

BAB III

METODOLOGI PENELITIAN

Bab ini menjelaskan tentang gambaran tahapan yang dilakukan untuk menganalisis data untuk menjawab perumusan masalah sehingga dapat mencapai tujuan sebenarnya dari penelitian. Pada metodologi penelitian akan dijelaskan tentang tahapan dari penelitian.

3.1. Analisis Permasalahan

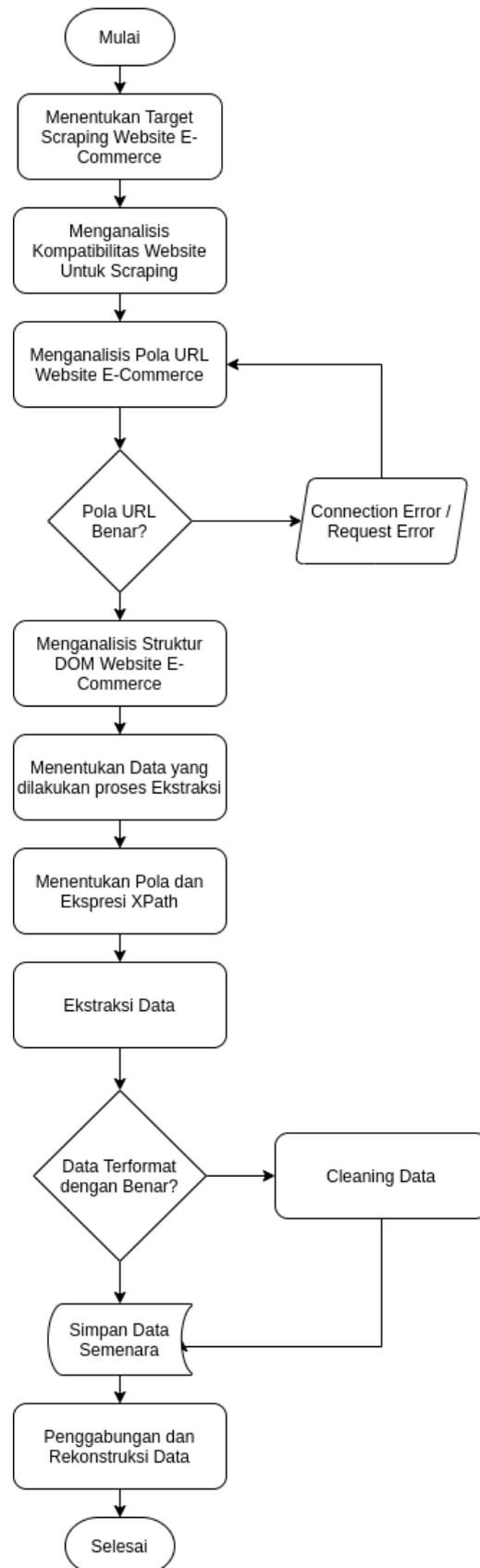
Permasalahan yang sering terjadi pada saat melakukan pencarian produk di situs belanja online yaitu banyaknya jumlah situs belanja online serta produk yang ditawarkan sangat banyak dan beragam yang kemudian menimbulkan kebingungan bagi konsumen untuk mencari produk yang diinginkan sehingga mengakibatkan kita harus melakukan pencarian produk di tiap-tiap situs belanja online dan berdampak menghabiskan terlalu banyak waktu hanya untuk melakukan pencarian produk diberbagai situs belanja online.

Selain produk yang ditawarkan cukup banyak, kita terkadang sulit membandingkan antara produk yang satu dengan produk lain yang serupa terutama antar situs belanja karena selain struktur produk yang berbeda dari tiap situs belanja, ketika melakukan perbandingan antara produk yang satu dengan yang lain terkadang menimbulkan keraguan pada saat membeli produk tersebut dikarenakan faktor faktor tertentu seperti kualitas maupun harga yang beragam dari situs belanja.

3.2. Solusi yang Diusulkan

Berdasarkan analisis tersebut, solusi yang diusulkan adalah mengimplementasikan Sistem Informasi Pencarian Produk dengan teknik webscraping pada website *e-commerce* dengan menggunakan metode XPath untuk membantu dalam melakukan proses pencarian produk dari berbagai macam situs belanja online. XPath merupakan salah satu metode yang digunakan untuk melakukan ekstraksi atau pengambilan data pada halaman markup XML maupun HTML, ekstraksi data dilakukan dengan menerapkan sintaks atau ekspresi XPath pada halaman markup tersebut sehingga mendapatkan data yang ingin kita ambil.

Untuk menerapkan solusi yang telah dijelaskan sebelumnya, maka perlu adanya kerangka kerja atau *framework* yang sistematis. *Framework* merupakan model alur kerja sistem yang akan dilakukan dalam penyelesaian masalah yang akan dibahas. Berikut merupakan *framework* yang tertera pada gambar 3.1.



Gambar 3. 1 Model Sistem Yang Diusulkan

3.2.1. Menentukan Target Scraping Website *E-Commerce*.

Terdapat banyak sekali website *e-commerce* yang tersedia, pada tahapan awal ini kita menentukan target website yang akan kita lakukan proses scraping. Pada mulanya target awal website scraping yaitu seperti Bukalapak, Tokopedia, Elevenia, Shopee, OLX, Uqishop, Zalora, dan Sorabel.

3.2.2. Menganalisis Kompatibilitas Website *E-Commerce*.

Menganalisis Kompatibilitas Website E-Commerce dilakukan untuk melakukan pengecekan apakah website tersebut dapat dilakukan scraping atau tidak. Langkah sederhana yang dapat dilakukan yaitu dengan mengunjungi website tersebut dan mengakses file robots.txt setelah *base http url*, contoh *https://example.com/robots.txt*. Robots.txt berfungsi untuk mengontrol dan mengatur halaman atau direktori mana yang boleh ditampilkan atau diindex oleh mesin pencari. File ini digunakan terutama untuk menghindarkan situs Anda dari beban request atau permintaan yang terlampau banyak. berikut ini daftar robots.txt dari target awal situs belanja yang akan dilakukan webscraping:

1. Bukalapak

Berikut ini adalah file robots.txt yang terdapat pada situs Bukalapak seperti yang dapat dilihat pada gambar 3.2.

```

User-agent: *
Disallow: /*/admin_link$
Disallow: /*/edit_label
Disallow: /*/one_click_form$
Disallow: /*/seller_actions$
Disallow: /*/tiny_link_to_cart$
Disallow: /*/user_manage$
Disallow: /account_settings/
Disallow: /api/v2/
Disallow: /ca/
Disallow: /campaign/
Disallow: /cart/
Disallow: /css/
Disallow: /components/products/related_deal/
Disallow: /fb_login
Disallow: /feed/feed.rss
Disallow: /gplus_login
Disallow: /js/
Disallow: /messages/
Disallow: /online/
Disallow: /palaver/
Disallow: /payment/
Disallow: /products/*/bookmarker$
Disallow: /products/*/recommended_bid_value
Disallow: /products/*_*
Disallow: /products/load_image_gallery
Disallow: /product_discussion/vote/*
Disallow: /product_reviews/votes/*
Disallow: /product_reviews/vote/*
Disallow: /recomendations/*/owner_actions
Disallow: /system/
Disallow: /system2/
Disallow: /system3/
Disallow: /system4/
Disallow: /users/*/inactivity
Disallow: /users/login_menu
Disallow: /users/*/report
Disallow: /v2/
Disallow: /u/*/label/bukamobil
Disallow: /cdn-cgi/

```

Gambar 3. 2 File Robots.txt Website Bukalapak

2. Tokopedia

Berikut ini adalah file robots.txt yang terdapat pada situs Tokopedia seperti yang dapat dilihat pada gambar 3.3.

```
User-agent: *
Disallow: /search?*
Disallow: /search/*
Disallow: /flight/search/*
Disallow: /events/search*
Disallow: /image-search/
Disallow: /kartu-kredit?*id=*
Disallow: /*.pl
Disallow: /amp/api/*
Disallow: /feed?sc=*
Disallow: /graphql
Disallow: /reputationapp/*
Disallow: /insight/*
Disallow: */tokopedia-lite-production/
Disallow: /similar-products*
Disallow: /provi/check*
Disallow: /**/talk
Disallow: /**/review
Disallow: /content/*
Disallow: /archive-*
Disallow: /wishlist?*
Disallow: /tokopoints
Disallow: /rekomendasi*/d/
Allow: /tokopoints/intro/
Disallow: /careers/*
Allow: /careers/jobs/
Allow: /careers/function/internship/
Disallow: /people/*
Disallow: /p/tour-travel
```

Gambar 3. 3 File Robots.txt Website Tokopedia

3. Elevenia

Berikut ini adalah file robots.txt yang terdapat pada situs Elevenia seperti yang dapat dilihat pada gambar 3.4.

```

User-agent: *
Disallow: /login*
Disallow: /myelevenia*
Disallow: /cart*

Disallow: /promo-kesehatan-puasa-675945
Disallow: /promo-lagu-religi-islami-674958
Disallow: /promo-tentang-puasa-lebaran-676951
Disallow: /promo-menu-buka-puasa-sahur-674959
Disallow: /promo-tips-cantik-hijab-676952
Disallow: /promo-model-baju-rambut-lebaran-674960
Disallow: /promo-panduan-resep-kue-sayur-lebaran-676953
Disallow: /promo-info-arus-mudik-674961
Disallow: /promo-selamat-hari-ray-idul-fitri-674962
Disallow: /promo-peta-mudik-lebaran-674954
Disallow: /promo-peta-mudik-jakarta-676946
Disallow: /promo-peta-mudik-semarang-676947
Disallow: /promo-peta-mudik-yogyakarta-674955
Disallow: /promo-peta-mudik-solo-676948
Disallow: /promo-peta-mudik-malang-676949
Disallow: /promo-peta-mudik-surabaya-676950
Disallow: /promo-peta-mudik-bali-674956
Disallow: /promo-peta-mudik-bandung-674957

```

Gambar 3. 4 File Robots.txt Website Elevenia

4. Shopee

Berikut ini adalah file robots.txt yang terdapat pada situs Shopee seperti yang dapat dilihat pada gambar 3.5.

```

User-Agent:*
Crawl-delay:1
Disallow: /cart/
Disallow: /checkout/
Disallow: /buyer/
Disallow: /user/
Disallow: /me/
Disallow: /order/
Disallow: /daily_discover/
Disallow: /mall/just-for-you/
Disallow: /mall/*-cat.
Disallow: /from_same_shop/
Disallow: /you_may_also_like/
Disallow: *-i.*/similar
Disallow: /find_similar_products/
Disallow: /top_products
Disallow: /search*searchPrefill
Disallow: /index.html

```

Gambar 3. 5 File Robots.txt Website Shopee

5. OLX

Berikut ini adalah file robots.txt yang terdapat pada situs OLX seperti yang dapat dilihat pada gambar 3.6.

```
User-agent: *
Disallow: /api/
Disallow: /post/
Disallow: /edit/
Disallow: /account/
Disallow: /chat/
Disallow: /profile/
Disallow: /payments/
Disallow: /nf/
Disallow: /items$

#Base Filters
Disallow: /*?*[*
Disallow: /*?*%5B
Disallow: *_and_*
Disallow: /*?*filter=*price_
Disallow: /*?*filter=*condition_

#Cars Filters
Disallow: /*?*filter=*cartype_
Disallow: /*?*filter=*mileage_
Disallow: /*?*filter=*transmission_
Disallow: /*?*filter=*year_
Disallow: /*?*filter=*m_exchange_eq_
Disallow: /*?*filter=*m_engine_capacity_eq

#RE Filters
Disallow: /*?*filter=*bathrooms_
Disallow: /*?*filter=*bedrooms_
Disallow: /*?*filter=*brokerfree_
Disallow: /*?*filter=*pets_
Disallow: /*?*filter=*surface_
Disallow: /*?*filter=*furnished_
```

Gambar 3. 6 File Robots.txt Website OLX

6. Uqishop

Berikut ini adalah file robots.txt yang terdapat pada situs Uqishop seperti yang dapat dilihat pada gambar 3.7.

```
User-agent: *
Disallow: /wp-admin/
Allow: /wp-admin/admin-ajax.php
```

Gambar 3. 7 File Robots.txt Website Uqishop

7. Zalora

Berikut ini adalah file robots.txt yang terdapat pada situs Zalora seperti yang dapat dilihat pada gambar 3.8.

```
User-agent: *
Allow: /_c/v1/lite/list_filters
Allow: /_c/v1/desktop/list_catalog_full
Disallow: /assets/
Disallow: /images/
Disallow: /product-index/*?
Disallow: /_c/
```

Gambar 3. 8 File Robots.txt Website Zalora

8. Sorabel

Berikut ini adalah file robots.txt yang terdapat pada situs Sorabel seperti yang dapat dilihat pada gambar 3.9.

```
User-agent: *
Disallow: /profile/
Disallow: /menu/
Disallow: /order/
Disallow: /account/
Disallow: /cart/
Disallow: /undang/
Sitemap: https://www.sorabel.com/sitemap.xml
```

Gambar 3. 9 File Robots.txt Website Sorabel

Dari beberapa File Robots.txt diatas dapat disimpulkan bahwa ada beberapa situs belanja yang tidak memperbolehkan bots, scraper maupun crawler untuk mengakses situs untuk melakukan operasi pencarian produk, salah satu contoh pada situs shopee, pada web tersebut *User Agent: ** artinya semuanya yang mengakses ada beberapa yang tidak diperbolehkan seperti *Disallow: /search*searchPrefills* untuk pencarian produk sehingga menyebabkan website tersebut tidak dapat kita lakukan scraping secara langsung dengan memasukkan parameter request, selain shopee ada website yang lain yang tidak dapat kita lakukan scraping seperti Bukalapak,

Tokopedia maupun situs belanja yang lain. Sehingga mengerucutkan penelitian ini pada 4 Website yaitu Elevenia, Olx, Uqishop, dan Sorabel.

3.2.3. Menganalisis Pola Url Target Website E-Commerce

Dari ke 4 website tersebut kemudian kita kembangkan dengan melakukan analisis dengan melakukan pencarian pada tiap-tiap website dan menemukan pola url yang cocok untuk mendukung dalam proses pencarian. Pola url tersebut dapat dilihat seperti pada Tabel 3.1.

Tabel 3. 1 Pola URL Website E-Commerce Dengan Keyword Pencarian

No	Website	Base URL	Joined Keyword
1	OLX	https://www.olx.co.id	/items/q-{search_keyword}
2	Elevenia	https://www.elevenia.co.id	/search?q={search_keyword}
3	Uqishop	https://www.uqishop.com	?s={search_keyword}&post_type=product
4	Sorabel	https://www.sorabel.com	/products/search?query={search_keyword}

Dari Tabel diatas dapat kita cermati bahwa tiap-tiap website mempunyai ciri khas tersendiri dalam menangani request pencarian pada website tersebut. Kita dapat melakukan hal yang sama untuk website *e-commerce* yang lain, namun harus menyesuaikan dengan target url website masing-masing yang akan kita lakukan proses scraping karena pola tiap-tiap url berbeda.

Setelah mendapatkan pola-pola tersebut kita dapat melakukan uji coba dengan melakukan request dengan menggunakan web browser apakah pola-pola tersebut sudah benar, jika tidak maka halaman pencarian tidak akan

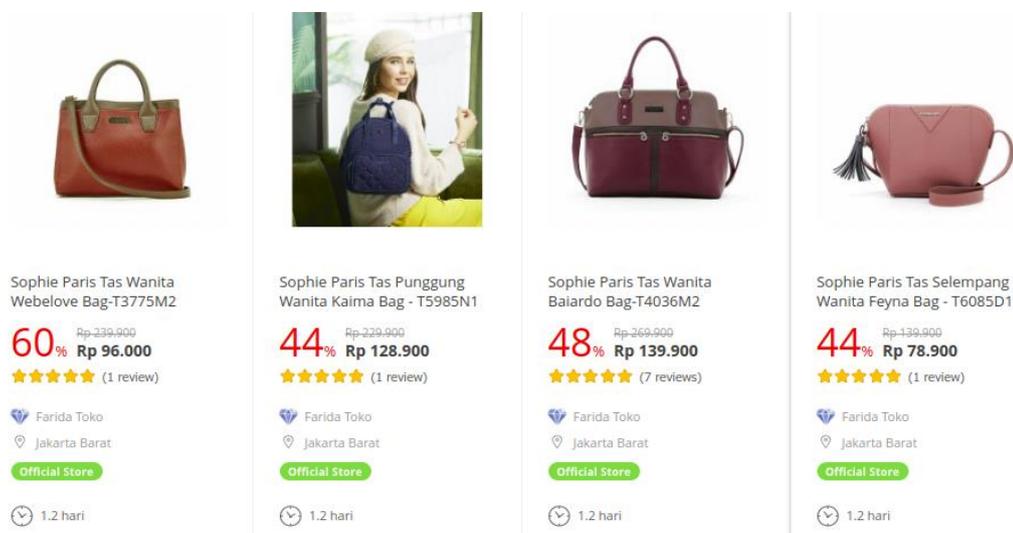
muncul atau bahkan muncul pesan error bahwa halaman yang diakses tidak tersedia, sehingga kira perlu melakukan *tweak* atau perbaikan dan menemukan pola yang optimal untuk pencarian produk pada tiap-tiap situs.

3.2.4. Menganalisis Struktur DOM Website *E-Commerce*

Menganalisis Struktur DOM dilakukan untuk mengetahui Tree HTML yang terkandung pada website yang akan kita analisa, pada tahapan ini kita melakukan analisis struktur untuk mengetahui jenis tag, atribut, properti maupun data yang akan kita ambil, data yang akan kita ambil yaitu terkait produk, sehingga kita hanya melakukan analisis terhadap struktur DOM produk tersebut, berikut merupakan struktur DOM dari tiap-tiap website yang akan kita lakukan proses scraping:

1. Elevenia

Berikut ini merupakan tampilan antarmuka pencarian produk pada halaman website Elevenia seperti yang dapat dilihat pada Gambar 3. 10.



Gambar 3. 10 Tampilan Antarmuka Pencarian Produk Halaman Elevenia

Dari antarmuka halaman pencarian pada website Elevenia tersebut kemudian didapatkan struktur HTML produknya yang dapat dilihat pada Gambar 3. 11.

```
<li id="prod_28635549" class="itemList" data-prdno="28635549">
  <div class="group">
    <a href="http://www.elevenia.co.id/
    prd-sophie-paris-tas-wanita-webelove-bag-t3775m2-28635549" class="img">
      
    </a>
    <a href="http://www.elevenia.co.id/
    prd-sophie-paris-tas-wanita-webelove-bag-t3775m2-28635549" class="pordLink notranslate">
    Sophie Paris Tas Wanita Webelove Bag-T3775M2</a>
    <div class="price notranslate"> ***
    </div>
    <div class="rankingArea"> ***
    </div>
    <ul class="seller notranslate"> ***
    </ul>
    <ul class="sellerPlace"> ***
    </ul>
    <ul class="officialBadge"> ***
    </ul>
  </div>
</li>
```

Gambar 3. 11 Potongan Struktur HTML Pencarian Produk Halaman Elevenia

2. OLX

Berikut ini merupakan tampilan antarmuka pencarian produk pada halaman website OLX seperti yang dapat dilihat pada Gambar 3. 12.



Gambar 3. 12 Tampilan Antarmuka Pencarian Produk Halaman OLX

Dari antarmuka halaman pencarian pada website OLX tersebut kemudian didapatkan struktur HTML produknya yang dapat dilihat pada Gambar 3. 13.

```

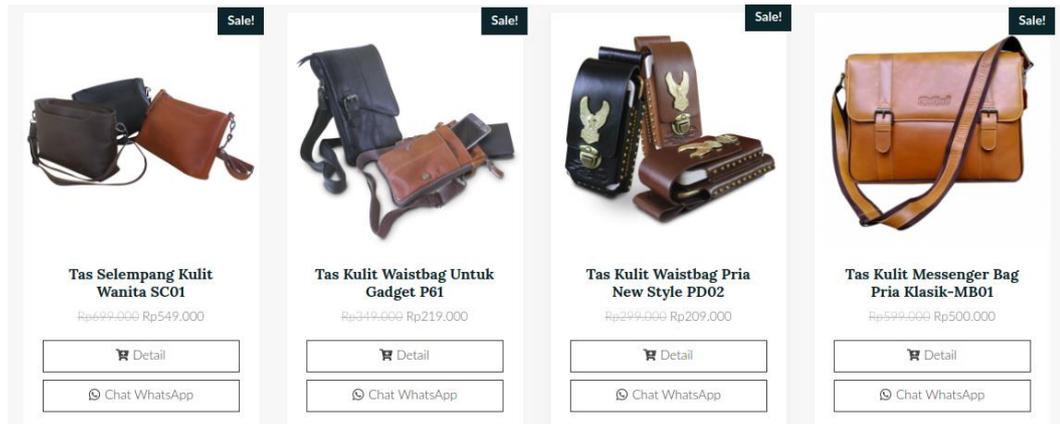
<li data-aut-id="itemBox" class="EIR5N">
  <a href="/item/fashi-tas-selempang-kasual-crossbody-bag-bahan-canvas-khaki-iid-784673936">
    <figure class="_2grx4" data-aut-id="ItemImage">
      
      <noscript></noscript>
    </figure>
    <div class="IKo3_">
      <span class="_89yzn" data-aut-id="itemPrice">Rp 80.000</span>
      <span class="_2tW1I" data-aut-id="itemTitle">Fashi Tas Selempang Kasual Crossbody Bag Bahan Canvas-Khaki</span>
      <div class="_1KOFM">
        <span class="tjgMj" data-aut-id="item-location">Batununggal, Bandung Kota</span>
        <span class="zLvFQ">
          <span>Hari ini</span>
        </span>
      </div>
    </div>
  </a>
  <span class="_2LYp_">
    <button type="button" class="rui-3mp0_" role="button" tabindex="0" data-aut-id="" title="Favourite">
      <svg width="24px" height="24px" viewBox="0 0 1024 1024" data-aut-id="icon" class="" fill-rule="evenodd">
      </svg>
    </button>
  </span>
</li>

```

Gambar 3. 13 Potongan Struktur HTML Pencarian Produk Halaman OLX

3. Uqishop

Berikut ini merupakan tampilan antarmuka pencarian produk pada halaman website Uqishop seperti yang dapat dilihat pada Gambar 3. 14.



Gambar 3. 14 Tampilan Antarmuka Pencarian Produk Halaman Uqishop

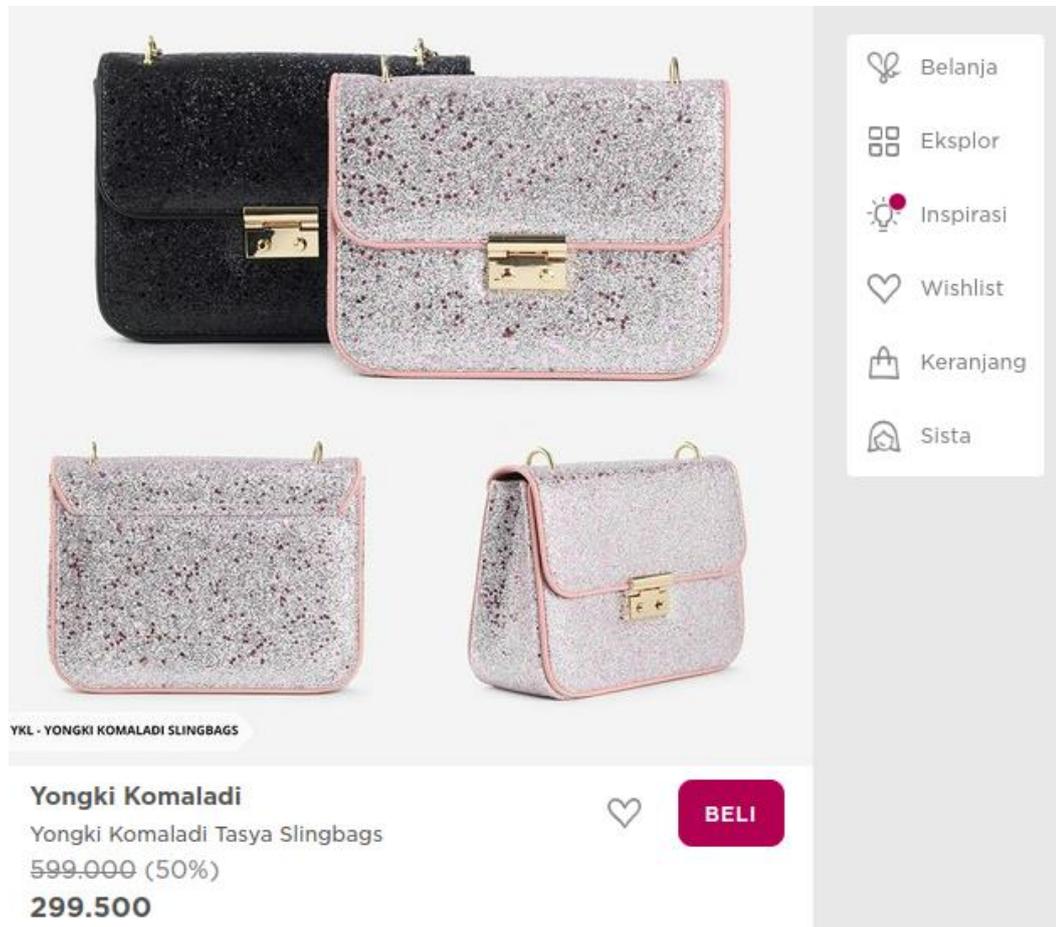
Dari antarmuka halaman pencarian pada website Uqishop tersebut kemudian didapatkan struktur HTML produknya yang dapat dilihat pada Gambar 3. 15.

```
<li class="product type-product post-4616 status-publish first instock product_cat-jual-produk-best-seller
product_cat-jual-tas-kulit product_tag-tas product_tag-tas-kulit product_tag-tas-selempang-kulit
product_tag-tas-selempang-wanita product_tag-tas-wanita has-post-thumbnail sale shipping-taxable purchasable
product-type-simple">
  <div class="product-inner card">
    <div class="product-image-box">
      <a href="https://uqishop.com/product/slingbag-women-sc01/" class="woocommerce-LoopProduct-link
woocommerce-loop-product__link">
        <span class="onsale">Sale!</span>
        
      </a>
    </div>
    <div class="product-detail-box card-body">
      <a href="https://uqishop.com/product/slingbag-women-sc01/" class="product-detail-link">
        <h2 class="woocommerce-loop-product__title">Tas Selempang Kulit Wanita SC01</h2>
        <span class="price"> ***
        </span>
      </a>
      <a href="https://uqishop.com/product/slingbag-women-sc01/" class="button button-shop-addtocart"> ***
      </a>
      <a rel="nofollow" href="https://uqishop.com/?tp_wa_rotator=1&tp_product_id=4616" class="button
button-shop-whatapp button-track-event" data-fbevent="Lead" data-fbtrack="track" data-fbparams="{&
quot;content_name&quot;:&quot;Tas Selempang Kulit Wanita SC01&quot;,&quot;value&quot;:&quot;549000.00
&quot;,&quot;currency&quot;:&quot;IDR&quot;,&quot;source&quot;:&quot;adoraid-wp&quot;,&quot;
source_action&quot;:&quot;button-click&quot;,&quot;source_position&quot;:&quot;single-product&quot;,&
quot;version&quot;:&quot;2.2.0&quot;,&quot;domain&quot;:&quot;uqishop.com/&quot;}"> ***
      </a>
    </div>
  </div>
</li>
```

Gambar 3. 15 Potongan Struktur HTML Pencarian Produk Halaman Uqishop

4. Sorabel

Berikut ini merupakan tampilan antarmuka pencarian produk pada halaman website Sorabel seperti yang dapat dilihat pada Gambar 3. 16.



Gambar 3. 16 Tampilan Antarmuka Pencarian Produk Halaman Sorabel

Dari antarmuka halaman pencarian pada website Sorabel tersebut kemudian didapatkan struktur HTML produknya yang dapat dilihat pada Gambar 3. 17.

```

<div data-qa-id="products_search?query=Tas_2_0_Product_product_score_desc">
  <div style="background-color: rgb(255, 255, 255); border-width: 0px;">
    <div style="border-width: 0px;">
      <div data-qa-id="row:layout:1:0" style="border-width: 0px;">
        <div style="margin-left: 0px; margin-right: 0px; box-sizing: border-box; display: flex; flex-flow
          : row wrap; align-items: flex-start;">
          <div style="padding-left: 0px; padding-right: 0px; width: 100%; vertical-align: middle;
            box-sizing: border-box;">
            <div data-qa-id="col:spanned:1" style="background-color: rgb(255, 255, 255); border-width
              : 0px;">
              <div data-qa-id="products_search?query=Tas_2_0_Product_product_score_desc" style="
                width: 100%; margin-bottom: 0px; background-color: white;">
                <a data-qa-id="qa_common_link" href="/products/yongki-komaladi-tasya-slingbags"
                  style="background-color: transparent; width: 100%; padding: 0px; border-width: 0
                    px; cursor: pointer; outline: none; text-decoration: none; vertical-align:
                      top;">
                  <div style="position: relative;">
                  </a>
                  <div data-qa-id="qa_product_info" style="padding: 8px 16px; height: 140px;
                    width: 100%; display: flex; align-items: flex-start; box-sizing:
                      border-box; flex-direction: row; justify-content: space-between;">
                    <div style="display: table-cell; vertical-align: top; white-space: nowrap
                      ; overflow: hidden; text-overflow: ellipsis;">
                    <a data-qa-id="qa_product_title_Yongki Komaladi Tasya Slingbags" href
                      ="/products/yongki-komaladi-tasya-slingbags" style="background-color:
                        transparent; padding: 0px; border-width: 0px; cursor: pointer;
                          outline: none; text-decoration: none; vertical-align: top;">
                    </a>
                    <p data-qa-id="qa_common_text" type="p" style="margin: 0px;
                      vertical-align: middle; font-family: GothamSSm-Book; color: rgb(
                        82, 82, 82); font-size: 14px; line-height: 21px; font-weight:

```

Gambar 3. 17 Potongan Struktur HTML Pencarian Produk Halaman Sorabel

3.2.5. Menentukan Data yang Akan Dilakukan Proses Ekstraksi

Data yang akan dilakukan ekstraksi dari ke-4 website tersebut kita melakukan ekstraksi data-data terkait yang penting seperti **nama produk**, **gambar produk**, **harga produk**, **asal produk**, **link produk** dan **source produk**, tahap ini merupakan lanjutan tahap lanjutan setelah menganalisis struktur DOM, kita melakukan penentuan bagian struktur yang akan kita ambil terkait data-data terkait yang penting tersebut.

1. Elevenia

Berikut merupakan data yang akan kita tentukan yang nanti akan dilakukan ekstraksi data pada website Elevenia:

a. Gambar Produk

Data gambar dapat dilihat pada sub produk yang mengandung tag `img`, dan dari tag tersebut terdapat properti `src` yang dapat kita lakukan ekstraksi. Seperti yang dapat dilihat pada gambar 3. 18.

```

```

Gambar 3. 18 Tag dan Data Gambar Produk Halaman Elevenia

b. Nama Produk

Data nama produk dapat dilihat pada sub produk yang mengandung informasi mengenai tag tentang nama produk yang kemudian dapat kita lakukan ekstraksi. Seperti yang dapat dilihat pada gambar 3. 19.

```
<a href="http://www.elevenia.co.id/prd-sophie-paris-tas-wanita-webelove-bag-t3775m2-28635549" class="pordLink notranslate">Sophie Paris Tas Wanita Webelove Bag-T3775M2</a>
```

Gambar 3. 19 Tag dan Data Nama Produk Halaman Elevenia

c. Harga Produk

Data harga produk dapat dilihat pada sub produk yang mengandung informasi mengenai tag tentang harga produk yang kemudian dapat kita lakukan ekstraksi. Seperti yang dapat dilihat pada gambar 3. 20.

```

<div class="price notranslate">
  <span>60<sup>%</sup></span>
  <div class="prc">
    <p><del>Rp 239.900</del></p>
    <strong>Rp 96.000</strong>
  </div>
</div>

```

Gambar 3. 20 Tag dan Data Harga Produk Halaman Elevenia

d. Asal Produk

Data asal produk dapat dilihat pada sub produk yang mengandung informasi mengenai tag tentang asal produk yang kemudian dapat kita lakukan ekstraksi. Seperti yang dapat dilihat pada gambar 3. 21.

```

<ul class="sellerPlace">
  <li>
    <em class="ico_prodType placeStore"></em>Jakarta Barat
  </li>
</ul>

```

Gambar 3. 21 Tag dan Data Asal Produk Halaman Elevenia

e. Link Produk

Data link produk dapat dilihat pada sub produk yang mengandung informasi mengenai tag link produk ke halaman *e-commerce*, yang kemudian dapat kita lakukan ekstraksi. Seperti yang dapat dilihat pada gambar 3. 22.

```

<a href="http://www.elevenia.co.id/
prd-sophie-paris-tas-wanita-webelove-bag-t3775m2-28635549" class="pordLink
notranslate">Sophie Paris Tas Wanita Webelove Bag-T3775M2</a>

```

Gambar 3. 22 Tag dan Data Link Produk Halaman Elevenia

2. OLX

Berikut merupakan data yang akan kita tentukan yang nanti akan dilakukan ekstraksi data pada website OLX:

a. Gambar Produk

Data gambar dapat dilihat pada sub produk yang mengandung tag `img`, dan dari tag tersebut terdapat properti `src` yang dapat kita lakukan ekstraksi. Seperti yang dapat dilihat pada gambar 3. 23.

```
<figure class="_2grx4" data-aut-id="ItemImage">

```

Gambar 3. 23 Tag dan Data Gambar Produk Halaman OLX

b. Nama Produk

Data nama produk dapat dilihat pada sub produk yang mengandung informasi mengenai tag tentang nama produk yang kemudian dapat kita lakukan ekstraksi. Seperti yang dapat dilihat pada gambar 3. 24.

```
<span class="_2tw1I" data-aut-id="itemTitle">Fashi Tas Selempang Kasual Crossbody
Bag Bahan Canvas-Khaki</span>
```

Gambar 3. 24 Tag dan Data Nama Produk Halaman OLX

c. Harga Produk

Data harga produk dapat dilihat pada sub produk yang mengandung informasi mengenai tag tentang harga produk yang kemudian dapat kita lakukan ekstraksi. Seperti yang dapat dilihat pada gambar 3. 25.

```
<span class="_89yzn" data-aut-id="itemPrice">Rp 80.000</span>
```

Gambar 3. 25 Tag dan Data Harga Produk Halaman OLX

d. Asal Produk

Data asal produk dapat dilihat pada sub produk yang mengandung informasi mengenai tag tentang asal produk yang kemudian dapat kita lakukan ekstraksi. Seperti yang dapat dilihat pada gambar 3. 26.

```
<span class="tjgMj" data-aut-id="item-location">Batununggal, Bandung Kota</span>
```

Gambar 3. 26 Tag dan Data Asal Produk Halaman OLX

e. Link Produk

Data link produk dapat dilihat pada sub produk yang mengandung informasi mengenai tag link produk ke halaman *e-commerce*, yang kemudian dapat kita lakukan ekstraksi. Seperti yang dapat dilihat pada gambar 3. 27.

```
<a href="/item/fashi-tas-selempang-kasual-crossbody-bag-bahan-canvas-khaki-iid-784673936">
```

Gambar 3. 27 Tag dan Data Link Produk Halaman OLX

3. Uqishop

Berikut merupakan data yang akan kita tentukan yang nanti akan dilakukan ekstraksi data pada website Uqishop:

a. Gambar Produk

Data gambar dapat dilihat pada sub produk yang mengandung tag `img`, dan dari tag tersebut terdapat properti `src` yang dapat kita lakukan ekstraksi. Seperti yang dapat dilihat pada gambar 3. 28.

```

```

Gambar 3. 28 Tag dan Data Gambar Produk Halaman Uqishop

b. Nama Produk

Data nama produk dapat dilihat pada sub produk yang mengandung informasi mengenai tag tentang nama produk yang kemudian dapat kita lakukan ekstraksi. Seperti yang dapat dilihat pada gambar 3. 29.

```
<h2 class="woocommerce-loop-product__title">Tas Selempang Kulit Wanita SC01</h2>
```

Gambar 3. 29 Tag dan Data Nama Produk Halaman Uqishop

c. *Harga Produk*

Data harga produk dapat dilihat pada sub produk yang mengandung informasi mengenai tag tentang harga produk yang kemudian dapat kita lakukan ekstraksi. Seperti yang dapat dilihat pada gambar 3. 30.

```
<span class="price">
  <del>
    <span class="woocommerce-Price-amount amount">
      <span class="woocommerce-Price-currencySymbol">Rp</span>699.000
    </span>
  </del>
  <ins>
    <span class="woocommerce-Price-amount amount">
      <span class="woocommerce-Price-currencySymbol">Rp</span>549.000
    </span>
  </ins>
</span>
```

Gambar 3. 30 Tag dan Data Harga Produk Halaman Uqishop

d. *Asal Produk*

Data asal produk dapat dilihat pada sub produk yang mengandung informasi mengenai tag tentang asal produk, namun pada halaman Uqishop di objek produkt tersebut tidak ada rincian mengenai asal produk, namun toko yang menangani penjualan berdasarkan deskripsi berlokasi di kota Malang, Jawa Timur, sehingga kita dapat melakukan

penambahan data secara otomatis untuk asal produk jika ada pencarian dan produk yang terdapat pada halaman website Uqishop.

e. Link Produk

Data link produk dapat dilihat pada sub produk yang mengandung informasi mengenai tag link produk ke halaman *e-commerce*, yang kemudian dapat kita lakukan ekstraksi. Seperti yang dapat dilihat pada gambar 3. 32.

```
<a href="https://uqishop.com/product/slingbag-women-sc01/" class="
woocommerce-LoopProduct-link woocommerce-loop-product__link">
```

Gambar 3. 31 Tag dan Data Link Produk Halaman Uqishop

4. Sorabel

Berikut merupakan data yang akan kita tentukan yang nanti akan dilakukan ekstraksi data pada website Sorabel:

a. Gambar Produk

Data gambar dapat dilihat pada sub produk yang mengandung tag *img*, dan dari tag tersebut terdapat properti *src* yang dapat kita lakukan ekstraksi. Seperti yang dapat dilihat pada gambar 3. 32.

```

```

Gambar 3. 32 Tag dan Data Gambar Produk Halaman Sorabel

b. Nama Produk

Data nama produk dapat dilihat pada sub produk yang mengandung informasi mengenai tag tentang nama produk yang kemudian dapat kita lakukan ekstraksi. Seperti yang dapat dilihat pada gambar 3. 33.

```
<span data-qa-id="qa_common_text" type="span" style="margin: 0px;
  vertical-align: middle; font-family: GothamSSm-Book; color:
  rgb(82, 82, 82); font-size: 12px; line-height: 18px;
  font-weight: normal; max-width: 480px; text-overflow:
  ellipsis; white-space: nowrap; overflow: hidden;">Yongki
  Komaladi Tasya Slingbags</span>
```

Gambar 3. 33 Tag dan Data Nama Produk Halaman Sorabel

c. *Harga Produk*

Data harga produk dapat dilihat pada sub produk yang mengandung informasi mengenai tag tentang harga produk yang kemudian dapat kita lakukan ekstraksi. Seperti yang dapat dilihat pada gambar 3. 34.

```
<a data-qa-id="qa_common_link" href="/products/
  yongki-komaladi-tasya-slingbags" style="background-color: transparent
  ; padding: 0px; border-width: 0px; cursor: pointer; outline: none
  ; text-decoration: none; vertical-align: top;">
  <div data-qa-id="
  qa_product_price_08e18c50-27e8-40d0-a797-7ffef495c18f" type="div"
  style="margin: 0px; vertical-align: middle; display: block; font
  -family: GothamSSm-Bold; color: rgb(82, 82, 82); font-size:
  16px; line-height: 24px; font-weight: normal; white-space:
  normal;">299.500
```

Gambar 3. 34 Tag dan Data Harga Produk Halaman Sorabel

d. *Asal Produk*

Data asal produk dapat dilihat pada sub produk yang mengandung informasi mengenai tag tentang asal produk, namun pada halaman Sorabel di objek produkt tersebut tidak ada rincian mengenai asal produk, namun toko yang menangani penjualan berdasarkan deskripsi berlokasi default di Indonesia sehingga kita dapat melakukan penambahan data

secara otomatis untuk asal produk jika ada pencarian dan produk yang terdapat pada halaman website Sorabel.

e. *Link Produk*

Data link produk dapat dilihat pada sub produk yang mengandung informasi mengenai tag link produk ke halaman *e-commerce*, yang kemudian dapat kita lakukan ekstraksi. Seperti yang dapat dilihat pada gambar 3. 35.

```
<a data-qa-id="qa_common_link" href="/products/yongki-komaladi-tasya-slingbags" style="background-color: transparent; width: 100%; padding: 0px; border-width: 0px; cursor: pointer; outline: none; text-decoration: none; vertical-align: top;">
```

Gambar 3. 35 Tag dan Data Link Produk Halaman Sorabel

3.2.6. Menentukan Pola dan Ekspresi XPath

Dari data yang akan telah kita tentukan diatas kita mendapatkan informasi mengenai tag serta data yang kita inginkan, langkah selanjutnya yaitu menemukan pola dan ekspresi XPath yang cocok agar kita mendapatkan data dan informasi yang kita inginkan, berikut ini merupakan pola dan ekspresi XPath pada tiap tiap situs *e-commerce*.

1. Elevenia

Berikut ini merupakan tabel pola dan ekspresi XPath berdasarkan data yang akan kita ambil pada halaman Elevenia, pola dan ekspresi Xpath halaman Elevenia dapat dilihat seperti pada tabel 3.2:

Tabel 3. 2 Pola dan Ekspresi XPath Halaman Elevenia

No	Data Produk	Ekspresi XPath
1	Gambar	//div[@class="group"]/a/img/@src

No	Data Produk	Ekspresi XPath
2	Nama	//a[@class="pordLink notranslate"]/text()
3	Harga	//div[@class="prc"]/strong/text()
4	Asal	//ul[@class="sellerPlace"]/li/text()
5	Link	//a[@class="pordLink notranslate"]/@href

2. OLX

Berikut ini merupakan tabel pola dan ekspresi XPath berdasarkan data yang akan kita ambil pada halaman OLX, pola dan ekspresi Xpath halaman OLX dapat dilihat seperti pada tabel 3.3:

Tabel 3. 3 Pola dan Ekspresi XPath Halaman OLX

No	Data	Ekspresi XPath
1	Gambar	//figure[@class="_2grx4"]//img//@src
2	Nama	//span[@class="_2tW1I"]/text()
3	Harga	//span[@class="_89yzn"]/text()
4	Asal	//div[@class="_1KOFM"]/span[@class="tjgMj"]/text()
5	Link	//span[@class="_2tW1I"]/ancestor::a/@href

3. Uqishop

Berikut ini merupakan tabel pola dan ekspresi XPath berdasarkan data yang akan kita ambil pada halaman Uqishop, pola dan ekspresi Xpath halaman Uqishop dapat dilihat seperti pada tabel 3.4:

Tabel 3. 4 Pola dan Ekspresi XPath Halaman Uqishop

No	Data	Ekspresi XPath
1	Gambar	//div[@class="product-image-box"]/a/img/@src
2	Nama	//h2[@class="woocommerce-loop-product_title"]/text()

No	Data	Ekspresi XPath
3	Harga	//span[@class="price"]/ins/span/text()
4	Asal	-
5	Link	//a[@class="product-detail-link"]/@href

4. Sorabel

Berikut ini merupakan tabel pola dan ekspresi XPath berdasarkan data yang akan kita ambil pada halaman Elevenia:

Tabel 3. 5 Pola dan Ekspresi XPath Halaman Sorabel

No	Data Produk	Ekspresi XPath
1	Gambar	//div[@data-qa-id="qa_common_card"]/a/div/div/img/@src
2	Nama	//div[@data-qa-id="qa_common_card"]/div[@data-qa-id="qa_product_info"]/div/a/span/text()
3	Harga	//div[@data-qa-id="qa_common_card"]/div[@data-qa-id="qa_product_info"]/div/a[2]/div/text()
4	Asal	-
5	Link	//div[@data-qa-id="qa_common_card"]/a/@href

3.2.7. Ekstraksi Data

Setelah mendapatkan pola dan ekspresi XPath, kita dapat melakukan request ke masing-masing website dan melakukan ekstraksi data sesuai dengan *keyword* yang kita inginkan, proses ekstraksi dilakukan secara otomatis oleh sistem, dengan mendeteksi apakah ada *GET* request dari sistem.

Pada saat melakukan ekstraksi data, *traffic* dan koneksi internet juga cukup berpengaruh dalam kecepatan melakukan ekstraksi data. Sistem yang dibangun menggunakan Framework Django, library lxml, requests untuk menangani XPath, sedangkan Framework Django saat ini belum mendukung untuk operasi multiple

request secara bersamaan (Asynchronous) sehingga proses ekstraksi dilakukan satu persatu (Synchronous) setelah setiap request dilakukan.

Data yang telah diekstrak tadi jika terdapat jenis data atau tipe data yang tidak sesuai selama proses ekstraksi maka kita lakukan proses cleaning data terlebih dahulu agar data antar situs belanja dapat kita presentasikan dalam bentuk yang konsisten.

3.2.8. Cleaning Data

Pada tahap ini kita melakukan analisa apakah data sudah dalam bentuk dan terformat seperti yang kita inginkan, salah satu contohnya yaitu pada saat melakukan ekstraksi data terutama data dengan format yang berbeda misalkan harga, terkadang pada saat ekstraksi terdapat data harga dengan format yang berbeda-beda seperti *Rp. 10.000,00-* , *10.0000*, *10000*, *Rp 10.000, 00*, *10.000,00*. maka perlu adanya cleaning data agar semua data yang kita peroleh dari masing-masing situs belanja tersebut terformat dengan baik dan benar, sehingga memudahkan kita dalam menampilkan data maupun dalam proses sorting kedepan.

3.2.9. Simpan Data

Jika data telah selesai dilakukan proses cleaning, langkah selanjutnya yaitu kita simpan data sementara tersebut, dalam hal ini data yang kita simpan secara bertahap karena request dilakukan secara bergantian (Synchronous),

3.2.10. Penggabungan dan Rekonstruksi Data

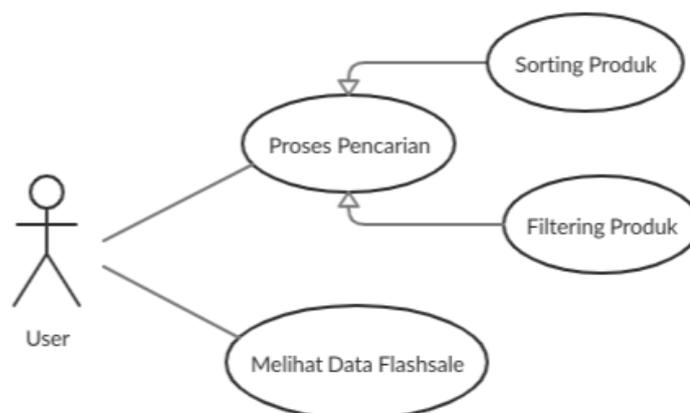
Jika data sudah disimpan, maka data tersebut akan kita rekonstruksi ulang susunan urutan datanya dengan menggunakan sorting sequence matcher atau sorting berdasarkan relevansi agar data yang kita gabung tidak berurutan sesuai urutan produk pada situs *e-commerce*, melainkan kita sorting ulang sehingga data tersebut kita sesuaikan dengan produk yang kita cari.

3.2.11. Output, Display Data

Data yang sudah digabungkan dan direkonstruksi dapat kita tampilkan ke dalam sistem yang akan kita bangun. Jika terdapat hasil pencarian maka akan menampilkan data berupa hasil pencarian namun jika tidak terdapat hasil pencarian atau pencarian tidak ditemukan maka sistem akan memberikan informasi produk tidak ditemukan.

3.2.12. Use Case Diagram

Permodelan *Use Case Diagram* dapat dilihat pada gambar 3. 36.



Gambar 3. 36 Use Case Diagram Sistem Pencarian

Penjelasan dari *Use Case Diagram* adalah sebagai berikut:

a. User

User berperan sebagai pengguna sistem pada Sistem Implementasi Webscraping untuk Pencarian Produk dengan menggunakan Metode XPath pada Halaman Website *e-commerce*, User dapat melakukan pencarian produk, melakukan sorting, filter dari pencarian produk, melihat flashsale atau diskon dari situs belanja. Yang berperan sebagai user adalah Pengguna maupun Konsumen yang sedang ingin melakukan kegiatan belanja online.

b. Use Case

1. Proses Pencarian

Proses melakukan pencarian produk dengan menginputkan *keyword* pencarian yang diinginkan, dan memberikan informasi hasil pencarian.

2. Sorting Produk

Proses melakukan sorting hasil pencarian produk berdasarkan relevansi, harga tertinggi atau harga terendah dari pencarian produk.

3. Filtering Produk

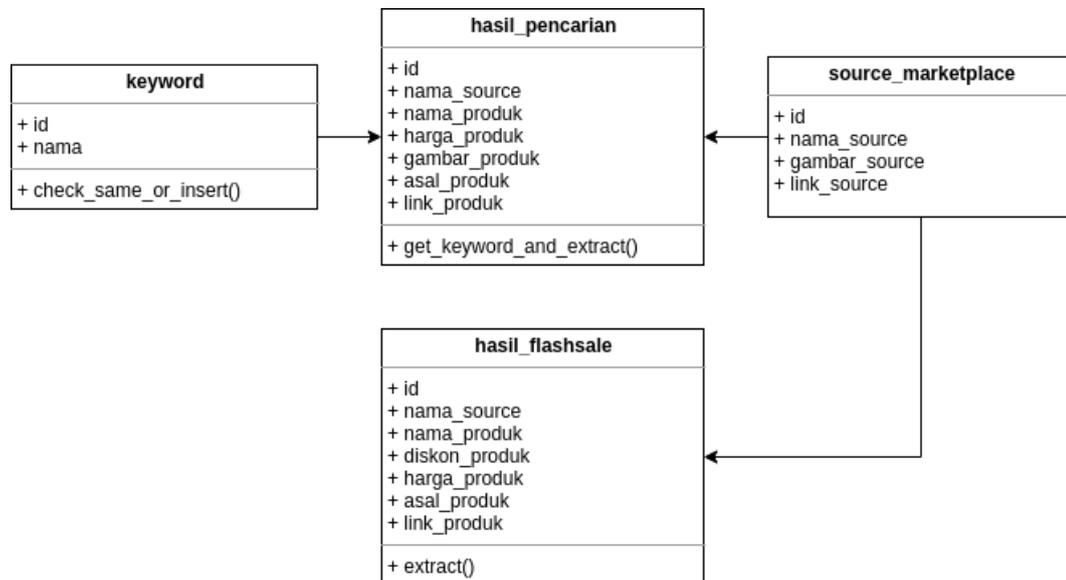
Proses melakukan filter hasil pencarian produk berdasarkan source marketplace.

4. Melihat Data Flashsale

Proses melihat data produk dari kategori flashsale, promo, diskon atau daily deals berdasarkan data yang telah diekstrak dari situs *e-commerce*.

3.2.13. Class Diagram

Permodelan *Class Diagram* dapat dilihat pada gambar 3. 37.



Gambar 3. 37 Class Diagram Sistem Pencarian

3.2.14. Activity Diagram

1. Activity Diagram Proses Pencarian

Activity diagram proses pencarian merupakan aktivitas untuk melakukan pencarian produk. Untuk dapat melakukan pencarian maka *actor* perlu mengisi keyword pencarian pada halaman beranda atau pada halaman pencarian. Activity diagram proses pencarian dapat dilihat pada tabel 3. 6.

Tabel 3. 6 Activity Diagram Proses Pencarian

Usecase	
Proses Pencarian	
Aktor Yang Terlibat Dalam Sistem	
User	
Keadaan Sebelum Proses	
User belum melakukan pencarian atau sedang berada di halaman pencarian	
Flow of event	
Main Event	<ol style="list-style-type: none"> 1. Mulai 2. Halaman Beranda 3. Masukkan keyword ke kolom pencarian 4. Muncul hasil pencarian 5. Selesai
Activity Diagram	<pre> graph TD Start(()) --> A(Halaman Beranda) A --> B(Masukkan keyword pencarian) B --> C(Muncul hasil pencarian) C --> End((())) </pre>
Keadaan Sesudah Proses	
Tampil hasil pencarian	

2. Activity Diagram Proses Sorting Produk

Activity diagram proses sorting merupakan aktivitas untuk melakukan sorting produk. Untuk dapat melakukan sorting User telah melakukan pencarian produk terlebih dahulu, dan kemudian user dapat memilih untuk melakukan sorting berdasarkan relevansi(default), harga tertinggi atau harga terendah. Activity diagram proses pencarian dapat dilihat pada tabel 3. 7.

Tabel 3. 7 Activity Diagram Proses Sorting Produk

Usecase	
Proses Sorting Produk	
Aktor Yang Terlibat Dalam Sistem	
User	
Keadaan Sebelum Proses	
User telah melakukan pencarian atau sedang berada dihalaman pencarian	
Flow of event	
Main Event	<ol style="list-style-type: none"> 1. Mulai 2. Halaman Pencarian 3. Pilih menu dropdown untuk melakukan sorting, pilihan yang disediakan ada 3 yaitu relevansi, harga tertinggi dan harga terendah 4. Terapkan 5. Selesai
Activity Diagram	<pre> graph TD Start(()) --> A(Halaman Pencarian) A --> B(Pilih Menu Sorting) B --> C1(Relevance) B --> C2(High Price) B --> C3(Low Price) C1 --> D(Terapkan) C2 --> D C3 --> D D --> End((())) </pre>
Keadaan Sesudah Proses	
Tampil hasil pencarian sesuai dengan sorting yang telah dipilih	

3. Activity Diagram Proses Filtering Produk

Activity diagram proses filtering merupakan aktivitas untuk melakukan filter produk. Untuk dapat melakukan filtering User telah melakukan pencarian produk terlebih dahulu, dan kemudian user dapat memilih untuk melakukan filter berdasarkan marketplace yang terselekt. Activity diagram proses pencarian dapat dilihat pada tabel 3. 8.

Tabel 3. 8 Activity Diagram Proses Filtering

Usecase	
Proses Filtering	
Aktor Yang Terlibat Dalam Sistem	
User	
Keadaan Sebelum Proses	
User belum melakukan pencarian atau sedang berada dihalaman pencarian	
Flow of event	
Main Event	<ol style="list-style-type: none"> 1. Mulai 2. Halaman Pencarian 3. Pilih source marketplace yang ingin di filter 4. Muncul hasil pencarian 5. Selesai
Activity Diagram	<pre> graph TD Start(()) --> A(Halaman Pencarian) A --> B(Pilih Filter) B --> C(Selected Marketplace) C --> D(Terapkan) D --> End((())) </pre>
Keadaan Sesudah Proses	
Tampil hasil pencarian sesuai dengan filter yang dipilih	

4. Activity Diagram Melihat Data Flashsale

Activity diagram melihat data flashsale merupakan aktivitas untuk melihat informasi produk dengan promo, diskon, dailydeals atau penawaran lainnya dari situs belanja. Activity diagram proses pencarian dapat dilihat pada tabel 3. 8.

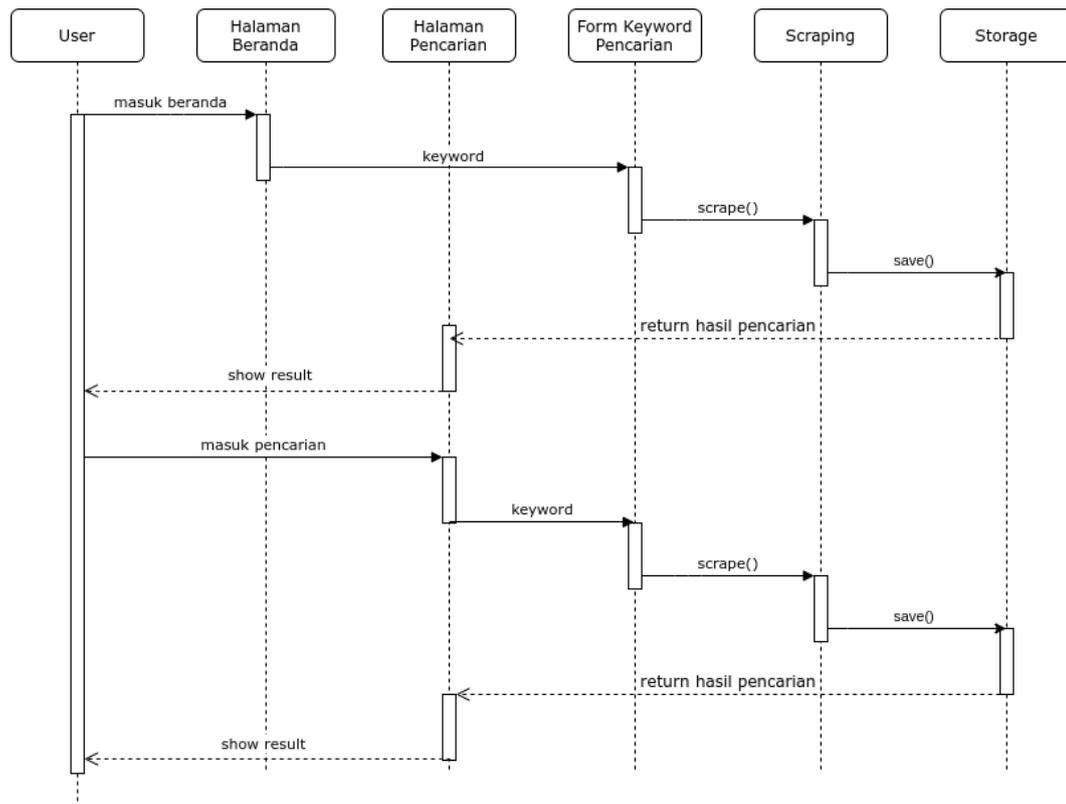
Usecase	
Proses Melihat Data Flashsale	
Aktor Yang Terlibat Dalam Sistem	
User	
Keadaan Sebelum Proses	
User tidak berada pada halaman flashsale	
Flow of event	
Main Event	<ol style="list-style-type: none"> 1. Mulai 2. Klik menu flashsale 3. Muncul hasil produk flashsale 4. Selesai
Activity Diagram	<pre> graph TD Start(()) --> A(Halaman Beranda atau Halaman Pencarian) A --> B(Pilih Menu Halaman Flashsale) B --> C(Muncul Hasil Flashsale) C --> End((())) </pre>
Keadaan Sesudah Proses	
Tampil hasil produk flashsale	

3.2.15. Sequence Diagram

Pemodelan *sequence diagram* berdasarkan *activity diagram* dapat ditampilkan pada tabel berikut ini:

1. Sequence Diagram Proses Pencarian

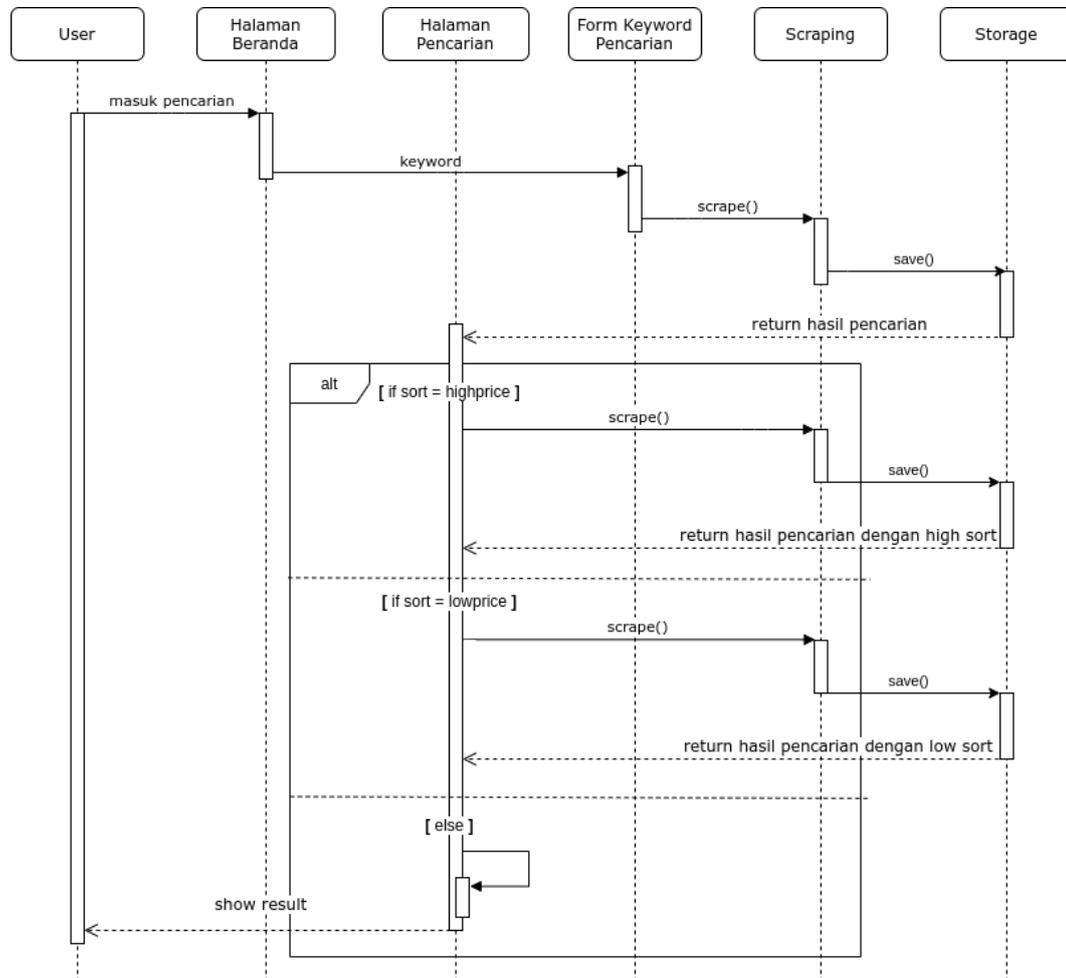
Berikut ini merupakan *sequence diagram* proses pencarian yang dapat dilihat pada gambar 3.38:



Gambar 3. 38 Sequence Diagram Proses Pencarian

2. Sequence Diagram Proses Sorting Produk

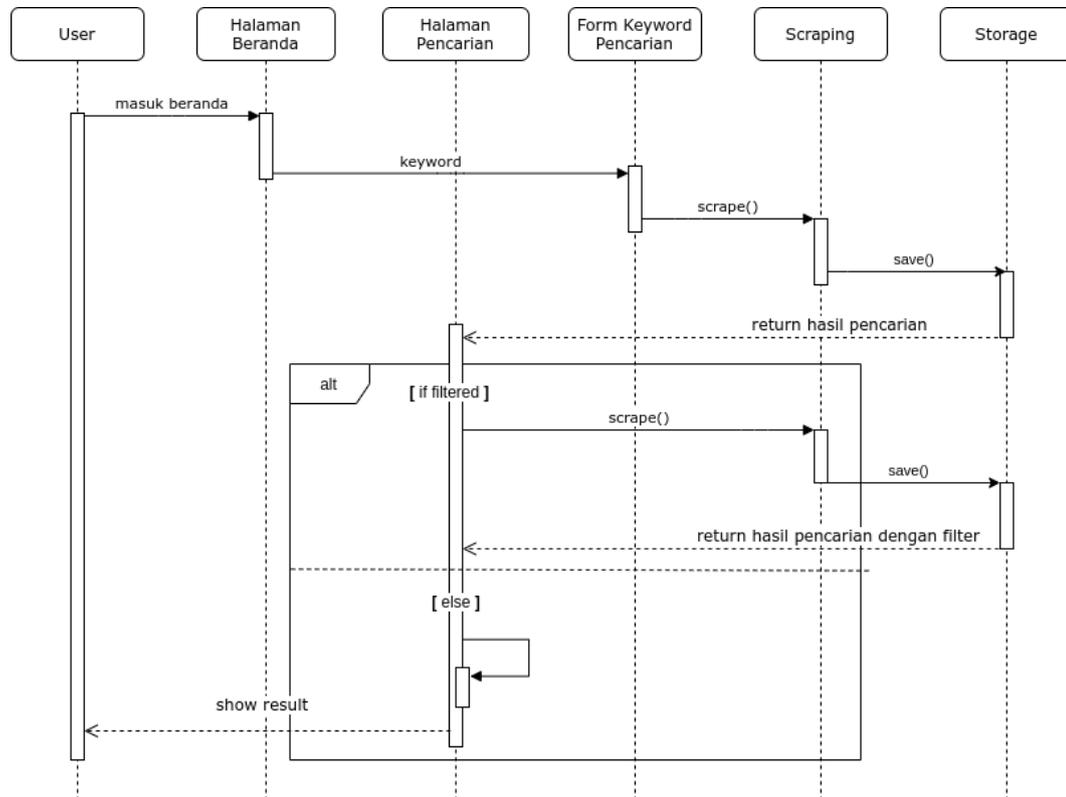
Berikut ini merupakan sequence diagram proses sorting produk yang dapat dilihat pada gambar 3.39:



Gambar 3. 39 Sequence Diagram Proses Sorting Produk

3. Sequence Diagram Proses Filtering Produk

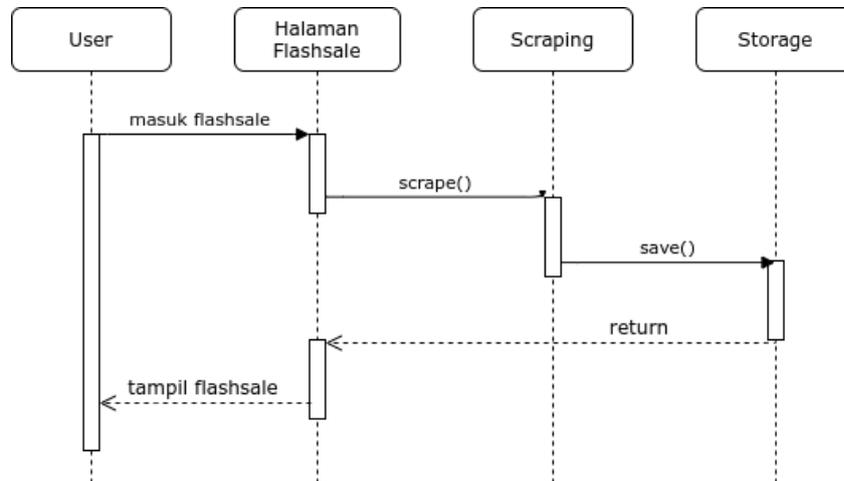
Berikut ini merupakan sequence diagram filtering produk yang dapat dilihat pada gambar 3.40:



Gambar 3. 40 Sequence Diagram Proses Filtering

4. Sequence Diagram Melihat Data Flashsale

Berikut ini merupakan sequence diagram melihat data flashsale yang dapat dilihat pada gambar 3.41:

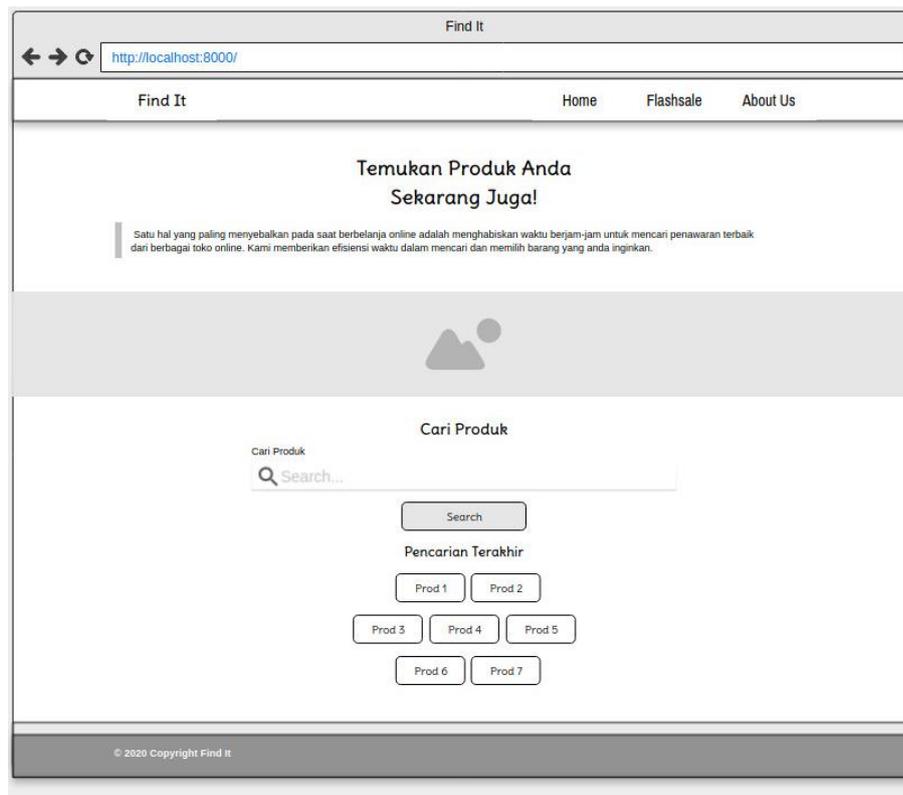


Gambar 3. 41 Sequence Diagram Melihat Data Flashsale

3.2.16. Perancangan Antarmuka Sistem (*Interface System*)

3.2.16.1. Rancangan Desain Halaman Home

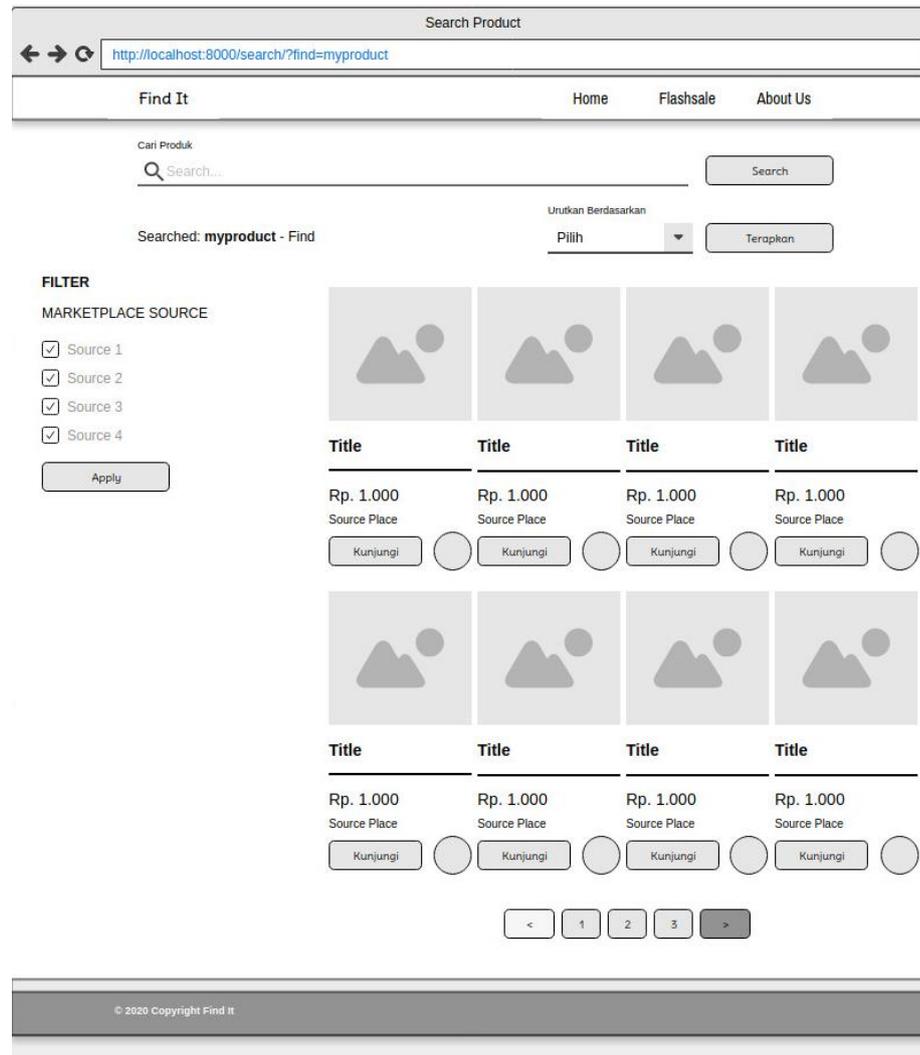
Halaman *Home* merupakan halaman utama sekaligus halaman yang pertama kali muncul pada saat sistem diakses. Perancangan untuk halaman home dapat dilihat pada gambar 3. 38.



Gambar 3. 42 Rancangan Desain Halaman Beranda (*Home*)

3.2.16.2. Rancangan Desain Halaman Search

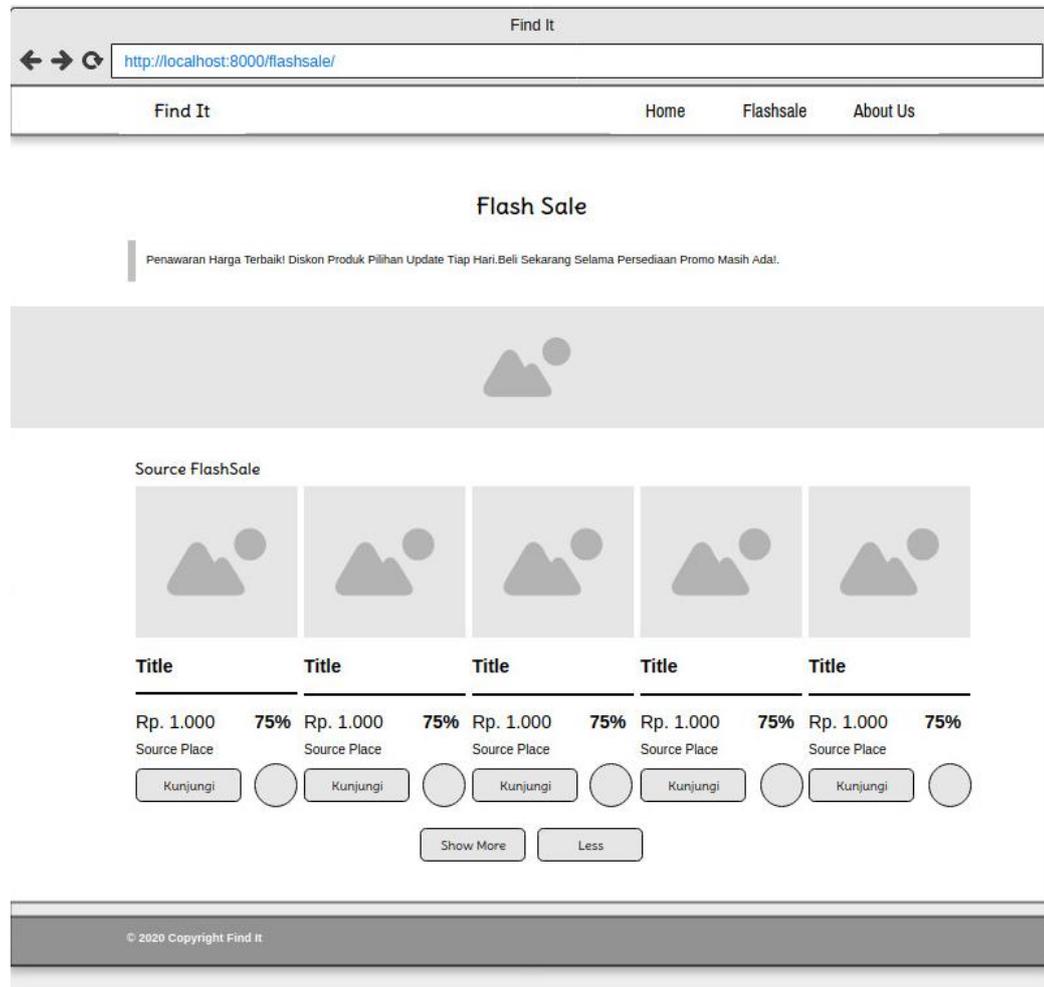
Halaman Search merupakan halaman yang menampilkan hasil pencarian ketika user meninputkan keyword pencarian. Perancangan untuk halaman search dapat dilihat pada gambar 3. 39.



Gambar 3. 43 Rancangan Desain Halaman *Search*

3.2.16.3. Rancangan Desain Halaman Flashsale

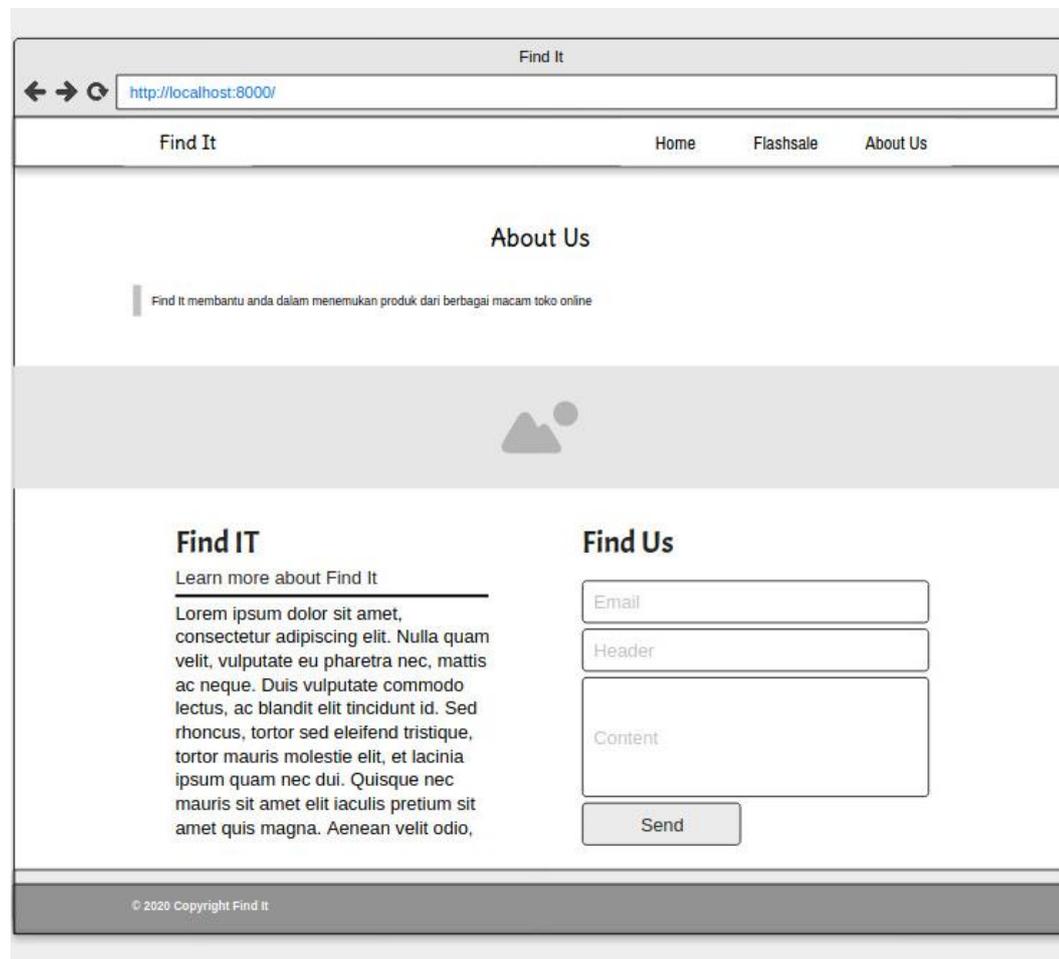
Halaman *Flashsale* merupakan halaman yang menyediakan beragam promo, diskon atau *flashsale* dari berbagai macam toko online. Perancangan untuk halaman *flashsale* dapat dilihat pada gambar 3. 40.



Gambar 3. 44 Rancangan Desain Halaman *Flashsale*

3.2.16.4. Rancangan Desain Halaman About Us

Halaman *About* merupakan halaman yang menyediakan informasi seputar website. Perancangan untuk halaman *about* dapat dilihat pada gambar 3. 41.



Gambar 3. 45 Rancangan Desain Halaman About Us