

2. seminarska naloga pri predmetu iskanje in ekstrakcija podatkov s spleta

Anže Gregorc, Jaka Kordež

Maj 2019

1 Uvod

V okviru druge seminarske naloge je bilo potrebno izdelati program za pridobivanje informacij iz znanih spletnih strani na tri različne načine. Poleg dveh primerkov podanih spletnih strani iz naslova rrvslo.si in overstock.com sva izbrala še dve časovnici iz Twitterja. V nadaljevanju bodo opisane uporabljene tehnologije in podrobnosti implementacije.

2 Twitter spletna stran

Kot omenjeno sva izbrala dve Twitter spletni strani. Ta vsebuje seznam twittov z različnimi vsebinami. S teh dveh straneh sva pridobila naslednje elemente: "Handle", "Content", "Timestamp", "Replies", "Retweets" in "Favourites". Na sliki 1 so prikazane pozicije vseh teh elementov.



Figure 1: Twitter.com vzorec spletne strani

3 Implementacija

Tako kot pri prvi nalogi sva se tudi tokrat odločila za uporabo Visual C# in programskega okolja .NET Core. Podpora regularnim izrazom je že vgrajena v standardno knjižnico, za uporabo jezika XPath sva uporabila knjižnico HtmlAgilityPack, algoritem Road runner pa sva implementirala sama.

3.1 Regularni izrazi

Regularni izrazi nam omogočajo enostavno in hitro poizvedovanje po besedilu. Vsak izraz je predstavljen z nizom, rezultat poizvedbe pa je množica ujemanj. Pri portalu rtslo.si je bilo dokaj enostavno zapisati potrebne izraze. Naslov članka se npr. vedno pojavi v znački h1, vsak naslednji element (podnaslov, avtor, ...) pa vsebuje atribut class z ustreznim imenom. Bistveno težja je bila naloga pri strani overstock.com. Stran je namreč izdelana z velikim številom tabel, ki se med seboj težko razlikujejo. Na koncu sva si najbolj pomagala z besedilom za ob vsakem artiklu ("List Price"). Tudi twitter.com ima dokaj lepo strukturirano časovnico, vendar pa se določeni elementi pojavijo na strani večkrat. Problem je bil na primer pri identifikatorju uporabnika (handle), ki se poleg vsakega tvita pojavi tudi na desni med predlogi za sledenje (Who to follow). Nekoliko bolj natančen regularni izraz je nato zadostoval tudi za te izjeme.

3.2 XPath

Jezik XPath je posebej namenjen iskanju po strukturiranih dokumentih tipa XML. Iskanje z njim je bilo bistveno lažje kot z regularnimi izrazi, saj sistem "pozna" XML elemente in zato ignoriranje vmesnih presledkov in novih vrstic ni potrebno. Spet je bila najenostavnejša implementacija za rtslo.si in twitter.com, medtem ko sva si pri overstock.com spet pomagala z besedilom v elementih.

3.3 Road Runner

Road Runner je metoda, ki vrne človeku-berljivo ovojnico, ki jo lahko uporabimo za nadaljnjo ekstrakcijo podatkov pri podobnih spletnih straneh. Naša ovojnica uporablja tako imenovano union-free regularni izraz.

Implementacija Road Runnerja se predvsem zanaša na predstavitev in prosojnico iz predavanj. Torej išče tri različne tipe neujemanj med spletnima stranema: tekstovno neujemanje, opcijski element in iterator. Če algoritem najde tekstovno neujemanje, se na standardni izhod izpiše #text. V primeru opcijskega elementa (tag), se le ta ovije z oklepajema ter na koncu doda ?. Pri iteratorju pa se element ovije okoli oklepajema ter na koncu +. Iterator tudi poišče vse ostale elemente, ki sodijo v iterator ter jih odstrani iz rešitve (regularnega izraza). Dodali pa smo tudi preverjanje atributov znotraj elementov. V primeru, da je atribut opcijski, se ga temu primerno ovije z oklepajema ter ?. Če pa se atribut vsakič pojavi, ampak je vrednost drugačna, pa se vrednost spremeni na #text.

Ker naše spletne strani vsebujejo ogromno elementov, smo za lažje razpoznavanje rezultatov strani nekoliko preprocesirali. Poskusili smo odstraniti vse neuporabne elemente. Neuporabni elementi so bili za nas tisti elementi, ki niso vsebovali nobenega podatka, ki ga mi pridobivamo. Eliminacijo smo izvedli ročno za vsako domeno posebej. Pri Overstock.com smo eliminirali dele, ki so služili le kot odmik med vrsticam. Pri Rtslo.si smo eliminirali le nekaj manjših

delov, kar bi lahko označili kot nepomemben popravek. Pri Twitter.com pa smo vzeli le element, ki je vseboval tabelo twittov.

Kljub preprocesiranju pa je nastalo mnogo podatkov, ki bi jih ekstrakcija s pomočjo naše rešitve zajela. Vsak element, ki je podoben, ampak vsebuje različno vsebino bi bil zajet. Tako, da sva se osredotočila bolj na to, da so zajeti vsi podatki, ki so za nas zanimivi. Pri Overstock.com je algoritem lepo našel element, ki se ponavlja in ga ovil v iterator. Zato nam je pri tej spletni strani uspelo zajeti vse podatke, ki jih pridobivamo, tudi zaradi tega, ker so vrednosti pomembnih elementov vsebovali drugačne podatke. Pri Rtvsl.si pa nam na žalost ni uspelo najti regularnega izraza, s katerim bi lahko ekstrahirali vse podatke. Problem je pri tem, da je avtor obeh spletnih strani ena oseba. Ker se vrednost elementa, pri katerem je zapisan avtor ni spremenila, ga naš algoritem ni zaznal kot pomembnega ter ga ni označil kot `#text`. Pri Twitter.com pa smo zopet imeli to srečo, da smo našli vse željene podatke.

3.4 Psevdo koda Road Runner algoritma

Prikazana je psevdo koda algoritma RoadRunner. Algoritem je kar precej obširen, zato je prikazana samo glavna funkcija tega algoritma. Ta funkcija sprejme dva elementa, ki sta korena dreves, ki vsebujeta vse elementa html datoteke. Funkcija je rekurzivna, saj je potrebno drevo generalizirati tudi v globino.

4 Zaključek

Z našim delom smo zadovoljni. Pri tej nalogi smo se veliko naučili. Spoznali smo različne načine, kako pridobivati podatke iz spletnih strani. To znanje je trenutno zelo uporabno v realnem svetu, kar nam je dalo še dodatnega zagona za delo na tem projektu.

5 Dodatek

V dodatku prilagamo še vse izhode našega programa. Nadalnja podpoglavja bodo vsebovala rezultate, ki nam jih določena metoda vrne pri neki spletni strani. Žal je podatkov zelo veliko in je zaradi tega onemogočen lep prikaz rezultatov naših metod.

5.1 Izhod Regularnega izraza pri Overstock.com 1.

```
1  [
2    {
3      "Title": "10-kt. Seven Diamond Ladies Heart Ring (0.08 Tw)",
4      "ListPrice": "$149.00",
5      "Price": "$69.99",
```

Algorithm 1 RoadRunner

```
1: procedure GENERALIZE(baseElement, element2) oba elementa sta drevesi,  
   ki vsebujeta elemente html dokumenta  
2:    $idx2 \leftarrow 0$   
3:   for baseOtrok in baseElement.otroci do  
4:      $otrok2 \leftarrow element2.otrok[idx2]$   
5:     while baseOtrok ni enak kot otrok2 do  
6:       if prejsnjiBaseOtrok je enak kot otrok2 then  
7:         Generalize(prejsnjiBaseOtrok, otrok2);  
8:         Ovij prejsnjiBaseOtrok z Iteratorjem;  
9:         Odstranjuj prejšnje otroke, ki spadajo v ta iterator;  
10:      else  
11:        Ovij baseOtrok z opsijskim elementom;  
12:      end if  
13:       $idx2 \leftarrow idx2 + 1$   
14:    end while  
15:    if  $idx2 \geq \text{število otrok elementa2}$  then  
16:      Ovij baseOtrok z opsijskim elementom;  
17:    else  
18:      if baseOtrok in otrok2 sta Tekstovna elementa then  
19:        baseOtrok generaliziraj z otrok2;  
20:      else if baseOtrok in otrok2 sta Html elementa  
21:        baseOtrok generaliziraj z otrok2;      ▷ generaliziraj attribute  
22:        elementov  
23:        Generalize(baseOtrok, otrok2);      ▷ rekurzivno generaliziraj  
24:        elementa  
25:      end if  
26:    end if  
27:    end forendfor  
28:    while  $idx2 < \text{število otrok elementa2}$  do      ▷ Procesiraj še preostale  
29:      otroke iz element2  
30:       $otrok2 \leftarrow element2.otrok[idx2]$   
31:      if zadnjiBaseOtrok je enak kot  $otrok2$  then  
32:        Generalize(zadnjiBaseOtrok,  $otrok2$ );  
33:        Ovij zadnjiBaseOtrok z Iteratorjem;  
34:        Odstranjuj prejšnje otroke, ki spadajo v ta iterator;  
35:      else  
36:        Ovij  $otrok2$  z opsijskim elementom ter ga dodaj k baseElement;  
37:      end if  
38:       $idx2 \leftarrow idx2 + 1$   
39:    end while  
40:  end procedure
```

```

6      "Saving": "$79.01",
7      "SavingPercent": "(53%)",
8      "Content": "This ladies fashion ring dazzles\r\nwith hearts and diamonds. The gold
9  },
10 {
11     "Title": "10-Kt. Diamond Ring (.25 TW)",
12     "ListPrice": "$250.00",
13     "Price": "$74.90",
14     "Saving": "$175.10",
15     "SavingPercent": "(70%)",
16     "Content": "Nineteen round diamonds accent this 10-karat yellow gold ring with fil
17 },
18 {
19     "Title": "10-kt. Pearl and Diamond Butterfly Earrings",
20     "ListPrice": "$149.00",
21     "Price": "$42.99",
22     "Saving": "$106.01",
23     "SavingPercent": "(71%)",
24     "Content": "Perfectly proportioned 5.5- to\r\n6-mm cultured pearls on 10-karat yel
25 },
26 {
27     "Title": "14-kt. Diamond 'S' Tennis Bracelet (2.00 TW)",
28     "ListPrice": "$1,539.99",
29     "Price": "$499.99",
30     "Saving": "$1,040.00",
31     "SavingPercent": "(67%)",
32     "Content": "Invest in a swirl of light with\r\nthis diamond 'S' tennis bracelet. C
33 },
34 {
35     "Title": "10-kt. Diamond Band Fashion Ring (.11 TW)",
36     "ListPrice": "$179.99",
37     "Price": "$79.99",
38     "Saving": "$100.00",
39     "SavingPercent": "(55%)",
40     "Content": "Crafted in white and yellow gold,\r\nthis ring displays a band of seve
41 },
42 {
43     "Title": "14-kt. White Gold, Pearl and Diamond Ring",
44     "ListPrice": "$419.99",
45     "Price": "$149.99",
46     "Saving": "$270.00",
47     "SavingPercent": "(64%)",
48     "Content": "Show your romantic side with this\r\n14-karat diamond and pearl ring.
49 },
50 {
51     "Title": "14-kt. Gold Diamond Present Future Pendant (.25TW)",

```

```

52     "ListPrice": "$299.00",
53     "Price": "$149.99",
54     "Saving": "$149.01",
55     "SavingPercent": "(49%)",
56     "Content": "Designed with three large, sparkling\r\nndiamonds to represent past, pr
57 },
58 {
59     "Title": "14-kt. Diamond Solitaire Pendant (.33 TW)",
60     "ListPrice": "$1,019.99",
61     "Price": "$319.99",
62     "Saving": "$700.00",
63     "SavingPercent": "(68%)",
64     "Content": "In this simple, yet elegant pendant,\r\na round brilliant diamond (0.3
65 },
66 {
67     "Title": "14-kt. Diamond Solitaire Earrings (0.33 TW)",
68     "ListPrice": "$639.99",
69     "Price": "$199.99",
70     "Saving": "$440.00",
71     "SavingPercent": "(68%)",
72     "Content": "Dazzle your way into her heart,\r\nwith these classic diamond solitair
73 },
74 {
75     "Title": "14-kt. Diamond Cross Pendant (.06 TW)",
76     "ListPrice": "$305.00",
77     "Price": "$119.99",
78     "Saving": "$185.01",
79     "SavingPercent": "(60%)",
80     "Content": "Over a cleanly sculpted Roman\r\ncross of 14-karat white gold drapes a
81 },
82 {
83     "Title": "14-kt. Diamond Solitaire Stud Earrings (.50 TW)",
84     "ListPrice": "$999.99",
85     "Price": "$359.99",
86     "Saving": "$640.00",
87     "SavingPercent": "(64%)",
88     "Content": "Every jewelry collection needs\r\na classic pair of diamond solitaire
89 },
90 {
91     "Title": "14-kt. Cultured Pearl Diamond Earrings",
92     "ListPrice": "$508.99",
93     "Price": "$179.99",
94     "Saving": "$329.00",
95     "SavingPercent": "(64%)",
96     "Content": "Create an elegant appearance with\r\nthese pearl and diamond stud earr
97 },

```

```

98     {
99         "Title": "14-kt. Diamond 7.5-8 mm Pearl Pendant",
100        "ListPrice": "$196.99",
101        "Price": "$69.99",
102        "Saving": "$127.00",
103        "SavingPercent": "(64%)",
104        "Content": "Add a classic to your jewelry\r\ncollection with this 14-karat gold, d
105    },
106    {
107        "Title": "14-kt. Diamond Solitaire Earrings (.50 TW)",
108        "ListPrice": "$1,369.99",
109        "Price": "$409.99",
110        "Saving": "$960.00",
111        "SavingPercent": "(70%)",
112        "Content": "This earring set has two brilliant-cut\r\ndiamonds (0.50 total carat w
113    },
114    {
115        "Title": "14-kt White Gold Diamond Band (0.50 TW)",
116        "ListPrice": "$1,635.00",
117        "Price": "$609.99",
118        "Saving": "$1,025.01",
119        "SavingPercent": "(62%)",
120        "Content": "Crafted of 14-karat white gold,\r\nthis stylish ring features a bright
121    }
122 ]

```

5.2 Izhod Regularnega izraza pri Overstock.com 2.

```

1    [
2        {
3            "Title": "14-kt. Green Jade Hoops",
4            "ListPrice": "$90.00",
5            "Price": "$46.99",
6            "Saving": "$43.01",
7            "SavingPercent": "(47%)",
8            "Content": "Hoops of cool green jade rest\r\nbetween 14-karat yellow gold endpiece
9        },
10       {
11           "Title": "14-kt. Jade Doughnut Pendant",
12           "ListPrice": "$150.00",
13           "Price": "$48.99",
14           "Saving": "$101.01",
15           "SavingPercent": "(67%)",
16           "Content": "The 25-mm disk hangs delicately\r\nfrom a 14-karat gold chain. The dis

```

```

17     },
18     {
19         "Title": "14-kt. Charcoal Jade and Ruby Elephant Pendant",
20         "ListPrice": "$100.00",
21         "Price": "$28.99",
22         "Saving": "$71.01",
23         "SavingPercent": "(71%)",
24         "Content": "Carved of rich dark grey jade,\r\nthis elephant pendant has 14-karat y
25     },
26     {
27         "Title": "14-kt. Carved Lavender Jade Earrings",
28         "ListPrice": "$80.00",
29         "Price": "$39.99",
30         "Saving": "$40.01",
31         "SavingPercent": "(50%)",
32         "Content": "Luscious 8-mm lavender jade balls, carved with intricate Asian style,
33     },
34     {
35         "Title": "14-kt. Jade Cross Pendant",
36         "ListPrice": "$150.00",
37         "Price": "$49.99",
38         "Saving": "$100.01",
39         "SavingPercent": "(66%)",
40         "Content": "Green jade and gold create this\r\nbeautiful cross pendant. Cylindrica
41     },
42     {
43         "Title": "14-kt. Multicolored Jade Earrings",
44         "ListPrice": "$375.00",
45         "Price": "$99.99",
46         "Saving": "$275.01",
47         "SavingPercent": "(73%)",
48         "Content": "A delicate wrapping of 14-karat\r\nyellow gold wire holds six 6 x 4 pe
49     },
50     {
51         "Title": "14-kt. Multicolored Jade Ring",
52         "ListPrice": "$250.00",
53         "Price": "$56.99",
54         "Saving": "$193.01",
55         "SavingPercent": "(77%)",
56         "Content": "A delicate wrapping of 14-karat\r\nyellow gold wire holds six 6 x 4 ov
57     },
58     {
59         "Title": "14-kt. Onyx and Ruby Elephant Pendant",
60         "ListPrice": "$100.00",
61         "Price": "$35.99",
62         "Saving": "$64.01",

```



```

63     "SavingPercent": "(64%)",
64     "Content": "Carved of rich black onyx, this\r\n\nelephant pendant has 14-karat yellow
65 }
66 ]

```

5.3 Izhod xPath pri Overstock.com 1.

```

1  [
2    {
3      "Title": "10-kt. Seven Diamond Ladies Heart Ring (0.08 TW)",
4      "ListPrice": "$149.00",
5      "Price": "$69.99",
6      "Content": "This ladies fashion ring dazzles\r\n\nwith hearts and diamonds. The gold
7      "Saving": "$79.01",
8      "SavingPercent": "(53%)"
9    },
10   {
11     "Title": "10-Kt. Diamond Ring (.25 TW)",
12     "ListPrice": "$250.00",
13     "Price": "$74.90",
14     "Content": "Nineteen round diamonds accent this 10-karat yellow gold ring with fil
15     "Saving": "$175.10",
16     "SavingPercent": "(70%)"
17   },
18   {
19     "Title": "10-kt. Pearl and Diamond Butterfly Earrings",
20     "ListPrice": "$149.00",
21     "Price": "$42.99",
22     "Content": "Perfectly proportioned 5.5- to\r\n\n6-mm cultured pearls on 10-karat yel
23     "Saving": "$106.01",
24     "SavingPercent": "(71%)"
25   },
26   {
27     "Title": "14-kt. Diamond 'S' Tennis Bracelet (2.00 TW)",
28     "ListPrice": "$1,539.99",
29     "Price": "$499.99",
30     "Content": "Invest in a swirl of light with\r\n\nthis diamond 'S' tennis bracelet. C
31     "Saving": "$1,040.00",
32     "SavingPercent": "(67%)"
33   },
34   {
35     "Title": "10-kt. Diamond Band Fashion Ring (.11 TW)",
36     "ListPrice": "$179.99",
37     "Price": "$79.99",

```

```

38     "Content": "Crafted in white and yellow gold,\r\nthis ring displays a band of seven
39     "Saving": "$100.00",
40     "SavingPercent": "(55%)"
41 },
42 {
43     "Title": "14-kt. White Gold, Pearl and Diamond Ring",
44     "ListPrice": "$419.99",
45     "Price": "$149.99",
46     "Content": "Show your romantic side with this\r\n14-karat diamond and pearl ring.
47     "Saving": "$270.00",
48     "SavingPercent": "(64%)"
49 },
50 {
51     "Title": "14-kt. Gold Diamond Present Future Pendant (.25TW)",
52     "ListPrice": "$299.00",
53     "Price": "$149.99",
54     "Content": "Designed with three large, sparkling\r\nndiamonds to represent past, pr
55     "Saving": "$149.01",
56     "SavingPercent": "(49%)"
57 },
58 {
59     "Title": "14-kt. Diamond Solitaire Pendant (.33 TW)",
60     "ListPrice": "$1,019.99",
61     "Price": "$319.99",
62     "Content": "In this simple, yet elegant pendant,\r\na round brilliant diamond (0.3
63     "Saving": "$700.00",
64     "SavingPercent": "(68%)"
65 },
66 {
67     "Title": "14-kt. Diamond Solitaire Earrings (0.33 TW)",
68     "ListPrice": "$639.99",
69     "Price": "$199.99",
70     "Content": "Dazzle your way into her heart,\r\nwith these classic diamond solitair
71     "Saving": "$440.00",
72     "SavingPercent": "(68%)"
73 },
74 {
75     "Title": "14-kt. Diamond Cross Pendant (.06 TW)",
76     "ListPrice": "$305.00",
77     "Price": "$119.99",
78     "Content": "Over a cleanly sculpted Roman\r\ncross of 14-karat white gold drapes a
79     "Saving": "$185.01",
80     "SavingPercent": "(60%)"
81 },
82 {
83     "Title": "14-kt. Diamond Solitaire Stud Earrings (.50 TW)",

```

```

84     "ListPrice": "$999.99",
85     "Price": "$359.99",
86     "Content": "Every jewelry collection needs\r\na classic pair of diamond solitaire
87     "Saving": "$640.00",
88     "SavingPercent": "(64%)"
89 },
90 {
91     "Title": "14-kt. Cultured Pearl Diamond Earrings",
92     "ListPrice": "$508.99",
93     "Price": "$179.99",
94     "Content": "Create an elegant appearance with\r\nthese pearl and diamond stud earr
95     "Saving": "$329.00",
96     "SavingPercent": "(64%)"
97 },
98 {
99     "Title": "14-kt. Diamond 7.5-8 mm Pearl Pendant",
100    "ListPrice": "$196.99",
101    "Price": "$69.99",
102    "Content": "Add a classic to your jewelry\r\ncollection with this 14-karat gold, d
103    "Saving": "$127.00",
104    "SavingPercent": "(64%)"
105 },
106 {
107     "Title": "14-kt. Diamond Solitaire Earrings (.50 TW)",
108     "ListPrice": "$1,369.99",
109     "Price": "$409.99",
110     "Content": "This earring set has two brilliant-cut\r\ndiamonds (0.50 total carat w
111     "Saving": "$960.00",
112     "SavingPercent": "(70%)"
113 },
114 {
115     "Title": "14-kt White Gold Diamond Band (0.50 TW)",
116     "ListPrice": "$1,635.00",
117     "Price": "$609.99",
118     "Content": "Crafted of 14-karat white gold,\r\nthis stylish ring features a bright
119     "Saving": "$1,025.01",
120     "SavingPercent": "(62%)"
121 }
122 ]

```

5.4 Izhod xPath pri Overstock.com 2.

```

1  [
2  {

```

```

3      "Title": "14-kt. Green Jade Hoops",
4      "ListPrice": "$90.00",
5      "Price": "$46.99",
6      "Content": "Hoops of cool green jade rest\r\nbetween 14-karat yellow gold endpieces",
7      "Saving": "$43.01",
8      "SavingPercent": "(47%)"
9  },
10 {
11     "Title": "14-kt. Jade Doughnut Pendant",
12     "ListPrice": "$150.00",
13     "Price": "$48.99",
14     "Content": "The 25-mm disk hangs delicately\r\nfrom a 14-karat gold chain. The dis",
15     "Saving": "$101.01",
16     "SavingPercent": "(67%)"
17 },
18 {
19     "Title": "14-kt. Charcoal Jade and Ruby Elephant Pendant",
20     "ListPrice": "$100.00",
21     "Price": "$28.99",
22     "Content": "Carved of rich dark grey jade,\r\nthis elephant pendant has 14-karat y",
23     "Saving": "$71.01",
24     "SavingPercent": "(71%)"
25 },
26 {
27     "Title": "14-kt. Carved Lavender Jade Earrings",
28     "ListPrice": "$80.00",
29     "Price": "$39.99",
30     "Content": "Luscious 8-mm lavender jade balls, carved with intricate Asian style,",
31     "Saving": "$40.01",
32     "SavingPercent": "(50%)"
33 },
34 {
35     "Title": "14-kt. Jade Cross Pendant",
36     "ListPrice": "$150.00",
37     "Price": "$49.99",
38     "Content": "Green jade and gold create this\r\nbeautiful cross pendant. Cylindrica",
39     "Saving": "$100.01",
40     "SavingPercent": "(66%)"
41 },
42 {
43     "Title": "14-kt. Multicolored Jade Earrings",
44     "ListPrice": "$375.00",
45     "Price": "$99.99",
46     "Content": "A delicate wrapping of 14-karat\r\nyellow gold wire holds six 6 x 4 pe",
47     "Saving": "$275.01",
48     "SavingPercent": "(73%)"

```

```

49     },
50     {
51         "Title": "14-kt. Multicolored Jade Ring",
52         "ListPrice": "$250.00",
53         "Price": "$56.99",
54         "Content": "A delicate wrapping of 14-karat\r\nyellow gold wire holds six 6 x 4 ov
55         "Saving": "$193.01",
56         "SavingPercent": "(77%)"
57     },
58     {
59         "Title": "14-kt. Onyx and Ruby Elephant Pendant",
60         "ListPrice": "$100.00",
61         "Price": "$35.99",
62         "Content": "Carved of rich black onyx, this\r\nelephant pendant has 14-karat yello
63         "Saving": "$64.01",
64         "SavingPercent": "(64%)"
65     }
66 ]

```

5.5 Izhod xPath pri Rtvvslo.si 1.

```

1  {
2      "Title": "Audi A6 50 TDI quattro: nemir v premijskem razredu",
3      "Subtitle": "Test nove generacije",
4      "Lead": "To je novi audi A6. V razred najdražjih in najbolj premijskih zrebcev je vn
5      "Author": "Miha Merljak",
6      "PublishedTime": "| 28. december 2018 ob 08:51\t\t\t\r\n\t",
7      "Content": "Samo poglejte njegovo masko - to ogromno satovje z radarji na takem polc
8  }

```

5.6 Izhod xPath pri Rtvvslo.si 2.

```

1  {
2      "Title": "Volvo XC 40 D4 AWD momentum: suvereno med najboljše v razredu",
3      "Subtitle": "Test novega modela",
4      "Lead": "XC 40 je najmanjši Volvov SUV, ki se oblikovno skoraj v celotni naslanja na
5      "Author": "Miha Merljak",
6      "PublishedTime": "| 25. januar 2019 ob 15:23\t\t\t\r\n\t",
7      "Content": "Volvo XC 40 D4 AWD momentum_ suvereno med najboljše v razredu - RTVSL0.
8  }

```

5.7 Izhod Regularnega izraza pri Rtv slo.si 1.

```
1 {
2   "Title": "Audi A6 50 TDI quattro: nemir v premijskem razredu",
3   "Subtitle": "Test nove generacije",
4   "Lead": "To je novi audi A6. V razred najdražjih in najbolj premijskih zrebcev je vn
5   "Author": "Miha Merljak",
6   "PublishedTime": "| 28. december 2018 ob 08:51\t\t\t\t\r\n\t",
7   "Content": "Samo poglejte njegovo masko - to ogromno satovje z radarji na takem polo
8 }
```

5.8 Izhod Regularnega izraza pri Rtv slo.si 2.

```
1 {
2   "Title": "Volvo XC 40 D4 AWD momentum: suvereno med najboljše v razredu",
3   "Subtitle": "Test novega modela",
4   "Lead": "XC 40 je najmanjši Volvov SUV, ki se oblikovno skoraj v celotni naslanja na
5   "Author": "Miha Merljak",
6   "PublishedTime": "| 25. januar 2019 ob 15:23\t\t\t\t\r\n\t",
7   "Content": "Volvo XC 40 D4 AWD momentum_ suvereno med najboljše v razredu - RTVSLO.
8 }
```

5.9 Izhod Regularnega izraza in xPath pri twitter.com 1.

```
1 [
2   {
3     "Handle": "Dadsaysjokes",
4     "Content": "If you want a good laugh order our book now. <img class=\"Emoji Emoji-
5     "Timestmap": "14 Nov 2018",
6     "Replies": "10",
7     "Retweets": "28",
8     "Favourites": "275"
9   },
10  {
11    "Handle": "Dadsaysjokes",
12    "Content": "Why is 16 always full? \r\n\r\nBecause it 8 and 8.",
13    "Timestmap": "2h",
14    "Replies": "4",
15    "Retweets": "22",
16    "Favourites": "158"
17  },
18  {
```

```

19     "Handle": "Dadsaysjokes",
20     "Content": "I didn't think vodka could help my problems. \r\n\r\nBut it was worth",
21     "Timestmap": "2h",
22     "Replies": "",
23     "Retweets": "25",
24     "Favourites": "114"
25 },
26 {
27     "Handle": "Dadsaysjokes",
28     "Content": "Sometimes I wake up grumpy. \r\n\r\nBut other times I let her sleep in",
29     "Timestmap": "17h",
30     "Replies": "5",
31     "Retweets": "79",
32     "Favourites": "409"
33 },
34 {
35     "Handle": "Dadsaysjokes",
36     "Content": "My son just accused me of lying. \r\n\r\nI wouldn't mind but I don't e",
37     "Timestmap": "17h",
38     "Replies": "",
39     "Retweets": "16",
40     "Favourites": "125"
41 },
42 {
43     "Handle": "Dadsaysjokes",
44     "Content": "My so asked: \"Do you want to hear a really good Batman impression?\" \r\n\r\nI said:",
45     "Timestmap": "17h",
46     "Replies": "4",
47     "Retweets": "31",
48     "Favourites": "305"
49 },
50 {
51     "Handle": "Dadsaysjokes",
52     "Content": "My wife asked me: \"What starts with F and ends in K. \r\n\r\nI said:",
53     "Timestmap": "17h",
54     "Replies": "2",
55     "Retweets": "43",
56     "Favourites": "375"
57 },
58 {
59     "Handle": "Dadsaysjokes",
60     "Content": "People always told my dad his pride would be the death of him. \r\n\r\n",
61     "Timestmap": "Apr 28",
62     "Replies": "3",
63     "Retweets": "23",

```

```

64         "Favourites": "202"
65     },
66     {
67         "Handle": "Dadsaysjokes",
68         "Content": "Job interviewer: "At the start you'll be earning £20,000; later that w
69         "Timestmap": "Apr 28",
70         "Replies": "",
71         "Retweets": "30",
72         "Favourites": "202"
73     },
74     {
75         "Handle": "Dadsaysjokes",
76         "Content": "What do you call 52 pieces of bread? \r\n\r\nA deck of carbs.",
77         "Timestmap": "Apr 28",
78         "Replies": "4",
79         "Retweets": "43",
80         "Favourites": "244"
81     },
82     {
83         "Handle": "Dadsaysjokes",
84         "Content": "What do you say when you tickle a tiny millionaire? \r\n\r\nGucci Gucci
85         "Timestmap": "Apr 28",
86         "Replies": "1",
87         "Retweets": "45",
88         "Favourites": "255"
89     },
90     {
91         "Handle": "Dadsaysjokes",
92         "Content": "What do you call a man with a flatfish on his head? \r\n\r\nRay.",
93         "Timestmap": "Apr 28",
94         "Replies": "2",
95         "Retweets": "7",
96         "Favourites": "83"
97     },
98     {
99         "Handle": "Dadsaysjokes",
100        "Content": "When I become a superhero I'm going to be called "ironic". \r\n\r\nSo
101        "Timestmap": "Apr 27",
102        "Replies": "4",
103        "Retweets": "75",
104        "Favourites": "373"
105    },
106    {
107        "Handle": "Dadsaysjokes",
108        "Content": "Why did the monkey and Jane fight over Tarzan? \r\n\r\nBecause they he

```



```

109     "Timestmap": "Apr 27",
110     "Replies": "1",
111     "Retweets": "15",
112     "Favourites": "112"
113 },
114 {
115     "Handle": "Dadsaysjokes",
116     "Content": "My wife told me she thought we'd have less arguments if I wasn't so pe
117     "Timestmap": "Apr 27",
118     "Replies": "3",
119     "Retweets": "41",
120     "Favourites": "300"
121 },
122 {
123     "Handle": "Dadsaysjokes",
124     "Content": "Spiders are the only web developers who love finding bugs.",
125     "Timestmap": "Apr 27",
126     "Replies": "1",
127     "Retweets": "164",
128     "Favourites": "591"
129 },
130 {
131     "Handle": "Dadsaysjokes",
132     "Content": "My weird boss has assigned designated toilet breaks for all employees
133     "Timestmap": "Apr 27",
134     "Replies": "1",
135     "Retweets": "16",
136     "Favourites": "148"
137 },
138 {
139     "Handle": "Dadsaysjokes",
140     "Content": "The Gluteus Maximus is the largest muscle in your body, \r\n\r\nIn fac
141     "Timestmap": "Apr 27",
142     "Replies": "3",
143     "Retweets": "45",
144     "Favourites": "271"
145 },
146 {
147     "Handle": "Dadsaysjokes",
148     "Content": "What did the German sausage say to his friend? \r\n\r\nYou are the wur
149     "Timestmap": "Apr 26",
150     "Replies": "2",
151     "Retweets": "30",
152     "Favourites": "242"
153 },
154 {

```

```

155     "Handle": "Dadsaysjokes",
156     "Content": "What is made out of brass and sounds like Tom Jones? \r\n\r\nTrombones",
157     "Timestmap": "Apr 26",
158     "Replies": "3",
159     "Retweets": "19",
160     "Favourites": "143"
161 },
162 {
163     "Handle": "Dadsaysjokes",
164     "Content": "Officer: "I'm arresting you for downloading the entire Wikipedia." \r\n",
165     "Timestmap": "Apr 26",
166     "Replies": "2",
167     "Retweets": "92",
168     "Favourites": "577"
169 }
170 ]

```

5.10 Izhod Regularnega izraza in xPath pri twitter.com

2.

```

1  [
2  {
3      "Handle": "kratkazgodba",
4      "Content": "Ko sem s tabo, si z užitkom solim kavo.",
5      "Timestmap": "Apr 27",
6      "Replies": "",
7      "Retweets": "3",
8      "Favourites": "7"
9  },
10 {
11     "Handle": "kratkazgodba",
12     "Content": "Oblacila se je v crno, a bila je mavrica v zivljenju mnogih ljudi.",
13     "Timestmap": "Apr 27",
14     "Replies": "1",
15     "Retweets": "5",
16     "Favourites": "23"
17 },
18 {
19     "Handle": "kratkazgodba",
20     "Content": "Bila je zmedljivo lepa. \r\nTako lepa je bila, da je zmedla vse okoli",
21     "Timestmap": "Apr 25",
22     "Replies": "2",
23     "Retweets": "4",
24     "Favourites": "19"

```

```

25 },
26 {
27   "Handle": "kratkazgodba",
28   "Content": "Gospod sodnik, nisem bil jaz. Melanholija je bila strel sprozila.",
29   "Timestmap": "Apr 24",
30   "Replies": "",
31   "Retweets": "1",
32   "Favourites": "6"
33 },
34 {
35   "Handle": "kratkazgodba",
36   "Content": "Vsak konec je prinesel možnost novega zacetka.",
37   "Timestmap": "Apr 14",
38   "Replies": "",
39   "Retweets": "1",
40   "Favourites": "11"
41 },
42 {
43   "Handle": "kratkazgodba",
44   "Content": "Nad Piranom se je pojavil NLP. \r\nNekaj minut je utripal v ponavljajo",
45   "Timestmap": "Apr 11",
46   "Replies": "",
47   "Retweets": "1",
48   "Favourites": "8"
49 },
50 {
51   "Handle": "kratkazgodba",
52   "Content": "Doziveto je razlagala zakaj je morala kupiti nove cevlje: \"Bila je ra",
53   "Timestmap": "Apr 10",
54   "Replies": "1",
55   "Retweets": "",
56   "Favourites": "21"
57 },
58 {
59   "Handle": "kratkazgodba",
60   "Content": "Vcasih se je izgubila v njegovi temi, ampak ni ga bila pripravljena za",
61   "Timestmap": "Apr 8",
62   "Replies": "1",
63   "Retweets": "2",
64   "Favourites": "14"
65 },
66 {
67   "Handle": "kratkazgodba",
68   "Content": "\"Ob tebi imam celo smisel za humor boljsi!\" \r\n\"Ker ti stose prevaj",
69   "Timestmap": "Apr 7",
70   "Replies": "",

```

```

71     "Retweets": "3",
72     "Favourites": "12"
73 },
74 {
75     "Handle": "kratkazgodba",
76     "Content": "Ne papir, voscenke najbolj privlaci na sveze prebarvana stena.\r\nStena",
77     "Timestmap": "Apr 6",
78     "Replies": "",
79     "Retweets": "1",
80     "Favourites": "6"
81 },
82 {
83     "Handle": "kratkazgodba",
84     "Content": "Romanticno vecerjo je pripravila. \r\nLacna je bila.\r\nDve uri je zamujala",
85     "Timestmap": "Apr 3",
86     "Replies": "1",
87     "Retweets": "3",
88     "Favourites": "28"
89 },
90 {
91     "Handle": "kratkazgodba",
92     "Content": "Razocarana hci je odrasla. \r\nZ barvikami se je nehala igrati.\r\nSedela",
93     "Timestmap": "Mar 29",
94     "Replies": "1",
95     "Retweets": "5",
96     "Favourites": "24"
97 },
98 {
99     "Handle": "kratkazgodba",
100    "Content": "Povecala je obiske fitnesa, zacela tecaj plesa in kulinarike . To pole",
101    "Timestmap": "Mar 28",
102    "Replies": "",
103    "Retweets": "",
104    "Favourites": "16"
105 },
106 {
107    "Handle": "kratkazgodba",
108    "Content": "Sovrazita se do dna duse, njuni telesi se pa se vedno pogresata.",
109    "Timestmap": "Mar 27",
110    "Replies": "",
111    "Retweets": "4",
112    "Favourites": "14"
113 },
114 {
115    "Handle": "kratkazgodba",
116    "Content": "Neskoncno trpljenje ob pricakovanju uzitka, ki ga lahko ustvari le meh

```

```

117     "Timestmap": "Mar 26",
118     "Replies": "",
119     "Retweets": "3",
120     "Favourites": "26"
121 },
122 {
123     "Handle": "kratkazgodba",
124     "Content": "Koliko stane resnica?\r\nKako zgleda?\r\nObraz ima?\r\nResnica?\r\n\r\n",
125     "Timestmap": "Mar 21",
126     "Replies": "",
127     "Retweets": "3",
128     "Favourites": "17"
129 },
130 {
131     "Handle": "kratkazgodba",
132     "Content": "\"Nisem vec froc, vem da magija ne obstaja, je samo druzbeni konstrukt\"",
133     "Timestmap": "Mar 13",
134     "Replies": "",
135     "Retweets": "4",
136     "Favourites": "10"
137 },
138 {
139     "Handle": "kratkazgodba",
140     "Content": "2030 - Proticepilci so zmagali. Cepljenje otrok ni brezplačna pravica.",
141     "Timestmap": "Mar 10",
142     "Replies": "1",
143     "Retweets": "2",
144     "Favourites": "12"
145 },
146 {
147     "Handle": "kratkazgodba",
148     "Content": "Umetno inteligenco smo imeli za nepravo, manjvredno. \r\nZatem smo nal...",
149     "Timestmap": "Mar 9",
150     "Replies": "2",
151     "Retweets": "1",
152     "Favourites": "13"
153 },
154 {
155     "Handle": "kratkazgodba",
156     "Content": "Politik je bil porocen z resnico. To mu je nemalokrat povzrocilo poslo...",
157     "Timestmap": "Mar 7",
158     "Replies": "",
159     "Retweets": "4",
160     "Favourites": "11"
161 },
162 {

```

```

163     "Handle": "kratkazgodba",
164     "Content": "Najboljsi izdelovalec beneskih mask ponosno nosi iznakazen obraz.",
165     "Timestmap": "Mar 5",
166     "Replies": "",
167     "Retweets": "2",
168     "Favourites": "4"
169 },
170 {
171     "Handle": "kratkazgodba",
172     "Content": "\"Hm . Rada potujem,\" je na vprasanje: \"Kaj najraje pocnes?\" odgovorila.",
173     "Timestmap": "Feb 27",
174     "Replies": "",
175     "Retweets": "",
176     "Favourites": "2"
177 },
178 {
179     "Handle": "kratkazgodba",
180     "Content": "Iskal je besede, ki bi bralcu ustvarile največji vtis. \r\nNazadnje je našel.",
181     "Timestmap": "Feb 26",
182     "Replies": "",
183     "Retweets": "5",
184     "Favourites": "17"
185 },
186 {
187     "Handle": "kratkazgodba",
188     "Content": "\"Kajenje skrajša življenje,\" je vedela že od prvega cigareta. To dejstvo jo je spodbudilo.",
189     "Timestmap": "Feb 23",
190     "Replies": "",
191     "Retweets": "",
192     "Favourites": "7"
193 },
194 {
195     "Handle": "kratkazgodba",
196     "Content": "Najbolj je uživala, ko jo je slacil z besedami.",
197     "Timestmap": "Feb 21",
198     "Replies": "",
199     "Retweets": "5",
200     "Favourites": "16"
201 },
202 {
203     "Handle": "kratkazgodba",
204     "Content": "Vsak večer je brala. \r\nKo so se ji oči zaprle, se je zgodba prelila.",
205     "Timestmap": "Feb 20",
206     "Replies": "",
207     "Retweets": "",
208     "Favourites": "21"

```

```

209 },
210 {
211   "Handle": "kratkazgodba",
212   "Content": "Te dni so se prerekali o glasbenem tekmovanju, prejsnje o politicnem s
213   "Timestmap": "Feb 19",
214   "Replies": "",
215   "Retweets": "10",
216   "Favourites": "20"
217 },
218 {
219   "Handle": "kratkazgodba",
220   "Content": "Govoril je mnozico jezikov in dialektov, največ pa je povedal, ko je o
221   "Timestmap": "Feb 12",
222   "Replies": "1",
223   "Retweets": "7",
224   "Favourites": "33"
225 },
226 {
227   "Handle": "kratkazgodba",
228   "Content": "Nesrečno je strmel v okno. \r\n\"Spet dez! Sovrazim dez pozimi! Kje je
229   "Timestmap": "Feb 10",
230   "Replies": "1",
231   "Retweets": "4",
232   "Favourites": "18"
233 },
234 {
235   "Handle": "kratkazgodba",
236   "Content": "Nosila je le originale dragih blagovnih znamk. \r\nIzjemo je naredila
237   "Timestmap": "Feb 5",
238   "Replies": "",
239   "Retweets": "6",
240   "Favourites": "44"
241 },
242 {
243   "Handle": "kratkazgodba",
244   "Content": "Bila je dobra. \r\nPredobra. \r\nBila je tako dobra, da se je vzburla
245   "Timestmap": "Jan 30",
246   "Replies": "",
247   "Retweets": "1",
248   "Favourites": "9"
249 },
250 {
251   "Handle": "kratkazgodba",
252   "Content": "Disala je. \r\nPo parfumu in cigaretah. \r\nObozeval sem jo.\r\nPrevar
253   "Timestmap": "Jan 22",
254   "Replies": "1",

```

```

255         "Retweets": "4",
256         "Favourites": "27"
257     },
258     {
259         "Handle": "kratkazgodba",
260         "Content": "\"Vsaka odlocitev sprozi nov svet. Drugi ne odmre, nadaljuje svojo pot",
261         "Timestmap": "Jan 19",
262         "Replies": "",
263         "Retweets": "2",
264         "Favourites": "3"
265     },
266     {
267         "Handle": "kratkazgodba",
268         "Content": "\"Zakaj vleces prevleko s kavca?! Se bo unibil! Kaj ti je?!\"\\r\\n\\\"Dra",
269         "Timestmap": "Jan 15",
270         "Replies": "",
271         "Retweets": "1",
272         "Favourites": "3"
273     },
274     {
275         "Handle": "kratkazgodba",
276         "Content": "Med ciscenjem kavca je nekaj njenih las nasel. \\r\\nNasmeh se mu je na",
277         "Timestmap": "Jan 14",
278         "Replies": "",
279         "Retweets": "1",
280         "Favourites": "18"
281     },
282     {
283         "Handle": "kratkazgodba",
284         "Content": "\"Ko jo zagledam v kuhinji ustvarjati to ... To njeno kulinaricno magi",
285         "Timestmap": "Jan 13",
286         "Replies": "",
287         "Retweets": "1",
288         "Favourites": "13"
289     },
290     {
291         "Handle": "kratkazgodba",
292         "Content": "\"Vcasih sem jezna nanj. Posodo pusti na pultu ali ob umivanju le-te v",
293         "Timestmap": "Jan 13",
294         "Replies": "",
295         "Retweets": "1",
296         "Favourites": "8"
297     },
298     {
299         "Handle": "kratkazgodba",
300         "Content": "\"Kako bi v desetih besedah opisal pravo ljubezen?\"\\r\\n\\\"Le eno potre

```



```

301         "Timestmap": "Jan 12",
302         "Replies": "",
303         "Retweets": "4",
304         "Favourites": "17"
305     },
306     {
307         "Handle": "kratkazgodba",
308         "Content": "\"VSEC ti je.\"\\r\\n\\\"Nop, sam frenda sva.\"\\r\\n\\\"Mogoce celo VEC kot v
309         \"Timestmap\": \"Jan 11\",
310         \"Replies\": \"\",
311         \"Retweets\": \"5\",
312         \"Favourites\": \"34\"
313     },
314     {
315         \"Handle\": \"kratkazgodba\",
316         \"Content\": \"Sedel je. \\r\\nNervozno je preveril okolico. \\r\\nNic. \\r\\nSkeniral je s
317         \"Timestmap\": \"Jan 10\",
318         \"Replies\": \"\",
319         \"Retweets\": \"3\",
320         \"Favourites\": \"14\"
321     },
322     {
323         \"Handle\": \"kratkazgodba\",
324         \"Content\": \"Obupan je odprl predal.\\r\\nOb pistoli je zagledal rdec gumb, ob njemu
325         \"Timestmap\": \"Jan 9\",
326         \"Replies\": \"\",
327         \"Retweets\": \"2\",
328         \"Favourites\": \"23\"
329     },
330     {
331         \"Handle\": \"kratkazgodba\",
332         \"Content\": \"\"Ze dolgo se nisva potila v dvoje,\" se je nasmehnila in me pospremila
333         \"Timestmap\": \"Jan 8\",
334         \"Replies\": \"5\",
335         \"Retweets\": \"8\",
336         \"Favourites\": \"59\"
337     },
338     {
339         \"Handle\": \"kratkazgodba\",
340         \"Content\": \"Prijava sta se za roke, prepletla prste ... Ujemanje je bilo popolno.\\
341         \"Timestmap\": \"Jan 7\",
342         \"Replies\": \"1\",
343         \"Retweets\": \"3\",
344         \"Favourites\": \"10\"
345     },
346     {

```

```

347     "Handle": "kratkazgodba",
348     "Content": "Dezne kapljice mirno trkajo ob kristalno morsko gladino.\r\nGaleb mirno",
349     "Timestmap": "Jan 6",
350     "Replies": "",
351     "Retweets": "2",
352     "Favourites": "4"
353 },
354 {
355     "Handle": "kratkazgodba",
356     "Content": "\"Sovrazim te dolge kopleksne zalostne zgodbe z utrinki evforije. Kako",
357     "Timestmap": "Jan 5",
358     "Replies": "",
359     "Retweets": "6",
360     "Favourites": "27"
361 },
362 {
363     "Handle": "kratkazgodba",
364     "Content": "\"Kako ti uspevajo novoletne zaobljube?\"\r\n\"Super! V predalu imam 1",
365     "Timestmap": "Jan 4",
366     "Replies": "",
367     "Retweets": "4",
368     "Favourites": "19"
369 },
370 {
371     "Handle": "kratkazgodba",
372     "Content": "Kot darilo za 2019, kot je obicaj, vas bom v naslednjih dneh razveselj",
373     "Timestmap": "Jan 3",
374     "Replies": "",
375     "Retweets": "",
376     "Favourites": "11"
377 },
378 {
379     "Handle": "kratkazgodba",
380     "Content": "Orkan slabih odlocitev je odnesel vse kar ji je bilo v zivljenju domac",
381     "Timestmap": "Jan 1",
382     "Replies": "",
383     "Retweets": "6",
384     "Favourites": "15"
385 },
386 {
387     "Handle": "kratkazgodba",
388     "Content": "Opozorilo casovnim popotnikom - Ne voscite \"Srecno novo leto\" tistim",
389     "Timestmap": "31 Dec 2018",
390     "Replies": "",
391     "Retweets": "2",
392     "Favourites": "9"

```

```

393 },
394 {
395   "Handle": "kratkazgodba",
396   "Content": "Vcasih mi zelje spolzijo med besede. Vcasih med njimi znas prebrati, j
397   "Timestmap": "29 Dec 2018",
398   "Replies": "",
399   "Retweets": "2",
400   "Favourites": "15"
401 },
402 {
403   "Handle": "kratkazgodba",
404   "Content": "Z branjem je zbirala besede za bolj zgovorno prihodnost.",
405   "Timestmap": "28 Dec 2018",
406   "Replies": "",
407   "Retweets": "1",
408   "Favourites": "4"
409 },
410 {
411   "Handle": "kratkazgodba",
412   "Content": "Zbudil se je. Pocen gong mu je v glavi donel. \"Sindikalka!\" Spomnil
413   "Timestmap": "22 Dec 2018",
414   "Replies": "",
415   "Retweets": "2",
416   "Favourites": "10"
417 },
418 {
419   "Handle": "kratkazgodba",
420   "Content": "Ni ga motilo, da vecina ni opazila njegovih kril. Ni jih zato pridobil
421   "Timestmap": "20 Dec 2018",
422   "Replies": "",
423   "Retweets": "2",
424   "Favourites": "5"
425 },
426 {
427   "Handle": "kratkazgodba",
428   "Content": "Odprl se je le ob njenih dotikih.",
429   "Timestmap": "19 Dec 2018",
430   "Replies": "1",
431   "Retweets": "",
432   "Favourites": "6"
433 },
434 {
435   "Handle": "kratkazgodba",
436   "Content": "\"Vcasih si zelim, da si ga ne bi tako mokro zelela ...\"\\r\\n\\\"Vcasih
437   "Timestmap": "18 Dec 2018",
438   "Replies": "1",

```

```

439     "Retweets": "2",
440     "Favourites": "9"
441 },
442 {
443     "Handle": "kratkazgodba",
444     "Content": "Vecino casa je vložila v raziskovanje kotov in kombinacij pod katerimi
445     "Timestmap": "16 Dec 2018",
446     "Replies": "",
447     "Retweets": "2",
448     "Favourites": "9"
449 },
450 {
451     "Handle": "kratkazgodba",
452     "Content": "\"Zakaj vleces prevleko s kavca?! Se bo unicil! Kaj ti je?!\"\\r\\n\\\"Dra
453     "Timestmap": "7 Dec 2018",
454     "Replies": "",
455     "Retweets": "2",
456     "Favourites": "17"
457 },
458 {
459     "Handle": "kratkazgodba",
460     "Content": "Svojih sovražnikov je bil navajen. \\r\\nDobro jih je poznal. \\r\\nV zivl
461     "Timestmap": "2 Dec 2018",
462     "Replies": "",
463     "Retweets": "4",
464     "Favourites": "14"
465 },
466 {
467     "Handle": "kratkazgodba",
468     "Content": "Vonj njene kože je bila njegova najljubša droga.",
469     "Timestmap": "23 Nov 2018",
470     "Replies": "2",
471     "Retweets": "5",
472     "Favourites": "17"
473 },
474 {
475     "Handle": "kratkazgodba",
476     "Content": "Vedno me je premocno stisnil. Vcasih tako mocno, da so moje kosti malo
477     "Timestmap": "20 Nov 2018",
478     "Replies": "1",
479     "Retweets": "2",
480     "Favourites": "12"
481 }
482 ]

```

5.11 Izhod RoadRunner pri Overstock.com

Zaradi boljše vidljivosti bo prikazan samo del, kjer se nahajajo pomembni podatki. Celotni regularni izraz je na voljo v mapi outputs.

```
1 <table>
2   <tbody>
3     <tr>
4       <td align="right" nowrap="nowrap">
5         <b>
6           List Price:
7         </b>
8       </td>
9       <td align="left" nowrap="nowrap">
10        <s>
11          #text
12        </s>
13      </td>
14    </tr>
15    <tr>
16      <td align="right" nowrap="nowrap">
17        <b>
18          Price:
19        </b>
20      </td>
21      <td align="left" nowrap="nowrap">
22        <span class="bigred">
23          <b>
24            #text
25          </b>
26        </span>
27      </td>
28    </tr>
29    <tr>
30      <td align="right" nowrap="nowrap">
31        <b>
32          You Save:
33        </b>
34      </td>
35      <td align="left" nowrap="nowrap">
36        <span class="littleorange">
37          #text
38        </span>
39      </td>
40    </tr>
```

```

41     </tbody>
42 </table>

```

5.12 Izhod RoadRunner pri Rtv slo.si

Zaradi boljše vidljivosti bo prikazan samo del, kjer se nahajajo pomembni podatki. Celotni regularni izraz je na voljo v mapi outputs.

```

1  <div class="author">
2    <div class="author-name">
3      Miha Merljak
4    </div>
5  </div>
6  ...
7  <header class="article-header">
8    <div class="#text">
9      <h3 class="section-title animated-circles-onhover">
10       <svg version="1.1" class="heading-circles" xmlns="http://www.w3.org/2000/s
11         <g class="gray">
12           <circle cx="2.63" cy="10" r="1.6"/>
13           <circle cx="15.9" cy="10" r="3.49"/>
14           <circle cx="31.9" cy="10" r="5.44"/>
15           <circle cx="53.9" cy="10" r="7.63"/>
16         </g>
17         <g class="color-overlay">
18           <circle cx="2.63" cy="10" r="1.6"/>
19           <circle cx="15.9" cy="10" r="3.49"/>
20           <circle cx="31.9" cy="10" r="5.44"/>
21           <circle cx="53.9" cy="10" r="7.63"/>
22         </g>
23         <circle class="color" cx="77.51" cy="10" r="9.46"/>
24         <circle class="gray-overlay" cx="77.51" cy="10" r="9.46"/>
25       </svg>
26       <a href="#text">
27         #text
28       </a>
29     </h3>
30     <a class="comments-icon ml-auto" data-comments-number="#text" href="#text"/>
31   </div>
32   <h1>
33     #text
34   </h1>
35   (
36     <div class="article-meta-mobile">

```

```

37      (
38          #text
39      ) ?
40      (
41          <div class="#text" ( style="transition-duration: 0ms; transform: trans
42              (
43                  <figure class="photoswipe swiper-slide mobile-show-only swiper
44                      <a href="./Audi A6 50 TDI quattro_ nemir v premijskem razr
45                          
48                          <span class="icon-photo"/>
49                          Audi A6 je avtomobil za direktorje, ki se mu na avtoce
50                      </figcaption>
51                  </figure>
52              ) ?
53              (
54                  <strong>
55                      Miha Merljak
56                  </strong>
57              ) ?
58              (
59                  #text
60              ) ?
61          </div>
62      ) ?
63      (
64          <div class="#text">
65              (
66                  <span class="swiper-pagination-bullet swiper-pagination-bullet
67              ) ?
68              (
69                  Ljubljana - MMC RTV SLO
70              ) ?
71          </div>
72      ) ?
73  </div>
74  ) +
75  ...

```

5.13 Izhod RoadRunner pri Twitter.com

Zaradi boljše vidljivosti bo prikazan samo del, kjer se nahajajo pomembni podatki. Celotni regularni izraz je na voljo v mapi outputs.

```

1  <ol class="stream-items js-navigable-stream" id="stream-items-id">
2  (
3  ...
4  <div class="content">
5  (
6  <div class="#text">
7  (
8  <a class="account-group js-account-group js-action-profile js-user
9  
11 <strong class="fullname show-popup-with-id u-textTruncate
12 <#text
13 </strong>
14 <span>
15 ?
16 </span>
17 <span class="UserBadges"/>
18 <span class="UserNameBreak">
19 &nbsp;
20 </span>
21 </span>
22 <span class="username u-dir u-textTruncate" dir="ltr" data-ari
23 @
24 <b>
25 <#text
26 </b>
27 </span>
28 </a>
29 ) ?
30 ...
31 <p class="TweetTextSize TweetTextSize--normal js-tweet-text tweet-text" da
32 <#text
33 ...
34 ...
35 ) +
36 </ol>

```