

Etalase Online Pedagang Produk UMKM dan Fashion di Wilayah Kota Malang dalam Penerapan UMKM Berbasis Teknologi Menggunakan Information Retrieval

Dwija Wisnu Brata^{1*}, Hariz Farisi²

^{1,2}Universitas Brawijaya

¹wisnubrata@ub.ac.id, ²hariz_farisi.ub.ac.id

ABSTRAK. Aktivitas belanja online yang dilakukan oleh pembeli senantiasa memiliki berbagai respon, misalnya saja pembeli berharap ketika berbelanja bisa mendapatkan diskon, lalu mendapatkan juga gratis ongkos kirim dan berharap sesuai terhadap barang yang datang dengan pesanan yang dilakukan. Kebiasaan pembeli juga biasanya senantiasa membanding-bandingkan produk yang dicari dan dicari mana yang dianggap sesuai (baik harga ataupun ongkos kirimnya). Hal ini mengakibatkan waktu yang dibutuhkan oleh pembeli pada saat membandingkan produk yang diinginkan dari satu situs jual beli online kesitus jual beli online lainnya mungkin cukup memakan waktu, oleh karena itu dibutuhkan sebuah program yang dapat memberikan informasi mengenai perbandingan data produk dari beberapa situs jual beli online sekaligus, apalagi ketika mencari produk yang terdekat (diwilayah kota Malang) yang memiliki produk yang sama seperti yang dijual diwilayah lain dan ongkos kirimnya lebih murah. Maka akan jauh lebih membantu dalam menyediakan informasi pelapak asli malang yang memasarkan produk-produknya dalam media online. Pada penelitian yang dilakukan, dirancang sebuah aplikasi yang dapat menyatukan banyak pelapak-pelapak kota Malang yang terdapat di Marketplace, sehingga warga kota Malang ketika akan berbelanja pakaian ataupun produk umkm akan tahu dan lebih mudah dalam mencari informasi pelapak fashion di wilayah kota Malang. Aplikasi Etalase ini juga dilakukan pengujian recall dan precision untuk tingkat keberhasilan mendapatkan nilai perolehan rata-rata (recall) sebesar 89% dan rata-rata tingkat ketepatan (precision) yang dihasilkan sebesar 90%. Hal ini dapat disimpulkan bahwa implementasi aplikasi dan metode yang digunakan cukup baik dalam menampilkan dan kesesuaian pencarian, baik berdasarkan kata, atau kategori.

Kata Kunci: Etalase Online, Umkm Malang, Information Retrieval

ABSTRACT. Online shopping activities carried out by buyers always have various responses. For example, buyers hope that when shopping, they can get discounts, then they also get free shipping and hope that the goods that come with their orders will match. Buyers also usually always compare the products they are looking for and which ones are considered appropriate (both price and shipping costs). This results in the time required by buyers when comparing the products they want from one online buying and selling site to another online buying and selling site may be quite time-consuming. Therefore a program is needed that can provide information on product data comparisons from several online buying and selling sites. At the same time, especially when looking for the closest product (in the Malang city area), which has the same product as those sold in other areas and cheaper shipping costs. Then it will be much more helpful in providing information on the original poor sellers who market their products in online media. In the research conducted, an application was designed that can unite many Malang city sellers in the Marketplace so that Malang city residents, when going to shop for clothes or MSME products, will know and find it easier to find information on fashion sellers in the Malang city area. This storefront application is also tested for recall and precision for the success rate of getting an average recall value of 89% and an average precision level of 90%. It can be concluded that the implementation of the application and the methods used are quite good in displaying and suitability of searches, either by word or category.

Keywords: Etalase Online, Umkm Malang, Information Retrieval

PENDAHULUAN

Kreatifitas generasi muda di kota Malang sudah semakin meningkat, terlihat semakin meningkatnya komunitas milenial yang bermacam-macam dan para pejuang muda (pengusaha muda) dari tahun ke tahun. Jika dikaitkan dengan teknologi, pengaruh perangkat dan kemampuan senantiasa berkembang mengikuti industry 4.0, sehingga generasi milenial dengan teknologi tersebut memiliki daya jangkauan yang luas, serta peningkatan kreatifitas. Kehandalan teknologi digunakan tidak hanya sebagai komunikasi, chatting, ataupun update status, tapi sekarang penggunaannya jauh lebih bermanfaat dan menguntungkan. Misalnya saja, di Malang banyak sekali distro pakaian yang dimiliki oleh anak muda. Produk distro yang dimiliki juga dipasarkan dimedia online ataupun media sosial. Di tahun 2019 hingga sekarang sudah sangat marak usaha-usaha kemilenialan yang mana konsumen terbanyak adalah anak muda, contoh saja kopi. Contoh lain ukm yang juga bergerak dalam bidang fashion wanita (seperti distro) yaitu penjualan gamis syari yang sangat populer dikalangan anak muda (milenial) ataupun ibu-ibu

rumah tangga. Para pegiat usaha yang dicontohkan memasarkan tidak hanya pada toko yang dimiliki tetapi lebih luas dengan memanfaatkan internet dan marketplace.

Permasalahan akan timbul ketika seorang pembeli yang ingin membeli produk (baik itu fashion ataupun jenis produk ukm lainnya), ketika produknya sudah dikirim dan diterima terkadang terjadi ketidaksesuaian. Kejadian tersebut menimbulkan kerugian kedua sisi, jika dari sisi pembeli akan mengalami kerugian dari barang yang diterima, apalagi ketika pembeli ingin mengembalikan ke penjual dimana harus mengirimkan produk yang salah dengan ongkos pribadi. Dari sisi penjual, kerugian akan dialami ketika barang akan dikirim lagi. Hal ini perlu adanya solusi dalam menyelesaikan permasalahan dalam ongkos kirim, kedekatan pembeli asli malang dengan penjual di wilayah malang dimana kesesuaian produk yang dicari, hal ini juga berdampak terhadap meminimalisir ongkos kirim yang didapat ataupun gratis

Pada penelitian yang dilakukan, dirancang sebuah aplikasi yang dapat menyatukan banyak pelapak-pelapak kota Malang yang terdapat di Marketplace, sehingga warga kota Malang ketika akan berbelanja pakaian akan tahu dan lebih mudah dalam mencari informasi pelapak fashion di wilayah kota Malang.

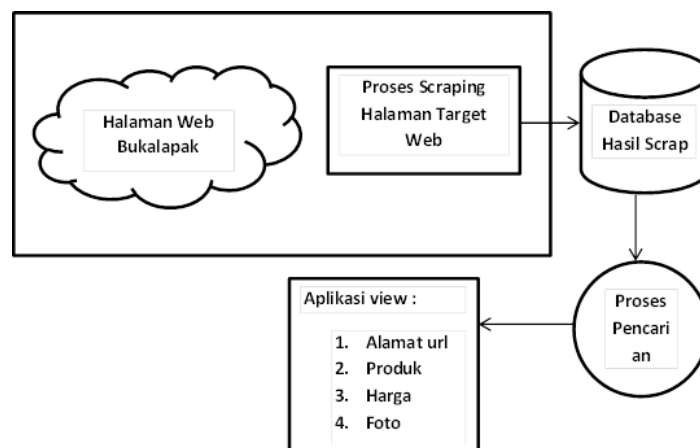
Metode scraping yang dilakukan pada html dom oleh [1] dengan tujuan untuk mendapatkan data korpus dua bahasa secara otomatis dan menambahkan ke database yang sudah ada, sehingga terjadi varian data dari dua bahasa yang pernah dikumpulkan dengan baru dikumpulkan.

Teknik scraping pada penelitian [2] dilakukan untuk mengekstrak data dari sebuah web toko online dimana toko tersebut menjual produk-produk jenis dan segala macam model smartphone. Adapun hasil akhir ekstrak data yang diperoleh yaitu merk, harga, dan url web. Data tersebut akan digunakan dan ditampilkan pada halaman web yang dirancang, lalu pengunjung dapat melihat hasilnya, dan dapat membandingkan produk smartphone yang dimiliki pada beberapa toko online.

Penelitian yang dilakukan oleh [3] tentang penggalian data, juga teknik web scrapping pada penelitian, dan juga pendekatan web scrapping pada studi kasus biomedik. Informasi data yang menjadi rujukan scrapping data berasal dari website National Center for Biotechnology Information dan European Bioinformatics Institute. Hasil yang diperoleh adalah pembentukan data pada sebuah database terkait data-data biomedis, dan juga data menjadi referensi data yang dapat diolah kapanpun dan dengan metode apapun.

METODE PENELITIAN

Metodologi penelitian yang dilakukan menggunakan beberapa konsep perancangan yaitu perancangan blok sistem, rancangan penyimpanan url, dan proses scrapping. Blok sistem yang dirancang pada penelitian yang dilakukan menggambarkan secara garis besar alur dari sistem pengambilan data hingga sistem penyajian data sesuai dengan inputan dari user yang ditunjukkan pada Gambar 1.

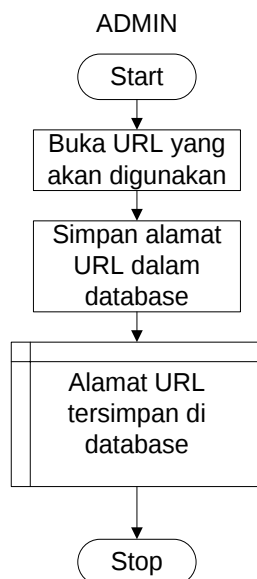


Gambar 1. Blok Sistem

Keterangan pada Gambar 1:

1. Pada bagian dari admin: admin bertugas untuk pengambilan data melalui proses scraping pada halaman situs dan kemudian hasil dari datanya disimpan kedalam database hasil scraping.
2. Pada bagian dari user: user dapat menginputkan merk dan type handphone sesuai keinginan kemudian sistem akan mencari kedalam database hasil scraping admin tadi. Setelah ditemukan maka data akan di tampilkan pada user sesuai dengan keyword yang diinputkan.

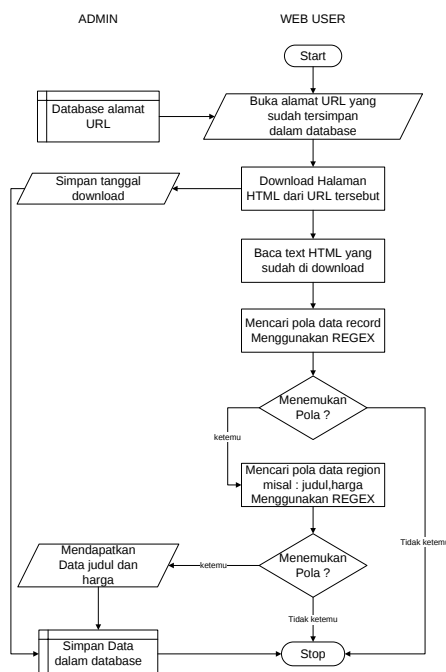
Pada sistem ini dirancang sebuah alur menggunakan flowchart dalam menyimpan url yang akan digunakan dalam menarik data umkm, desain flowchart admin bisa dilihat pada Gambar 2.



Gambar 2. Desain Flowchart Admin

Pada proses sumber data yang akan dicari, administrator/admin menelusuri data-data berkaitan dengan umkm pelapak kota Malang yang memiliki lapak online di e-commerce yang terkenal seperti bukalapak dan lain-lain. Url yang mengarah kepada halaman pelapak, maka url tersebut akan disimpan oleh sistem ke database.

Tahap berikutnya dalam mengoleksi data yaitu scrapping [4]. Hal ini terlihat pada rancangan alur proses scrapping pada Gambar 3:



Gambar 3. Desain Scrapping Data

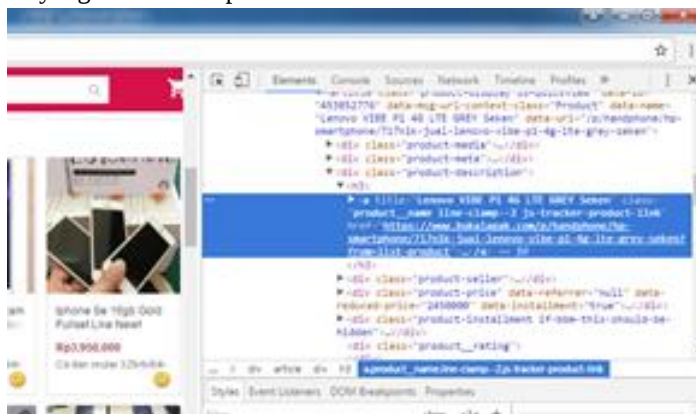
Pada proses scraping data, url yang sudah tersimpan pada database akan di-load dan dibaca format html yang terdapat pada halaman tersebut. Halaman web tersebut berisi data dan pola data recordnya. Proses analisis pola pada halaman tersebut digunakan Regular Expression. Pola yang sudah diketahui akan disimpan kedalam database.

HASIL DAN PEMBAHASAN

Pada hasil yang didapat pada penelitian, dibagi menjadi beberapa pembahasan, yaitu:

1. Analisis Struktur Data Web

Mengawali pencarian pola[5] dengan meng-inspect hasil url yang sudah ditentukan dan ditampilkan, hal dimaksudkan agar mengetahui struktur atau pola html pada web sehingga diperlukan untuk membaca halaman html dari url yang sudah disimpan.



Gambar 4. Inspect Web

2. Pencarian Pola Struktur HTML

Setelah membaca halaman html dari proses *inspect*[6] halaman web maka ditemukan struktur html [7][8] yang memuat keseluruhan data yang akan diambil. Berikut ini adalah struktur yang diambil pada halaman web bukalapak:

Tabel 1. Pola Data

No	Body Data	Record Data
1	a : <div class="basic-products basic-products—grid	a : <li class="product—sem col-12--2">
	b : <section class="endorsed-products	b :

Tabel 2. Pola Isi Struktur Data Record

No	Pola Isi
1	Nama : data-name="nama barang"
	Harga : harga barang
	Alamat url : data-url="alamat url"

3. HTML Menjadi Pola Regex

Struktur yang sudah diperoleh (baik dari body data sampai pola isi struktur data record) tidak langsung bisa digunakan untuk pengambilan data, harus dirubah terlebih dahulu kedalam pola *regular expression* [9][10][2]. Berikut adalah perubahan struktur sebelum dan sesudah dirubah kedalam pola *regular expression*:

Tabel 3. Perubahan Struktur ke Pola Regex

	HTML	REGEX
Bagian Body	A: <div class ="basic-products basic-products--grid b : <section class="endorsed-products	<div\s*class\s*=\\"basic-products+([^\"]*)\s*.*?<section class="endorsed-products
Bagian Record	a : <li class ="product— sem col-12--2"> b :	<li\s*class\s*=\\"product--sem\s*col-12--2\">\s*.*?
Nama	data-name="nama barang"	data-name=\"([^\"]+)
Harga	harga barang	<span\s*class=\\"amount positive\">([^\"]+)
URL	data-url="alamat url"	data-url=\"([^\"]+)
Foto	data-src="url foto"	data-src=\"([^\"]+)

Ketika melakukan pemotongan data menggunakan pola diatas, data masih terdapat didalam struktur sehingga diperlukan proses *Replace First*[7] untuk menghilangkan struktur html tersebut. Setelah dilakukan pembersihan menggunakan *Replace First*[11] barulah data disimpan kedalam database [12].

Tabel 4. Hasil Replace First (RF)

No	Pola	RF	Data
Nama	data-name="Samsung Galaxy J3 Pro	Data name\s*=\s*"\"s*	Samsung Galaxy J3 Pro
Harga	167500	^\<span\s*class=\\"amount positive\">	167500
URL	Data url="http://olx.co.id/iklan/samsung-galaxy-j3-pro-IDlmnVB.html#536f5d98c4	data-url\s*=\s*"\"s*	http://olx.co.id/iklan/samsung-galaxy-j3-pro-IDlmnVB.html#536f5d98c4
Foto	data-src="https://img.olx.biz.id/C66A/50794/294949705_1_261x203_samsung-galaxy-j3-pro-yogyakarta-kota.jpg	data-src\s*=\s*"\"s*	https://img.olx.biz.id/C66A/50794/294949705_1_261x203_samsung-galaxy-j3-pro-yogyakarta-kota.jpg

4. Pengujian dan Implementasi

Berikut ini adalah tabel hasil pengujian program komparasi data yaitu dengan melakukan 20 kali inputan atau percobaan dan kemudian dihitung menggunakan *Recall dan Precission*[13]. Berikut ini adalah hasil *Recall dan Precission* :

Tabel 5. Pengujian

No	Inputan	a	b	a+b	c	A+c	Recall	Precission
1	Samsung Grand Prime	4	0	4	0	4	100%	100%
2	Samsung J3	3	0	3	0	3	100%	100%
3	Xiaomi MiMax	3	0	3	1	4	75%	100%
4	Xiaomi Redmi 4	3	6	9	0	3	100%	33%
5	Asus Zenfone 5	1	0	1	2	3	33%	100%
6	Asus Zenfone 2	1	1	2	1	2	50%	50%
7	Nokia X	1	3	4	0	1	100%	25%
8	Nokia XL	1	0	1	0	1	100%	100%
9	Lenovo a2010	2	0	2	0	2	100%	100%
10	Lenovo Vibe c	1	0	1	0	1	100%	100%
11	Oppo F1s	5	0	5	0	5	100%	100%
12	Oppo Neo 7	2	0	2	1	3	67%	100%
13	Iphone 5c	1	0	1	0	1	100%	100%
14	Kaos Fashion Nike	18	1	19	0	18	100%	95%
15	Kaos Fashion Three Second Seven Sunday	1	0	1	0	1	100%	100%
16	Tas Selempang Wanita Import Sling Bag Import fashion	1	0	1	0	1	100%	100%
17	Tas Wanita Tas Fashion Meimou 56876	1	0	1	0	1	100%	100%
18	Advan M6 Barca	1	0	1	0	1	100%	100%
19	Kaos Fashion Tomny Hilfiger	3	0	3	0	3	100%	100%
20	Iphone 5s	1	0	1	1	2	50%	100%
Rata – rata							89%	90%

Dari hasil pengujian diatas dapat disimpulkan bahwa tingkat keefektifan temu kembali pada program komparasi data bisa dikatakan sudah efektif. Hal ini terlihat dari rata-rata tingkat perolehan (*recall*) sebesar 89% dan rata-rata tingkat ketepatan (*precision*) yang dihasilkan sebesar 90%. Dari sejumlah kata kunci yang dicari, kata kunci yang nilai *recall* dan *precision* yang tinggi terdapat pada kata kunci "Samsung Grand Prime", "Samsung J3", "Nokia XL", "Lenovo a2010", "Lenovo Vibe C", "Kaos Malang", "Oppo F1S", "Iphone 5c", "Advan M6 Barca", "Advan S5D", "Evercross Ysmart", "Evercross A74R" dan "Fashion Malang" yang masing-masing mendapat *recall* dan *precision* 100%.

Hal tersebut dikarenakan jumlah yang relevan lebih besar atau seimbang jika dibandingkan dengan yang tidak relevan. Hal ini membuktikan bahwa nilai *recall* dan *precision* tinggi[14] karena sistem dapat mengendalikan pencarian[15] yang dilakukan dengan menggunakan kumpulan istilah atau kata kunci yang lebih spesifik dan kata yang digunakan sesuai dengan data sistem.

KESIMPULAN

Proses akhir penelitian yang dilakukan berupa hasil kesimpulan yang diperoleh dari tahapan proses yang dilakukan, dalam hal hasil pengujian disimpulkan bahwa program yang dirancang sudah efektif yakni dilihat dari perhitungan tingkat perolehan (*recall*) sebesar 89% dan rata-rata tingkat ketepatan (*precision*) yang dihasilkan sebesar 90%. Tingkat keefektifan program tergantung dari data yang sudah diambil yaitu berdasarkan penulisan merk oleh pengguna situs jual beli online. Kemudian, semakin spesifik merk yang ditulis oleh pengguna situs jual beli online maka sistem komparasi data akan lebih tepat.

DAFTAR RUJUKAN

- [1] V. Mitra, H. Sujaini, and A. B. P. Negara, "Rancang Bangun Aplikasi Web Scraping untuk Korpus Paralel Indonesia - Inggris dengan Metode HTML DOM," *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, vol. 5, no. 1, 2017.
- [2] N. R. Dyah, P. Astuti, F. Noviyanto, and D. Soyusiawati, "FORENSIK DIGITAL METODE REGEX (REGULAR EXPRESSION) DARI GRAB GOOGLE SEARCH API DALAM PROSES PELACAKAN TERHADAP KEJAHATAN ONLINE," 2018.
- [3] D. Glez-Peña, A. Lourenço, H. López-Fernández, M. Reboiro-Jato, and F. Fdez-Riverola, "Web scraping technologies in an API world," *Brief Bioinform*, vol. 15, no. 5, pp. 788–797, May 2013, doi: 10.1093/bib/bbt026.
- [4] R. Raghavendra and M. Niranjanamurthy, "Web Information Extraction methods using Web Content Mining (WCM) for Webapplications," *International Journal of Computing and Digital Systems*, vol. 11, no. 1, pp. 609–623, 2022, doi: 10.12785/ijcds/110149.
- [5] W. B. Trihanto, R. Arifudin, and A. Muslim, "Information Retrieval System for Determining The Title of Journal Trends in Indonesian Language Using TF-IDF and Naïve Bayes Classifier," *Scientific Journal of Informatics*, vol. 4, no. 2, pp. 2407–7658, 2017.
- [6] R. A. Fink, D. R. Zaret, R. B. Stonehirsch, R. M. Seng, and S. M. Tyson, "Streaming, Plaintext Private Information Retrieval Using Regular Expressions on Arbitrary Length Search Strings," in *Proceedings - 2017 IEEE Symposium on Privacy-Aware Computing, PAC 2017*, Dec. 2017, vol. 2017-January, pp. 107–118. doi: 10.1109/PAC.2017.35.
- [7] M. van Opijnen and C. Santos, "On the concept of relevance in legal information retrieval," *Artif Intell Law (Dordr)*, vol. 25, no. 1, pp. 65–87, Mar. 2017, doi: 10.1007/s10506-017-9195-8.
- [8] S. Sulaiman, R. A. Wahid, and F. Morsidi, "Feature extraction using regular expression in detecting proper noun for Malay news articles based on KNN algorithm," *Journal of Fundamental and Applied Sciences*, vol. 9, no. 5S, p. 210, Jan. 2018, doi: 10.4314/jfas.v9i5s.16.
- [9] M. Berglund, B. van der Merwe, and S. van Litsenborgh, "Regular expressions with lookahead," *Journal of Universal Computer Science*, vol. 27, no. 4, pp. 324–340, 2021, doi: 10.3897/JUCS.66330.
- [10] T. P. Shabna and M. K. Haneefa, "Web-Based Information Retrieval Pattern of Doctoral Students in Universities in Kerala," *Journal of Knowledge & Communication Management*, vol. 6, no. 1, p. 1, 2016, doi: 10.5958/2277-7946.2016.00001.2.
- [11] J. Khan, "Web-Based Information Retrieval Patterns Among Under Graduates, Post Graduates & Faculty Members in Delhi University: A Study," *Journal of Library & Information Science*, vol. 4, no. 3, 2014.
- [12] M. J. Hamid Mughal, "Data mining: Web data mining techniques, tools and algorithms: An overview," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 6, pp. 208–215, 2018, doi: 10.14569/IJACSA.2018.090630.

- [13] D. Garg and D. Sharma, "Information Retrieval on the Web and its Evaluation," 2013.
- [14] R. Geethanjali and M. S. Muruganantham, "Web Based Information Retrieval Using Dynamic Classified Average Precision Crawling Approach," 2015.
- [15] T. Plissac and J. R. nd Supriya Susan Kurian, "WEB BASED INFORMATION RETRIEVAL PATTERN OF FACULTY MEMBERS IN ST.ALOYSIUS COLLEGE, EDATHUA: A STUDY," *International Journal of Library & Information Science (IJLIS)*, vol. 5, no. 1, pp. 5–6, 2016.