="btn btn-primary" type="submit">Prijava</
        button> <a class="register" href="https://moj.rtvslo.si/registracija">Registracija</a> </div></form>
```

RoadRunner Overstock

1 `<table> <tbody><tr><td valign="top"><table> <tbody><tr><td align="right" nowrap="nowrap">List Price:</td><td align="left" nowrap="nowrap"><s>$305.00</s></td></tr> <tr><td align="right" nowrap="nowrap">Price:</td><td align="left" nowrap="nowrap">$119.99</td></tr> <tr><td align="right" nowrap="nowrap">You Save:</td><td align="left" nowrap="nowrap">$185.01 (60%)</td></tr> </tbody></table></td></tr></table>`

2 `<tbody> <tr><td valign="top"><table> <tbody><tr><td align="right" nowrap="nowrap">List Price:</td><td align="left" nowrap="nowrap"><s>$305.00</s></td></tr> <tr><td align="right" nowrap="nowrap">Price:</td><td align="left" nowrap="nowrap">$119.99</td></tr> <tr><td align="right" nowrap="nowrap">You Save:</td><td align="left" nowrap="nowrap">$185.01 (60%)</td></tr> </tbody></table></td></tr></tbody>`

3 `<tr> <td valign="top"><table> <tbody><tr><td align="right" nowrap="nowrap">List Price:</td><td align="left" nowrap="nowrap"><s>$305.00</s></td></tr></tbody></table></td> <td align="right" nowrap="nowrap">Price:</td><td align="left" nowrap="nowrap">$119.99</td> <td align="right" nowrap="nowrap">You Save:</td><td align="left" nowrap="nowrap">$185.01 (60%)</td></tr>`

4 `<td> <table> <tbody><tr><td align="right" nowrap="nowrap">List Price:</td></tr></tbody></table> <s>$305.00</s> Price: $119.99 You Save: $185.01 (60%) Over a cleanly sculpted Roman cross of 14-karat white gold drapes a slender banner containing three bright prong-set round diamonds (0.06 total carat weight, H-I color, I clarity).
Click here to purchase.
</td>`

5 `<a> <span class="tiny"><b>Click here to purchase.</b></span></a>`

6 `<table> <tbody><tr><td> <table bgcolor="#d2dbfb" border="0" cellpadding="0" cellspacing="5" width="100%"> <tbody><tr><td align="right" nowrap="nowrap"> List Sorted By:</td><td nowrap="nowrap"> Top Sellers | Discount | Newest First | Price | Quantity | Markdowns</td></tr><tr><td align="right" nowrap="nowrap"> Items:</td><td nowrap="nowrap"> 16 - 30 of 296 | First Page | Previous 15 | Next 15 </td></tr> </tbody></table></td></tr></tbody></table>`

7 `<tbody> <tr><td> <table bgcolor="#d2dbfb" border="0" cellpadding="0" cellspacing="5" width="100%"> <tbody><tr><td align="right" nowrap="nowrap"> List Sorted By:</td><td nowrap="nowrap"> Top Sellers | Discount | Newest First | Price | Quantity | Markdowns</td></tr><tr><td align="right" nowrap="nowrap"> Items:</td><td nowrap="nowrap"> 16 - 30 of 296 | First Page | Previous 15 | Next 15 </td></tr> </tbody></table></td></tr></tbody>`

8 `<tr> <td> <table bgcolor="#d2dbfb" border="0" cellpadding="0" cellspacing="5" width="100%"> <tbody><tr><td align="right" nowrap="nowrap"> List Sorted By:</td><td nowrap="nowrap"> Top Sellers | Discount | Newest First | Price | Quantity | Markdowns</td></tr></tbody></table></td> <td align="right" nowrap="nowrap"> Items:</td><td nowrap="nowrap"> 16 - 30 of 296 | First Page | Previous 15 | Next 15 </td></tr>`

9 `<td> <table bgcolor="#d2dbfb" border="0" cellpadding="0" cellspacing="5" width="100%"> <tbody><tr><td align="right" nowrap="nowrap"> List Sorted By:</td></tr></tbody></table> Top Sellers | Discount | Newest First | Price | Quantity | Markdowns Items: 16 - 30 of 296 | First Page | Previous 15 | Next 15</td>`

10 `<a> Discount Newest First Price Quantity Markdowns First Page Previous 15 Next 15</a>`

RoadRunner Enaa

```
1   <a> Spominska kartica Kingston Canvas Select Plus microSDXC, 128 GB, 100 MB/s</a>
2   <form> <button class="btn btn-primary" id="addtobasket" type="submit"> <span class="material-icons">   </span> <span
        class="add-to-cart-btn-text">V vozi ek</span> </button></form>
3   <a> <span class="material-icons">favorite</span> Shrani na seznam izdelkov Dostava iz zaloge <span class="material-
        icons">email</span>Po lji in deli</a>
4   <form> <div class="single-product-side" itemprop="offers" itemscope="" itemtype="http://schema.org/Offer"> <div class="
        add-to-cart-button"> <button class="btn btn-lg btn-primary" id="submitButtonBasket" type="submit"> <span class="
        material-icons">   </span> <span class="add-to-cart-btn-text">Dodaj v vozi ek</span> </button> <div class="add-to
        -chart-text-2">nakup tudi brez registracije</div> </div> <div class="single-product-side-wishlist-button"> <a
        class="btn btn-outline-gray active" data-bs-content="Za shranjevanje izdelkov se morate prijavit v Enaa" data-bs-
        original-title="" data-bs-placement="bottom" data-bs-toggle="popover" data-bs-trigger="hover focus" href="
        javascript:void(0);" title=""> <span class="material-icons">favorite</span> Shrani na seznam izdelkov </a> </div>
        <hr/> <div class="single-product-side-info-item large green" id="
        avalMsg"> <span class="material-icons">check</span> <div class="single-product-side-info-item-text"><strong><a
        class="green" data-bs-content="Izdelek je na zalogi v enem izmed na ih skladi  ali skladi  pogodbenih
        dobaviteljev." data-bs-original-title="" data-bs-placement="bottom" data-bs-toggle="popover" data-bs-trigger="
        hover focus" href="javascript:Void(0);" title="">Dostava iz zaloge</a></strong></div> </div> <div class="sale-info
        -dobava" id="productDobavljivost"><div aria-describedby="popover206999" class="single-product-side-info-item" data
        -bs-content="Najhitreja dostava v Sloveniji. Po oddaji naro ila dostavni slu bi prejmete elektronsko po to in
         SMS sporo ilo ( e ste vnesli va o mobilno  tevilko  za obve anje o dostavi). Po iljko lahko na va em
        naslovu v Sloveniji pri akujete naslednji dan po prejemu sporo ila. Zaradi razli nih vzrokov na katere ne
        moremo vplivati: okvare, nesre e, napake, razmere na cesti, ... tudi kak en dan kasneje. " data-bs-original-
        title="" data-bs-placement="bottom" data-bs-toggle="popover" data-bs-trigger="hover focus" title=""><span class="
        material-icons">local_shipping</span><div class="single-product-side-info-item-text">Izdelek bomo predvidoma
        poslali do <strong class="green text-nowrap">6. 5. 2024</strong></div></div><div class="single-product-side-info-
        item" data-bs-content="Prevzem v nekaj sekundah na Enaa DriveIn ali hiter prevzem na  alterju  z zagotovljenim
        parkingom vsak delovnik od ponedeljka do sobote od 9:00 do 18:00 ure. Izberete lahko tudi non-stop 24 ur prevzem
        na Enaa paketomatih. Ko je naro ilo pripravljeno za izdajo, prejmete elektronsko po to in SMS sporo ilo ( e
        ste vnesli va o mobilno  tevilko  za obve anje o dostavi).  " data-bs-original-title="" data-bs-placement="
        bottom" data-bs-toggle="popover" data-bs-trigger="hover focus" title=""><span class="material-icons">
        local_shipping</span><div class="single-product-side-info-item-text">Brezpla en prevzem v Enaa centru, Savska 3a,
         Ljubljana predvidoma od <span class="green text-nowrap"><b>6. 5. 2024</b></span></div></div></div> <div class="
        single-product-side-bottom"> <p><strong> ifra  izdelka:</strong> 2l18ygtey32</p> <p> <b>S to kami zvestobe do
        popustov:</b> 1 to ko. </p> <a class="btn btn-sm btn-outline-gray" href="mailto:?subject=SanDisk Ultra microSDXC
        128GB 100MB/s Class 10 UHS-I - SDSQUNR-128G-GN3MN | enaA.com&amp;body=Zdravo, %0D%0A%0D%0ATule na Enaa sem na el
        eno res dobro in zanimivo ponudbo, ki jo priporo am:%0D%0A%0D%0A https://www.enaa.com/micro-sd-kartice/sandisk-
        ultra-microsdxc-128gb-100mb-s-class-10-uhs-i-sdsqunr-128g-gn3mn" id="poslji"> <span class="material-icons">email</
        span>Po lji in deli</a> </div> </div> </div></form>
5   <a> Pomnilni ka kartica Kingston Canvas Select Plus 128 GB microSDXC, UHS-I U1, Class 10, V10, A1,  rna </a>
6   <a> SanDisk 128GB Ultra microSDXC 100MB/s Class 10 UHS-I - SDSQUNR-128G-GN6MN</a>
7   <a> Enaa.com</a>
8   <a> SanDisk Ultra microSDXC 128GB + SD Adapter 100MB/s Class 10 UHS-I - SDSQUNR-128G-GN3MA</a>
9   <a> SanDisk Ultra microSDXC 128GB + SD Adapter 100MB/s Class 10 UHS-I - SDSQUNR-128G-GN3MA</a>
10  <a> SanDisk 128GB Ultra microSDXC 100MB/s Class 10 UHS-I - SDSQUNR-128G-GN6MN</a>
11  <a> Pomnilni ka kartica Patriot EP 128 GB microSDXC UHS-I U3, Class 10, V30, A1,  rna /rde a </a>
12  <a> SanDisk Ultra microSDXC 128GB + SD Adapter 100MB/s Class 10 UHS-I - SDSQUNR-128G-GN3MA</a>
13  <a> Intenso 128GB microSDXC UHS-I Class 10 Pro 90MB/s spominska kartica - 3433491</a>
14  <a> Pomnilni ka kartica Kingston Canvas Select Plus 128 GB microSDXC, UHS-I U1, Class 10, V10, A1,  rna </a>
15  <a> APACER microSD XC 128GB spominska kart. Class 30 Gaming AP128GMCSX10U7-RAGC</a>
16  <form> <button class="btn btn-primary" id="addtobasket" type="submit"> <span class="material-icons">   </span> <span
        class="add-to-cart-btn-text">V vozi ek</span> </button></form>
17  <a> SanDisk 128GB Ultra microSDXC 100MB/s Class 10 UHS-I - SDSQUNR-128G-GN6MN</a>
18  <a> Pomnilni ka kartica Kingston High Endurance 32 GB microSDHC UHS-I U1, Class 10</a>
19  <a> SanDisk Ultra microSDXC 128GB 100MB/s Class 10 UHS-I - SDSQUNR-128G-GN3MN</a>
20  <a> Spominska kartica Kingston Canvas Select Plus microSDXC, 128 GB, 100 MB/s</a>
21  <a> Spominska kartica Kingston Industrial 8GB microSDHC UHS-I U3, Class 10, V30, A1,  rna </a>
22  <a> Enaa.com</a>
23  <a> SanDisk Ultra microSDXC 128GB 100MB/s Class 10 UHS-I - SDSQUNR-128G-GN3MN</a>
24  <a> Spominska kartica Kingston Canvas Select Plus microSDXC, 128 GB, 100 MB/s</a>
25  <a> Pomnilni ka kartica Kingston High Endurance 32 GB microSDHC UHS-I U1, Class 10</a>
26  <a> Dostava iz zaloge <span class="material-icons">email</span>Po lji in deli</a>
27  <a> Pomnilni ka kartica Patriot EP 128 GB microSDXC UHS-I U3, Class 10, V30, A1,  rna /rde a </a>
28  <a> <span class="material-icons">favorite</span> Shrani na seznam izdelkov Dostava iz zaloge <span class="material-
        icons">email</span>Po lji in deli</a>
29  <a> <span class="material-icons">email</span>Po lji in deli</a>
30  <a> Spominska kartica Kingston Industrial 8GB microSDHC UHS-I U3, Class 10, V30, A1,  rna </a>
31  <a> Enaa.com</a>
32  <a> Pomnilni ka kartica Patriot EP 128 GB microSDXC UHS-I U3, Class 10, V30, A1,  rna /rde a </a>
33  <form> <button class="btn btn-primary" id="addtobasket" type="submit"> <span class="material-icons">   </span> <span
        class="add-to-cart-btn-text">V vozi ek</span> </button></form>
34  <a> Spominska kartica Kingston Industrial 8GB microSDHC UHS-I U3, Class 10, V30, A1,  rna </a>
35  <form> <button class="btn btn-primary" id="addtobasket" type="submit"> <span class="material-icons">   </span> <span
        class="add-to-cart-btn-text">V vozi ek</span> </button></form>
36  <a> Intenso 128GB microSDXC UHS-I Class 10 Pro 90MB/s spominska kartica - 3433491</a>
37  <a> Pomnilni ka kartica Kingston High Endurance 32 GB microSDHC UHS-I U1, Class 10</a>
38  <a> Spominske kartice in  italci  SanDisk</a>
39  <a> SanDisk 128GB Ultra microSDXC 100MB/s Class 10 UHS-I - SDSQUNR-128G-GN6MN</a>
40  <a> Spominska kartica Kingston Canvas Select Plus microSDXC, 128 GB, 100 MB/s</a>
41  <a> Pomnilni ka kartica Kingston High Endurance 32 GB microSDHC UHS-I U1, Class 10</a>
42  <a> Enaa Fotoaparati Spominske kartice in  italci  Micro SD kartice</a>
43  <a> SanDisk Ultra microSDXC 128GB 100MB/s Class 10 UHS-I - SDSQUNR-128G-GN3MN</a>
```