

Web Information Extraction and Retrieval - Assignment 2 Report

Roman Komac, Dik Medvešček Murovec, Žan Ožbot

May 2019

1 Introduction

Extracting useful data from a collection as large and diverse as the whole surface web is a difficult task. Manual wrapper generation or wrapper induction may have been a feasible solution for earlier stages of the internet. With a wide public adoption however, the internet rapidly became a collection of semi-structured and inconsistent documents. Still to this day a wide array of websites face issues such as improper use of HTML5 elements, misaligned or missing closing element tags or inconsistent overall structure due to embedded or imported elements such as social network functionalities or adverts.

The task of the second programming assignment was to create wrappers that translate semi-structured HTML data into structured relational form. For the first part of the assignment we have used two different approaches, the first being data extraction using regular expressions and the second using XPath library. Since these types of wrappers require manual creation and a lot of labour to sustain they are mostly suitable for collections of resources that are consistent and change only on longer time intervals. The second part of the assignment was to implement an algorithm that automatically adapts to changes in document structure across several websites. Our approach for this part of the assignment is based on the original paper that describes the roadrunner algorithm. [1]

1.1 External libraries and tools

The first part of the implementation relies on lxml toolkit [2], which is a python binding of an xml parser. It was used for implementing the XPath extraction method since it contains a standard-compliant XPath implementation. No other libraries have been used.

1.2 Webpage selection

We have selected two web pages from the *mimovrste.com* web site. Content-wise they most closely resemble the web pages from *overstock.com* web site. From each page we extract the following data:

- Title - the name of the product
- Rating - rating in the interval [0,5]
- EnergyLevel - energy efficiency category
- Brand - the brand of the product
- CatalogNumber - internal reference number
- Stickers - highlighted words
- Description - product description
- Price - the final price including VAT
- BasePrice - base price before discount
- Saving - the difference between Price and BasePrice
- SavingPercent - discount given in percentage
- Availability - whether the product is in stock
- MarketplacePartner - delivery service

A graphic overview of the attributes given above is available in Figure 1.

The screenshot shows the mimovrste.com website interface. The header includes navigation links: Domača stran, Prezemni mesta, Seznam pooblaščenih servisov, and Kontakt. Below the header is a search bar and a login/register button. The main content area displays the product details for a Philips TV sprejemnik 43PFS5503/12. The product title is highlighted in a red box. Below the title, there are several attributes: Rating (2.000), EnergyLevel (A+), Brand (Philips), and CatalogNumber (1244513). The product image is shown below these attributes. To the right of the image, there are several sections: Stickers (Brezplačna dostava, Odlična cena, Priporočamo), Description (Philips TV sprejemnik 43PFS5503/12, LED televizija z ločljivostjo Full HD in diagonalo 108 cm (43"), Pixel Plus HD integriran sprejemnik DVB T/T2 (HEVC)/C/S2, snemanje oddaje, Timeshift, USB predvajalnik, 2x HDMI, 1x USB, optični avdio izhod, priključek za slušalke, Celoten opis), Price (259 €), BasePrice (Redna cena: 399 €), Saving (Vrnilo 50 €), SavingPercent (15 %), Availability (Na zalogi – predvidena odprema: 6.5.), and MarketplacePartner (Dostava Mimovrste). At the bottom of the product details, there is a red button labeled V KOŠARICO and a link to SLEDI CENI. Below the product image, there are several small thumbnail images showing different views of the TV.

Figure 1: Identification of data items and records from selected mimovrste.com web pages.

2 Methods

In Section 2.1 we further describe our implementation using regular expressions and XPath. The RoadRunner-like implementation and pseudocode are disclosed in Section 2.2.

2.1 Regular expressions and XPath

Since we are using regular expressions for text extraction in the first part we require some preprocessing of web pages. A small proportion of putative matches obtained using regex can be erroneous as they may be part of a script rather than the actual website. Web pages from *overstock.com* are therefore preprocessed so that all `<script></script>` elements are removed. Conversely some tags on *rtvslo.si* and *mimovrste.com* are postprocessed to exclude all unwanted HTML elements that are contained amidst text.

For XPath wrapper we firstly build a DOM tree. Using the lxml XPath implementation we then define the paths to the elements that contain useful data. Since we are not searching directly through the text representation but through a well-formed tree with proper closing tags we can define rules more tersely and do not need to worry about unwanted matches.

In the following Sections 2.1.1, 2.1.2 and 2.1.3 the structures that we search for with either regex or XPath are portrayed in HTML form. Text preceded with # denotes the extracted part. The structures are not direct copies of elements present in the document, they just describe key attributes and tag names that are used by us for extraction.

2.1.1 RTV SLO HTML elements

- Author

```
<div class="author-name">#Author</div>
```

- PublishedTime

```
<div class="publish-meta">#PublishedTime</div>
```

- Title

```
<header class="article-header"><h1>#Title</h1></header>
```

- SubTitle

```
<div class="subtitle">#SubTitle</div>
```

- Lead

```
<p class="lead">#Lead</p>
```

- Content

```
<article class="article">
  <figure></figure>
  #Content
  <div class="gallery"></div>
</article>
```

2.1.2 Overstock HTML elements

Overstock contains a collection of items. The following search criteria applies for multiple items and not for just a single match.

- Title

```
<td valign="top"><a><b>#Title</b></a></td>
```

- ListPrice

```
<td align="left" nowrap="nowrap"><s>#ListPrice</s></td>
```

- Price

```
<span class="bigred"><b>#Price</b></span>
```

- Saving, SavingPercent

```
<span class="littleorange">$#Saving (#SavingPercent)</span>
```

- Content

```
<span class="normal">#Content</span>
```

2.1.3 Mimovrste HTML elements

The selected web pages of *mimovrste.com* contain only information about one item. We therefore search only for the first match of each term. For the Description we postprocess extracted text by removing all non-text elements.

- Title

```
<section class="pro-column"><h1>#Title</h1></section>
```

- Rating

```
<div class="pro-subtitle-info">
  <span>
    <span class="con-reader">
      #Rating
    </span>
  </span>
</div>
```

- EnergyLevel

```
<span title="Energijska nalepka">#EnergyLevel</span>
```

- Brand

```

<span class="brand-info">
  <a class="link--pro-subtitle-info">
    #Brand
  </a>
</span>

```

- CatalogNumber


```

<span class="catalog-number">#CatalogNumber</span>

```
- Stickers (multiple)


```

<em>#Sticker</em>

```
- Description


```

<p class="pro-description">#Description</p>

```
- Price


```

<b class="pro-price">#Price</b>

```
- BasePrice, Save, SavingPercent


```

<span class="base-price">
  <span></span>
  <del>#BasePrice</del>
  ,#Save
  <span>(#SavingPercent)</span>
</span>

```
- LastPiece


```

<em class="label--last-piece">#LastPiece</em>

```
- Availability


```

<a class="con-text--availability">#Availability</a>

```
- MarketplacePartner


```

<span class="marketplace-partner__name">
  <b>#MarketplacePartner</b>
</span>

```

2.2 RoadRunner

RoadRunner is a technique for extracting data from HTML sites through the use of automatically generated wrappers. For our purposes we developed a RoadRunner-like algorithm, that takes two similar web pages and returns a generalized wrapper, capable of parsing both sites effectively. Our implementation is based on the article by Valter Crescenzi ??.

The pseudocode for the algorithm is as follows.

```
Read wrapper (initially just HTML of the first page)
Read sample
Do preprocessing on HTML
Read sample line by line
    If the line matches the same line in wrapper, then just ignore it
    If it matches only in tag, but not in content -> Replace content with #PCDATA
    If it matches neither in content nor in tag:
        Check if it is an iterator by comparing it to HTML code that came before
        If it is, remove all other occurrences of this iterator and mark it with ()+
        If it is not, mark it as optional ()?
When you are done with every line in the sample, end
```

Our only requirements are, that the HTML code is structured according to the XHTML standard and that it uses no special characters. We included no additional heuristics and or rules, that were not already described in the article.

3 Results

Resulting json files are part of the appendix.

References

- [1] Valter Crescenzi, Giansalvatore Mecca, Paolo Merialdo, et al. Roadrunner: Towards automatic data extraction from large web sites. In *VLDB*, volume 1, pages 109–118, 2001.
- [2] lxml - xml and html with python. <https://lxml.de/>, 2019. [Online; accessed 7-May-2019].

A Appendix

A.1 Regular expressions

A.1.1 RTV SLO - Audi A6

```
{
"Author": "Miha Merljak",
"PublishedTime": "28. december 2018 ob 08:51",
"Title": "Audi A6 50 TDI quattro: nemir v premijskem razredu",
"SubTitle": "Test nove generacije",
"Lead": "To je novi audi A6. V razred najdražjih in najbolj premijskih
→ žrebcev je vnesel nemir, še preden je sploh zapeljal na parkirni
→ prostor, rezerviran za izvršnega direktorja.",
"Content": "Samo poglejte njegovo masko { to ogromno satovje z radarji na
→ takem položaju, da se ti na avtocesti tudi pri 120 km/h vsi
→ spoštljivo umikajo, saj so prepričani, da gre za Pahorjev ali Šarčev
→ avto. Seveda, novi A6 lahko cesto in promet skenira s kar petimi
→ radarji, petimi kamerami, infrardečo kamero za nočni vid,
→ dvanajstimi ultrazvočnimi senzorji in laserskim čitalnikom {
→ lidarjem. V glavnem vojaška tehnologija v službi varnosti za fante,
→ ki smo radi gledali Top Gun, Bonda in druge možakarja s finimi
→ igračami.Novo poglavjeVozniški delovni prostor je novo poglavje
→ digitalne dobe, z dvema ogromnima zaslonoma, ki tako kot naprednejši
→ telefoni dregnejo blazinice vaših prstov, kot se sprehajate po
→ steklu. A še bolj se nam zdi pomembno, da so osnovna stikala tam,
→ kjer jih pričakujete. Najprej so torej zagotovili enostavno osnovo,
→ tisti bolj \"advanced\" vozniki pa si lahko nato vse skupaj še
→ veliko bolj prilagodijo. Velik korak naprej pri kabinskem udobju
→ zaznavajo tudi na zadnji klopi, tam je prostora v vseh smereh precej
→ več.Če vam pogled na Audijev spisek dodatne opreme ne odvzame volje
→ do življenja, potem vsekakor toplo priporočamo nakup zračnega
→ vzmetenja, saj dobi z njim A6 več različnih in vozniško zelo
→ uporabnih karakterjev. Enako velja za seksi luči z inteligentno
→ matrično osvetlitvijo, pa za športno podvozje in vsekakor za
→ štirikolesno krmiljenje. S tem postane A6 med ovinki v občutku na
→ volanu še veliko krajši in bolj agilen. Vse naštetu smo preskušali v
→ družbi agregata 50 TDI, ki je v resnici klasični trilitrski dizel,
→ podkrepjen z elektromotorjem. Ja, ta audi je mehki hibrid z
→ izjemnim navorom in dovolj moči kadar koli in kjer koli. Si pa
→ mislimo, da bo največji del trga zadovoljil že učinkovit dvolitrski
→ mehki hibrid z močjo 150 kilovatov. Ključni tehnični podatki:- na
→ testu Audi A6 50 TDI quattro tiptronic Mere:- dolžina: 4,9 m-
→ medosna razdalja: 2,9 m- obračalni krog: 12,1 m- prtljažnik:
```

```

530 l- masa: 1.900 kgPogon:- trilitrski šestvaljni dizelski motor-
→ moč: 210 kW- navor: 620 Nm- 8-stopenjski samodejni menjalnik- pogon
→ na vsa štiri 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/kmStroški pri 15.000 km in 5-letni uporabi:- nakupna
→ cena: 69.080 EUR- stroški finančnega lizinga: 4.463 EUR/5 let-
→ stroški registracije: 10.829 EUR/5 let- stroški vzdrževanja: 1.926
→ EUR/5 let- stroški goriva: 6.702 EUR/75.000 km- strošek 1 kompleta
→ pnevmatik: 716 EUR- vrednosti po 5 letih po Eurotaxu: 33.964 EUR-
→ stroški skupaj: 1.001 EUR/mesec"
}

```

A.1.2 RTV SLO - Volvo XC 40

```

{
"Author": "Miha Merljak",
"PublishedTime": "25. januar 2019 ob 15:23",
"Title": "Volvo XC 40 D4 AWD momentum: suvereno med najboljše v razredu",
"SubTitle": "Test novega modela",
"Lead": "XC 40 je najmanjši Volvov SUV, ki se oblikovno skoraj v celotni
→ naslanja na oba večja predhodnika. Že samo s tem so mu vrata do
→ denarnic tistih kupcev, ki iščejo izstopajočo, a hkrati visoko
→ kultivirano in prečiščeno dizajnersko govorico, na pol odprta.",
"Content": "Volvo se je nižjih srednjih razredov v preteklosti izogibal
→ ali pa je vanje vstopal z zelo nišnimi produkti, ki niso pustili
→ večjega tržnega pečata. V primeru XC 40 ni težko napovedati, da bo
→ ta tradicija prekinjena. Ponuja namreč visoko kakovost končne
→ izdelave in v kabini odlično premišljeno funkcionalnost ter na dotik
→ prijetne materiale. Še posebej hvalimo število, iznajdljivost in
→ velikost različnih odlagalnih prostorov ter široke, čvrste in zelo
→ udobne sedeže. Intuitivno in enostavno logično je upravljanje z
→ velikim vmesnikom, ki z večfunkcijskim zaslonom na dotik kraljuje na
→ z roko lahko dostopnem mestu na sredinski armaturi. Razočaranj ne bo
→ niti v velikosti in uporabnosti prtljažnega prostora, ki s 460 litri
→ prostornine sicer ni med večjimi v razredu, a se v uporabniškem
→ smislu odkupi z dobro urejenostjo ter domiselnimi rešitvami
→ pregrajevanja. Foto: David ŠavliXC 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 nevtrarno in
→ predvidljivo lego. V premeru preskušane modela, ki je imel v
→ paketu R design vzmetenje še nekoliko bolj trdo, se je to samo še
→ bolj potrdilo, a je v tem primeru treba računati na manj udobno
→ vožnjo čez različne asfaltne grbine. Podoben razmislek velja
→ opraviti tudi pri izbiri motorja. Preskušani 2-litrski dizel s 190
→ KM predstavlja vrh ponudbe, ki z močjo, udobjem in tudi povprečno
→ porabo navduši predvsem pri avtocestnih dolgoprogaških izzivih, v
→ počasni mestni vožnji ter pri pogostih postankih in speljevanjih pa
→ deluje preveč robusten. XC 40 je s čvrsto gradnjo, funkcionalno

```


in udobno kabino ter številnimi asistenčnimi sistemi in izstopajočim
→ skandinavskim dizajnom v premišljenem trenutku vstopil na trg modnih
→ mestnih terencev, v katerem se brez ene same sence dvoma suvereno
→ postavi med najdražje in najbolj premijske v mestu. Ključni tehnični
→ podatki:- na testu Volvo XC40 2.0 TD avt awd momentumMere:- dolžina:
→ 4,4 m- medosna razdalja: 2,7 m- obračalni krog: 11,4 m- oddaljenost
→ od tal: 21 cm- prtljažnik: 432 l- masa: 2.250 kgPogon:- 2-litrski
→ 4-valjni bencinski motor- moč: 140 kW- navor: 400 Nm- 8-stopenjski
→ samodejni menjalnik- pogon na vsa štiri 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/kmStroški pri 15.000 km in 5-letni
→ uporabi:- nakupna cena: 43.619 EUR- stroški finančnega leasinga:
→ 3.268 EUR/5 let- stroški registracije: 8.701 EUR/5 let- stroški
→ vzdrževanja: 2.320 EUR/5 let- stroški goriva: 6.190 EUR/75.000 km-
→ strošek 1 kompleta pnevmatik: 923 EUR- vrednosti po 5 letih po
→ Eurotaxu: 18.886 EUR- stroški skupaj: 774 EUR/mesec"

}

A.1.3 Overstock - Jewelry 01

```
{
  "Jewel0": {
    "Title": "10-kt. Seven Diamond Ladies Heart Ring (0.08 TW)",
    "ListPrice": "$149.00",
    "Price": "$69.99",
    "Saving": "$79.01",
    "SavingPercent": "53%",
    "Content": "This ladies fashion ring dazzleswith hearts and
    ↪ diamonds. The gold band is crafted into delicate, open
    ↪ hearts.Seven brilliant-cut diamonds add a bit of sparkle.
    ↪ Click here to purchase."
  },
  "Jewel1": {
    "Title": "10-Kt. Diamond Ring (.25 TW)",
    "ListPrice": "$250.00",
    "Price": "$74.90",
    "Saving": "$175.10",
    "SavingPercent": "70%",
    "Content": "Nineteen round diamonds accent this 10-karat yellow
    ↪ gold ring with filigree accents. Click here to purchase."
  },
  "Jewel2": {
    "Title": "10-kt. Pearl and Diamond Butterfly Earrings",
    "ListPrice": "$149.00",
    "Price": "$42.99",
    "Saving": "$106.01",
    "SavingPercent": "71%",
    "Content": "Perfectly proportioned 5.5- to6-mm cultured pearls on
    ↪ 10-karat yellow gold settings highlight these
    ↪ petiteearrings. A dainty rhodium-plated gold butterfly
    ↪ studded with a diamond (0.02total carat weight, J-K color,
    ↪ I-2 clarity) rests atop each pearl. Click here to purchase."
  },
  "Jewel3": {
    "Title": "14-kt. Diamond 'S' Tennis Bracelet (2.00 TW)",
    "ListPrice": "$1,539.99",
    "Price": "$499.99",
    "Saving": "$1,040.00",
    "SavingPercent": "67%",
    "Content": "Invest in a swirl of light withthis diamond 'S'
    ↪ tennis bracelet. Crafted in 14-karat gold, the piece
    ↪ features49 diamonds for two full carats. The 7.25-inch
    ↪ bracelet closes with a pressureclasp. Click here to
    ↪ purchase."
  },
  "Jewel4": {
    "Title": "10-kt. Diamond Band Fashion Ring (.11 TW)",
    "ListPrice": "$179.99",
```

```

    "Price": "$79.99",
    "Saving": "$100.00",
    "SavingPercent": "55%",
    "Content": "Crafted in white and yellow gold,this ring displays a
    ↪ band of seven round diamonds. Order your new gold anddiamond
    ↪ fashion ring today at our low, online price. Click here to
    ↪ purchase."
  },
  "Jewel5": {
    "Title": "14-kt. White Gold, Pearl and Diamond Ring",
    "ListPrice": "$419.99",
    "Price": "$149.99",
    "Saving": "$270.00",
    "SavingPercent": "64%",
    "Content": "Show your romantic side with this14-karat diamond and
    ↪ pearl ring. Set in a domed band of 14-karat white gold,the
    ↪ ring features a 7-mm cultured pearl. Curved rows of diamonds
    ↪ flank thepearl. Click here to purchase."
  },
  "Jewel6": {
    "Title": "14-kt. Gold Diamond Present Future Pendant (.25TW)",
    "ListPrice": "$299.00",
    "Price": "$149.99",
    "Saving": "$149.01",
    "SavingPercent": "49%",
    "Content": "Designed with three large, sparklingdiamonds to
    ↪ represent past, present, and future, this stunning pendant
    ↪ isset in gleaming 14-karat gold. It incorporates a total of
    ↪ nine diamonds (0.25total carat weight, K color, I-2 to I-3
    ↪ clarity). Click here to purchase."
  },
  "Jewel7": {
    "Title": "14-kt. Diamond Solitaire Pendant (.33 TW)",
    "ListPrice": "$1,019.99",
    "Price": "$319.99",
    "Saving": "$700.00",
    "SavingPercent": "68%",
    "Content": "In this simple, yet elegant pendant,a round brilliant
    ↪ diamond (0.33 total carat weight, H-J color, I-1 to
    ↪ I-2clarity) is prong-set in 14-karat white gold. Click here
    ↪ to purchase."
  },
  "Jewel8": {
    "Title": "14-kt. Diamond Solitaire Earrings (0.33 TW)",
    "ListPrice": "$639.99",
    "Price": "$199.99",
    "Saving": "$440.00",
    "SavingPercent": "68%",

```

```

        "Content": "Dazzle your way into her heart,with these classic
        ↪ diamond solitaire earrings. Two brilliant-cut diamonds(0.33
        ↪ total carat weight, G-H color, I-1 to I-2 clarity) are set
        ↪ in fourprongs of 14-karat white gold. Click here to
        ↪ purchase."
    },
    "Jewel9": {
        "Title": "14-kt. Diamond Cross Pendant (.06 TW)",
        "ListPrice": "$305.00",
        "Price": "$119.99",
        "Saving": "$185.01",
        "SavingPercent": "60%",
        "Content": "Over a cleanly sculpted Romancross of 14-karat white
        ↪ gold drapes a slender banner containing three
        ↪ brightprong-set round diamonds (0.06 total carat weight, H-I
        ↪ color, I clarity). Click here to purchase."
    },
    "Jewel10": {
        "Title": "14-kt. Diamond Solitaire Stud Earrings (.50 TW)",
        "ListPrice": "$999.99",
        "Price": "$359.99",
        "Saving": "$640.00",
        "SavingPercent": "64%",
        "Content": "Every jewelry collection needsa classic pair of
        ↪ diamond solitaire earrings. Set in 14-karat gold,
        ↪ thesediamond stud earrings (0.50 total carat weight) have
        ↪ post backs with butterflyclasps. Click here to purchase."
    },
    "Jewel11": {
        "Title": "14-kt. Cultured Pearl Diamond Earrings",
        "ListPrice": "$508.99",
        "Price": "$179.99",
        "Saving": "$329.00",
        "SavingPercent": "64%",
        "Content": "Create an elegant appearance withthese pearl and
        ↪ diamond stud earrings. Set in 14-karat yellow gold,
        ↪ eachearring features an 8 to 8.5-mm cultured white pearl.
        ↪ Prong-set round diamondsaccent the pearls. Posts with
        ↪ butterfly clasps secure the earrings. Click here to
        ↪ purchase."
    },
    "Jewel12": {
        "Title": "14-kt. Diamond 7.5-8 mm Pearl Pendant",
        "ListPrice": "$196.99",
        "Price": "$69.99",
        "Saving": "$127.00",
        "SavingPercent": "64%",
    }

```

```

        "Content": "Add a classic to your jewelrycollection with this
        ↪ 14-karat gold, diamond, and pearl necklace. The 7.5-8mm
        ↪ cultured white pearl creates the focal point of the pendant,
        ↪ while a diamond(0.10 TW) adds sparkle. Click here to
        ↪ purchase."
    },
    "Jewel13": {
        "Title": "14-kt. Diamond Solitaire Earrings (.50 TW)",
        "ListPrice": "$1,369.99",
        "Price": "$409.99",
        "Saving": "$960.00",
        "SavingPercent": "70%",
        "Content": "This earring set has two brilliant-cutdiamonds (0.50
        ↪ total carat weight, G-H color, I-1 to I-2 clarity) set
        ↪ infour prongs of 14-karat white gold. Click here to
        ↪ purchase."
    },
    "Jewel14": {
        "Title": "14-kt White Gold Diamond Band (0.50 TW)",
        "ListPrice": "$1,635.00",
        "Price": "$609.99",
        "Saving": "$1,025.01",
        "SavingPercent": "62%",
        "Content": "Crafted of 14-karat white gold,this stylish ring
        ↪ features a bright row of 20 channel-set, princess-cut
        ↪ baguetteditiamonds. Treat her like royalty and save when you
        ↪ buy jewelry treasuresat Overstock.com. Click here to
        ↪ purchase."
    }
}

```

A.1.4 Overstock - Jewelry 02

```

{
    "Jewel0": {
        "Title": "14-kt. Green Jade Hoops",
        "ListPrice": "$90.00",
        "Price": "$46.99",
        "Saving": "$43.01",
        "SavingPercent": "47%",
        "Content": "Hoops of cool green jade restbetween 14-karat yellow
        ↪ gold endpieces. The hoops graduate in thickness from3 mm at
        ↪ the ends to 6 mm in the center, with approximately 29 mm
        ↪ overallldiameter. Click here to purchase."
    },
    "Jewel1": {
        "Title": "14-kt. Jade Doughnut Pendant",
        "ListPrice": "$150.00",
        "Price": "$48.99",
        "Saving": "$101.01",
    }
}

```

```

    "SavingPercent": "67%",
    "Content": "The 25-mm disk hangs delicately from a 14-karat gold
    ↪ chain. The disk features a dramatic gold Chinese character in
    ↪ the center, accompanied by four stylized gold bees. Click
    ↪ here to purchase."
  },
  "Jewel2": {
    "Title": "14-kt. Charcoal Jade and Ruby Elephant Pendant",
    "ListPrice": "$100.00",
    "Price": "$28.99",
    "Saving": "$71.01",
    "SavingPercent": "71%",
    "Content": "Carved of rich dark grey jade, this elephant pendant
    ↪ has 14-karat yellow gold applied to mark the feet, tusk,
    ↪ tail, and blanket. A 2-mm round faceted ruby in a gold bezel
    ↪ setting forms the eye. The pendant hangs from an 18-inch
    ↪ chain. Click here to purchase."
  },
  "Jewel3": {
    "Title": "14-kt. Carved Lavender Jade Earrings",
    "ListPrice": "$80.00",
    "Price": "$39.99",
    "Saving": "$40.01",
    "SavingPercent": "50%",
    "Content": "Luscious 8-mm lavender jade balls, carved with
    ↪ intricate Asian style, dangle from a 14-karat yellow gold
    ↪ French hook. Click here to purchase."
  },
  "Jewel4": {
    "Title": "14-kt. Jade Cross Pendant",
    "ListPrice": "$150.00",
    "Price": "$49.99",
    "Saving": "$100.01",
    "SavingPercent": "66%",
    "Content": "Green jade and gold create this beautiful cross
    ↪ pendant. Cylindrical bars of green jade feature caps
    ↪ and center of 14-karat yellow gold. Click here to purchase."
  },
  "Jewel5": {
    "Title": "14-kt. Multicolored Jade Earrings",
    "ListPrice": "$375.00",
    "Price": "$99.99",
    "Saving": "$275.01",
    "SavingPercent": "73%",
    "Content": "A delicate wrapping of 14-karat yellow gold wire holds
    ↪ six 6 x 4 pear shapes of jade in various shades:
    ↪ brilliant green, orange, lavender, black, pale yellow, and
    ↪ white. The post earrings have butterfly backs. Click here to
    ↪ purchase."
  },

```

```

    "Jewel6": {
      "Title": "14-kt. Multicolored Jade Ring",
      "ListPrice": "$250.00",
      "Price": "$56.99",
      "Saving": "$193.01",
      "SavingPercent": "77%",
      "Content": "A delicate wrapping of 14-karat yellow gold wire holds
        ↪ six 6 x 4 ovals of jade in various shades: brilliant green,
        ↪ orange, lavender, black, pale yellow, and white. A narrow
        ↪ gold band divides to support the setting. Click here to
        ↪ purchase."
    },
    "Jewel7": {
      "Title": "14-kt. Onyx and Ruby Elephant Pendant",
      "ListPrice": "$100.00",
      "Price": "$35.99",
      "Saving": "$64.01",
      "SavingPercent": "64%",
      "Content": "Carved of rich black onyx, this elephant pendant has
        ↪ 14-karat yellow gold applied to mark the feet, tusk, tail,
        ↪ and blanket. A 2-mm round faceted ruby in a gold bezel
        ↪ setting forms the eye. The pendant hangs from an 18-inch
        ↪ chain. Click here to purchase."
    }
  }
}

```

A.1.5 Mimovrste - Item 01

```
{
  "Title": "Philips TV sprejemnik 43PFS5503/12",
  "Rating": "Ocena: 4,5",
  "EnergyLevel": "A+",
  "Brand": "Philips",
  "CatalogNumber": "1244513",
  "Stickers": [
    "Brezplačna dostava",
    "Odlična cena",
    "Priporočamo"
  ],
  "Description": "LED televizija z ločljivostjo Full HD in
    ↳ diagonalo 108cm (43"). Pixel Plus HD integriran
    ↳ sprejemnik DVB T/T2 (HEVC)/C/S2, snemanje oddaje, Timeshift, USB
    ↳ predvajalnik, 2 HDMI, 1 USB, optični avdio izhod, priključek za
    ↳ slušalke.Celoten opis",
  "Price": "259 ",
  "BasePrice": "309 ",
  "Save": "Prihranite 50 ",
  "SavingPercent": "16%",
  "Availability": "Na zalogi { predvidena odprema:6.5.",
  "MarketplacePartner": "Mimovrste"
}
```

A.1.6 Mimovrste - Item 02

```
{
  "Title": "MAX MHC32R1B kabel HDMI - 2m, črn",
  "Rating": "Ocena: 5",
  "Brand": "MAX",
  "CatalogNumber": "1112206",
  "Stickers": [
    "Odlična cena",
    "Zadnji kos!"
  ],
  "Description": "Kabel HDMI različice 2.0b s pozlačenima in
    ↳ vrtljivima priključkoma ter dolžine 2m.Celoten opis",
  "Price": "5,49 ",
  "BasePrice": "11,71 ",
  "Save": "Prihranite 6,22 ",
  "SavingPercent": "53%",
  "LastPiece": "Zadnji kos!",
  "Availability": "Na zalogi { predvidena odprema:6.5.",
  "MarketplacePartner": "Mimovrste"
}
```


A.2 XPath

A.2.1 RTV SLO - Audi A6

```
{
  "Author": "Miha Merljak",
  "PublishedTime": "28. december 2018 ob 08:51",
  "Title": "Audi A6 50 TDI quattro: nemir v premijskem razredu",
  "SubTitle": "Test nove generacije",
  "Lead": "To je novi audi A6. V razred najdražjih in najbolj
    ↪ premijskih žrebcev je vnesel nemir, še preden je sploh zapeljal
    ↪ na parkirni prostor, rezerviran za izvršnega direktorja.",
  "Content": "Samo poglejte njegovo masko { to ogromno satovje z
    ↪ radarji na takem položaju, da se ti na avtocesti tudi pri 120
    ↪ km/h vsi spoštljivo umikajo, saj so prepričani, da gre za
    ↪ Pahorjev ali Šarčev avto. Seveda, novi A6 lahko cesto in promet
    ↪ skenira s kar petimi radarji, petimi kamerami, infrardečo kamero
    ↪ za nočni vid, dvanajstimi ultrazvočnimi senzorji in laserskim
    ↪ čitalnikom { lidarjem. V glavnem vojaška tehnologija v službi
    ↪ varnosti za fante, ki smo radi gledali Top Gun, Bonda in druge
    ↪ možakarja s finimi igračami. Če vam pogled na Audijev spisek
    ↪ dodatne opreme ne odvzame volje do življenja, potem vsekakor
    ↪ toplo priporočamo nakup zračnega vzmetenja, saj dobi z njim A6
    ↪ več različnih in vozniško zelo uporabnih karakterjev. Enako velja
    ↪ za seksi luči z inteligentno matrično osvetlitvijo, pa za
    ↪ športno podvozje in vsekakor za štirikolesno krmiljenje. S tem
    ↪ postane A6 med ovinki v občutku na volanu še veliko krajši in
    ↪ bolj agilen. Vse naštetost smo preskušali v družbi agregata 50
    ↪ TDI, ki je v resnici klasični trilitrski dizel, podkrepjen z
    ↪ elektromotorjem. Ja, ta audi je mehki hibrid z izjemnim navorom
    ↪ in dovolj moči kadar koli in kjer koli. Si pa mislimo, da bo
    ↪ največji del trga zadovoljil že učinkovit dvolitrski mehki
    ↪ hibrid z močjo 150 kilovatov.- na testu Audi A6 50 TDI quattro
    ↪ tiptronic- stroški skupaj: 1.001 EUR/mesec"
}
```

A.2.2 RTV SLO - Volvo XC 40

```
{
  "Author": "Miha Merljak",
  "PublishedTime": "25. januar 2019 ob 15:23",
  "Title": "Volvo XC 40 D4 AWD momentum: suvereno med najboljše v
    ↪ razredu",
  "SubTitle": "Test novega modela",
  "Lead": "XC 40 je najmanjši Volvov SUV, ki se oblikovno skoraj v
    ↪ celotni naslanja na oba večja predhodnika. Že samo s tem so mu
    ↪ vrata do denarnic tistih kupcev, ki iščejo izstopajočo, a hkrati
    ↪ visoko kultivirano in prečiščeno dizajnersko govorico, na pol
    ↪ odprta.",
}
```

"Content": "Volvo se je nižjih srednjih razredov v preteklosti
 ↳ izogibal ali pa je vanje vstopal z zelo nišnimi produkti, ki
 ↳ niso pustili večjega tržnega pečata. V primeru XC 40 ni težko
 ↳ napovedati, da bo ta tradicija prekinjena. Ponuja namreč visoko
 ↳ kakovost končne izdelave in v kabini odlično premišljeno
 ↳ funkcionalnost ter na dotik prijetne materiale. Še posebej
 ↳ hvalimo število, iznajdljivost in velikost različnih odlagalnih
 ↳ prostorov ter široke, čvrste in zelo udobne sedeže. Intuitivno
 ↳ in enostavno logično je upravljanje z velikim vmesnikom, ki z
 ↳ večfunkcijskim zaslonom na dotik kraljuje na z roko lahko
 ↳ dostopnem mestu na sredinski armaturi. Razočaranj ne bo niti v
 ↳ velikosti in uporabnosti prtljažnega prostora, ki s 460 litri
 ↳ prostornine sicer ni med večjimi v razredu, a se v uporabniškem
 ↳ smislu odkupi z dobro urejenostjo ter domiselnimi rešitvami
 ↳ pregrajevanja.XC 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 nevtrarno in predvidljivo
 ↳ lego. V premeru preskušane modela, ki je imel v paketu R
 ↳ design vzmetenje še nekoliko bolj trdo, se je to samo še bolj
 ↳ potrdilo, a je v tem primeru treba računati na manj udobno
 ↳ vožnjo čez različne asfaltne grbine. Podoben razmislek velja
 ↳ opraviti tudi pri izbiri motorja.Preskušani 2-litrski dizel s
 ↳ 190 KM predstavlja vrh ponudbe, ki z močjo, udobjem in tudi
 ↳ povprečno porabo navduši predvsem pri avtocestnih dolgoprogaških
 ↳ izzivih, v počasni mestni vožnji ter pri pogostih postankih in
 ↳ speljevanjih pa deluje preveč robusten.XC 40 je s čvrsto
 ↳ gradnjo, funkcionalno in udobno kabino ter številnimi
 ↳ asistenčnimi sistemi in izstopajočim skandinavskim dizajnom v
 ↳ premišljenem trenutku vstopil na trg modnih mestnih terencev, v
 ↳ katerem se brez ene same sence dvoma suvereno postavi med
 ↳ najdražje in najbolj premijske v mestu.Mere:Stroški pri 15.000
 ↳ km in 5-letni uporabi:"

}

A.2.3 Overstock - Jewelry 01

```
{
  "Jewel0": {
    "Title": "10-kt. Seven Diamond Ladies Heart Ring (0.08 TW)",
    "ListPrice": "$149.00",
    "Price": "$69.99",
    "Saving": "$79.01",
    "SavingPercent": "53%",
    "Content": "This ladies fashion ring dazzleswith hearts and
    ↪ diamonds. The gold band is crafted into delicate, open
    ↪ hearts.Seven brilliant-cut diamonds add a bit of sparkle."
  },
  "Jewel1": {
    "Title": "10-Kt. Diamond Ring (.25 TW)",
    "ListPrice": "$250.00",
    "Price": "$74.90",
    "Saving": "$175.10",
    "SavingPercent": "70%",
    "Content": "Nineteen round diamonds accent this 10-karat yellow
    ↪ gold ring with filigree accents."
  },
  "Jewel2": {
    "Title": "10-kt. Pearl and Diamond Butterfly Earrings",
    "ListPrice": "$149.00",
    "Price": "$42.99",
    "Saving": "$106.01",
    "SavingPercent": "71%",
    "Content": "Perfectly proportioned 5.5- to6-mm cultured pearls on
    ↪ 10-karat yellow gold settings highlight these
    ↪ petiteearrings. A dainty rhodium-plated gold butterfly
    ↪ studded with a diamond (0.02total carat weight, J-K color,
    ↪ I-2 clarity) rests atop each pearl."
  },
  "Jewel3": {
    "Title": "14-kt. Diamond 'S' Tennis Bracelet (2.00 TW)",
    "ListPrice": "$1,539.99",
    "Price": "$499.99",
    "Saving": "$1,040.00",
    "SavingPercent": "67%",
    "Content": "Invest in a swirl of light withthis diamond 'S'
    ↪ tennis bracelet. Crafted in 14-karat gold, the piece
    ↪ features49 diamonds for two full carats. The 7.25-inch
    ↪ bracelet closes with a pressureclasp."
  },
  "Jewel4": {
    "Title": "10-kt. Diamond Band Fashion Ring (.11 TW)",
    "ListPrice": "$179.99",
    "Price": "$79.99",
    "Saving": "$100.00",
```

```

    "SavingPercent": "55%",
    "Content": "Crafted in white and yellow gold,this ring displays a
    ↪ band of seven round diamonds. Order your new gold anddiamond
    ↪ fashion ring today at our low, online price."
  },
  "Jewel5": {
    "Title": "14-kt. White Gold, Pearl and Diamond Ring",
    "ListPrice": "$419.99",
    "Price": "$149.99",
    "Saving": "$270.00",
    "SavingPercent": "64%",
    "Content": "Show your romantic side with this14-karat diamond and
    ↪ pearl ring. Set in a domed band of 14-karat white gold,the
    ↪ ring features a 7-mm cultured pearl. Curved rows of diamonds
    ↪ flank thepearl."
  },
  "Jewel6": {
    "Title": "14-kt. Gold Diamond Present Future Pendant (.25TW)",
    "ListPrice": "$299.00",
    "Price": "$149.99",
    "Saving": "$149.01",
    "SavingPercent": "49%",
    "Content": "Designed with three large, sparklingdiamonds to
    ↪ represent past, present, and future, this stunning pendant
    ↪ isset in gleaming 14-karat gold. It incorporates a total of
    ↪ nine diamonds (0.25total carat weight, K color, I-2 to I-3
    ↪ clarity)."
  },
  "Jewel7": {
    "Title": "14-kt. Diamond Solitaire Pendant (.33 TW)",
    "ListPrice": "$1,019.99",
    "Price": "$319.99",
    "Saving": "$700.00",
    "SavingPercent": "68%",
    "Content": "In this simple, yet elegant pendant,a round brilliant
    ↪ diamond (0.33 total carat weight, H-J color, I-1 to
    ↪ I-2clarity) is prong-set in 14-karat white gold."
  },
  "Jewel8": {
    "Title": "14-kt. Diamond Solitaire Earrings (0.33 TW)",
    "ListPrice": "$639.99",
    "Price": "$199.99",
    "Saving": "$440.00",
    "SavingPercent": "68%",
    "Content": "Dazzle your way into her heart,with these classic
    ↪ diamond solitaire earrings. Two brilliant-cut diamonds(0.33
    ↪ total carat weight, G-H color, I-1 to I-2 clarity) are set
    ↪ in fourprongs of 14-karat white gold."
  },
  "Jewel9": {

```

```

        "Title": "14-kt. Diamond Cross Pendant (.06 TW)",
        "ListPrice": "$305.00",
        "Price": "$119.99",
        "Saving": "$185.01",
        "SavingPercent": "60%",
        "Content": "Over a cleanly sculpted Romancross of 14-karat white
        ↪ gold drapes a slender banner containing three
        ↪ brightprong-set round diamonds (0.06 total carat weight, H-I
        ↪ color, I clarity).",
    },
    "Jewel10": {
        "Title": "14-kt. Diamond Solitaire Stud Earrings (.50 TW)",
        "ListPrice": "$999.99",
        "Price": "$359.99",
        "Saving": "$640.00",
        "SavingPercent": "64%",
        "Content": "Every jewelry collection needsa classic pair of
        ↪ diamond solitaire earrings. Set in 14-karat gold,
        ↪ thesediamond stud earrings (0.50 total carat weight) have
        ↪ post backs with butterflyclasps.",
    },
    "Jewel11": {
        "Title": "14-kt. Cultured Pearl Diamond Earrings",
        "ListPrice": "$508.99",
        "Price": "$179.99",
        "Saving": "$329.00",
        "SavingPercent": "64%",
        "Content": "Create an elegant appearance withthese pearl and
        ↪ diamond stud earrings. Set in 14-karat yellow gold,
        ↪ eachearring features an 8 to 8.5-mm cultured white pearl.
        ↪ Prong-set round diamondsaccent the pearls. Posts with
        ↪ butterfly clasps secure the earrings.",
    },
    "Jewel12": {
        "Title": "14-kt. Diamond 7.5-8 mm Pearl Pendant",
        "ListPrice": "$196.99",
        "Price": "$69.99",
        "Saving": "$127.00",
        "SavingPercent": "64%",
        "Content": "Add a classic to your jewelrycollection with this
        ↪ 14-karat gold, diamond, and pearl necklace. The 7.5-8mm
        ↪ cultured white pearl creates the focal point of the pendant,
        ↪ while a diamond(0.10 TW) adds sparkle.",
    },
    "Jewel13": {
        "Title": "14-kt. Diamond Solitaire Earrings (.50 TW)",
        "ListPrice": "$1,369.99",
        "Price": "$409.99",
        "Saving": "$960.00",
        "SavingPercent": "70%",
    }
}

```

```

    "Content": "This earring set has two brilliant-cutdiamonds (0.50
    ↪ total carat weight, G-H color, I-1 to I-2 clarity) set
    ↪ infour prongs of 14-karat white gold."
  },
  "Jewel14": {
    "Title": "14-kt White Gold Diamond Band (0.50 TW)",
    "ListPrice": "$1,635.00",
    "Price": "$609.99",
    "Saving": "$1,025.01",
    "SavingPercent": "62%",
    "Content": "Crafted of 14-karat white gold,this stylish ring
    ↪ features a bright row of 20 channel-set, princess-cut
    ↪ baguettediamonds. Treat her like royalty and save when you
    ↪ buy jewelry treasuresat Overstock.com."
  }
}

```

A.2.4 Overstock - Jewelry 02

```

{
  "Jewel0": {
    "Title": "14-kt. Green Jade Hoops",
    "ListPrice": "$90.00",
    "Price": "$46.99",
    "Saving": "$43.01",
    "SavingPercent": "47%",
    "Content": "Hoops of cool green jade restbetween 14-karat yellow
    ↪ gold endpieces. The hoops graduate in thickness from3 mm at
    ↪ the ends to 6 mm in the center, with approximately 29 mm
    ↪ overallldiameter."
  },
  "Jewel1": {
    "Title": "14-kt. Jade Doughnut Pendant",
    "ListPrice": "$150.00",
    "Price": "$48.99",
    "Saving": "$101.01",
    "SavingPercent": "67%",
    "Content": "The 25-mm disk hangs delicatelyfrom a 14-karat gold
    ↪ chain. The disk features a dramatic gold Chinese characterin
    ↪ the center, accompanied by four stylized gold bees."
  },
  "Jewel2": {
    "Title": "14-kt. Charcoal Jade and Ruby Elephant Pendant",
    "ListPrice": "$100.00",
    "Price": "$28.99",
    "Saving": "$71.01",
    "SavingPercent": "71%",

```

```

        "Content": "Carved of rich dark grey jade,this elephant pendant
        ↪ has 14-karat yellow gold applied to mark the feet,tusk,
        ↪ tail, and blanket. A 2-mm round faceted ruby in a gold bezel
        ↪ settingforms the eye. The pendant hangs from an 18-inch
        ↪ chain."
    },
    "Jewel3": {
        "Title": "14-kt. Carved Lavender Jade Earrings",
        "ListPrice": "$80.00",
        "Price": "$39.99",
        "Saving": "$40.01",
        "SavingPercent": "50%",
        "Content": "Luscious 8-mm lavender jade balls, carved with
        ↪ intricate Asian style, dangle from a 14-karat yellow gold
        ↪ French hook."
    },
    "Jewel4": {
        "Title": "14-kt. Jade Cross Pendant",
        "ListPrice": "$150.00",
        "Price": "$49.99",
        "Saving": "$100.01",
        "SavingPercent": "66%",
        "Content": "Green jade and gold create thisbeautiful cross
        ↪ pendant. Cylindrical bars of green jade feature caps
        ↪ andcenter of 14-karat yellow gold."
    },
    "Jewel5": {
        "Title": "14-kt. Multicolored Jade Earrings",
        "ListPrice": "$375.00",
        "Price": "$99.99",
        "Saving": "$275.01",
        "SavingPercent": "73%",
        "Content": "A delicate wrapping of 14-karatyellow gold wire holds
        ↪ six 6 x 4 pear shapes of jade in various shades:
        ↪ brilliantgreen, orange, lavender, black, pale yellow, and
        ↪ white. The post earrings have butterfly backs."
    },
    "Jewel6": {
        "Title": "14-kt. Multicolored Jade Ring",
        "ListPrice": "$250.00",
        "Price": "$56.99",
        "Saving": "$193.01",
        "SavingPercent": "77%",
        "Content": "A delicate wrapping of 14-karatyellow gold wire holds
        ↪ six 6 x 4 ovals of jade in various shades: brilliantgreen,
        ↪ orange, lavender, black, pale yellow, and white. A narrow
        ↪ gold banddivides to support the setting."
    },
    "Jewel7": {
        "Title": "14-kt. Onyx and Ruby Elephant Pendant",

```

```

    "ListPrice": "$100.00",
    "Price": "$35.99",
    "Saving": "$64.01",
    "SavingPercent": "64%",
    "Content": "Carved of rich black onyx, this elephant pendant has
↳ 14-karat yellow gold applied to mark the feet, tusk, tail,
↳ and blanket. A 2-mm round faceted ruby in a gold bezel
↳ setting forms the eye. The pendant hangs from an 18-inch
↳ chain."
  }
}

```


A.2.5 Mimovrste - Item 01

```
{
  "Title": "Philips TV sprejemnik 43PFS5503/12",
  "Rating": "Ocena: 4,5",
  "EnergyLevel": "A+",
  "Brand": "Philips",
  "CatalogNumber": "1244513",
  "Stickers": [
    "Brezplačna dostava",
    "Odlična cena",
    "Priporočamo"],
  "Description": "LED televizija z ločljivostjo Full HD in diagonalo
↳ 108 cm (43\"). Pixel Plus HD integriran sprejemnik DVB T/T2
↳ (HEVC)/C/S2, snemanje oddaje , Timeshift, USB predvajalnik, 2
↳ HDMI , 1 USB, optični avdio izhod, priključek za slušalke.
↳ Celoten opis",
  "Price": "259 ",
  "BasePrice": "309 ",
  "SavingPercent": "16 %",
  "Availability": "Na zalogi { predvidena odprema:\t\t\t6.5.",
  "MarketplacePartner": "Mimovrste"
}
```

A.2.6 Mimovrste - Item 02

```
{
  "Title": "MAX MHC32R1B kabel HDMI - 2m, črn",
  "Rating": "Ocena: 5",
  "Brand": "MAX",
  "CatalogNumber": "1112206",
  "Stickers": [
    "Odlična cena"
  ],
  "Description": "Kabel HDMI različice 2.0b s pozlačenima in vrtljivima
↳ priključkoma ter dolžine 2 m. Celoten opis",
  "Price": "5,49 ",
  "BasePrice": "11,71 ",
  "SavingPercent": "53 %",
  "Availability": "Na zalogi { predvidena odprema:\t\t\t6.5.",
  "MarketplacePartner": "Mimovrste",
  "LastPiece": "Zadnji kos!"
}
```

A.3 RoadRunner

A.3.1 Example

```
{  "site1": {
    "content": "<HTML>
        Books of:
        <B>
            Paul Jones
        </B>
        <IMG src="#\" />
        <UL>
            <LI>
                <I>Title:</I>
                XML at Work
            </LI>
            <LI>
                <I> Title: </I>
                HTML Scripts
            </LI>
            <LI>
                <I>Title:</I>
                Javascript
            </LI>
        </UL>
        <b>Paul Jones has received no awards yet</b>
    </HTML>"
  }
  "site2": {
    "content": "<HTML>
        Books of:
        <B>
            Paul Jones
        </B>
        <IMG src="#\" />
        <UL>
            <LI>
                <I>Title:</I>
                XML at Work
            </LI>
            <LI>
                <I> Title: </I>
                HTML Scripts
            </LI>
            <LI>
                <I>Title:</I>
                Javascript
            </LI>
        </UL>
        <b>Paul Jones has received no awards yet</b>
```

```

        </HTML>"
    }
    "wrapper": "<html>Books of:<b>#PCDATA</b><(img
    ↪   src="#"/>)?<ul><(li><i>Title:</i>#PCDATA</li>)+</ul><(b>#PCDATA</b>)+</html>"
}

```