



YAYASAN NURUL JADID PAITON
**LEMBAGA PENERBITAN, PENELITIAN, &
PENGABDIAN KEPADA MASYARAKAT
UNIVERSITAS NURUL JADID
PROBOLINGGO JAWA TIMUR**

PP. Nurul Jadid
Karanganyar Paiton
Probolinggo 67291
☎ 0888-3077-077
e: lp3m@unuja.ac.id
w: <https://lp3m.unuja.ac.id>

SURAT KETERANGAN

Nomor : NJ-To6/0358/A.03/LP3M/02.2021

Lembaga Penerbitan, Penelitian, dan Pengabdian Masyarakat (LP3M) Universitas Nurul Jadid Probolinggo menerangkan bahwa artikel/karya tulis dengan identitas berikut ini:

Judul : *Implementasi Analisa Semantik Pada Webdriver Untuk Otomatisasi Scraping Profil Pengguna Facebook*
Penulis : ZAINAL ARIFIN, S.Kom., M.Kom.
Identitas : Jurnal Keilmuan Dan Aplikasi Teknik Informatika, Vol. XII, No.2, Desember 2020, E-ISSN : 2549-354X, Universitas Yudharta Pasuruan
No. Pemeriksaan : 012099

Telah selesai dilakukan *similarity check* dengan menggunakan perangkat lunak **Turnitin** pada 01 Desember 2020 dengan hasil sebagai berikut: Tingkat kesamaan diseluruh artikel (*Similarity Index*) adalah 24 % dengan publikasi yang telah diterbitkan oleh penulis pada Jurnal Keilmuan Dan Aplikasi Teknik Informatika (<https://jurnal.yudharta.ac.id/v2/index.php/EXPLORE-IT/article/view/2283/1726>)

Demikian Surat Keterangan ini dibuat untuk digunakan sebagaimana mestinya.

Probolinggo, 02 Februari 2021

Kepala LP3M,



ACHMAD FAWAID, M.A., M.A.

NIDN. 2123098702

Implementasi Analisa Semantik Pada Webdriver Untuk Otomatisasi Scraping Profil Pengguna Facebook

by Zainal Arifin

Submission date: 01-Dec-2020 09:12PM (UTC+0900)

Submission ID: 1461320472

File name: Jurnal_PDP_2020.docx (490.25K)

Word count: 2250

Character count: 14122



10

IMPLEMENTASI ANALISA SEMANTIK PADA WEBDRIVER UNTUK OTOMATISASI SCRAPING PROFIL PENGGUNA FACEBOOK

Maulidiansyah¹, Zainal Arifin²

¹ Teknologi Informasi, Universitas Nurul Jadid, Probolinggo, Indonesia

² Teknologi Informasi, Universitas Nurul Jadid, Probolinggo, Indonesia

email: ¹maulid@unuja.com, ²zainal@unuja.ac.id

INFO ARTIKEL

Sejarah Artikel:
Diterima 6 Mei 2020
Direvisi -
Disetujui 8 Mei 2020
Dipublikasi Juni 2020

Katakunci:
Analisa Semantik
Webdriver
Facebook Bot

ABSTRAK

Penelitian ini membahas tentang implementasi Analisa semantik pada webdriver untuk mendapatkan data profil pengguna facebook secara otomatis. Pada era digital ini, data adalah sesuatu yang sangat berharga. Hal ini dikarenakan data dapat di olah menjadi pengetahuan baru yang bisa digunakan dalam semua lini kehidupan seperti pendidikan, ekonomi, bisnis dan lain-lain. Facebook adalah salah satu tempat penyimpanan data dalam jumlah besar. Hanya saja tidak mudah mendapatkan data dari Facebook. Dibutuhkan sebuah teknik yang bisa digunakan oleh sebuah bot untuk bisa menjelajah halaman Facebook, mengambil data profil pengguna dan mengelompokkannya berdasarkan kategori profil. Teknik tersebut adalah membuat bot dari Webdriver dengan mengimplementasikan analisa semantik.

Tahapan proses dalam penelitian ini adalah : Memahami struktur HTML pada halaman Facebook untuk mengetahui pola data profil pada Facebook. Kemudian Bot akan dikembangkan agar bisa membuka halaman profil pengguna, mengambil data profil dan kemudian menganalisanya berdasarkan pola pada struktur HTML halaman Facebook. Berdasarkan hasil uji coba pada internet dengan kecepatan 20Mbps, bot ini dapat mendapatkan 100 data profil pengguna Facebook dalam waktu ± 20 menit.

(Justify, Times New Roman 8, spasi single, italic. Abstrak memuat ringkasan isi tulisan: permasalahan yang dibahas, pendekatan atau solusi yang diusulkan dan hasil dari solusi yang diajukan/simpulan)

ABSTRACT

This research discusses the implementation of semantic analysis on the webdriver to get Facebook user profile data automatically. In this digital era, data has become invaluable. Data can be utilized for everyday applications such as in education, economics and business among many other industries. Facebook is a storage place for large amounts of data. But, retrieving profile data from Facebook is a complicated endeavor. Therefore, unique methods must be engineered within the bot processes. These countermeasures swiftly browse, retrieve user profile data, and group them based on profile categories. This technique uses the webdriver to create the bot by implementing semantic analysis

We begin by studying the HTML structure on any given Facebook page to determine the pattern of profile data. Then the Bot will be developed so that it can open the user's profile page, retrieve profile data and then analyze it based on the patterns on the HTML structure of the page. With download speeds of 20Mbps, test results indicate this bot is capable of extracting 100 Facebook user profiles in approximately 20 minutes

(Justify, Times New Roman 10, single space, italic. Abstract contains of summary of the writing : the problem being discussed, approach or suggested solution and result of the solution/ conclusion)

@ 2020 diterbitkan oleh Prodi Teknik Informatika Universitas Yudharta Pasuruan

1. Pendahuluan

Keberadaan data menjadi sangat berharga pada era ini. Hal ini karena dengan pengolahan data dapat diperoleh pengetahuan baru yang bisa digunakan untuk banyak hal seperti penelitian, memprediksi harga saham, melihat ketertarikan pembeli, analisa sentimen dan lain-lain. Proses tersebut biasa disebut dengan data mining. Data mining adalah proses untuk mengubah sesuatu yang tidak bermanfaat menjadi sesuatu yang berharga dan dapat memberikan suatu informasi yang penting [1]. Salah satu tempat yang menyimpan data penting dalam jumlah besar adalah sosial media seperti Facebook. Tapi tidak mudah untuk mendapatkan dari situs tersebut. Karena facebook tidak mengizinkan bot penjelajah otomatis untuk mengakses halamannya. Maka dari itu butuh tool seperti webdriver untuk mendapatkan data pengguna Facebook secara otomatis dan analisa semantik untuk mengolah dan mengelompokkan data pengguna pada kategori profil seperti pendidikan, alamat, link media sosial, no telepon, whatsapps dan lain-lain. Jadi dengan teknik ini pengambilan data pengguna dari situs Facebook bisa lebih cepat dan akurat.

25

Semantik bertujuan untuk menerjemahkan kalimat menjadi bentuk representasi lain[2]. Sedangkan teknik yang digunakan untuk mengekstraksi data HTML pada halaman Facebook adalah Regular Expression. Regular Expression berfungsi adalah teknik yang dominan digunakan untuk mengekstrak informasi saat ini[3].

2. Kajian Teori

a. Web Semantik

Dalam dunia teknologi informasi penggunaan Word Wide Web (WWW) sangat banyak digunakan untuk mencari informasi berdasarkan metadata, dalam pencarian metadata tersebut lebih dikenal dengan WEB Semantik dimana metode tersebut adalah sebuah kumpulan teknologi dalam dunia komputer yang mampu membedakan sebuah informasi yang berlandaskan pada metadata. Hal ini diharapkan dapat membantu pengguna dalam melakukan pencarian informasi secara cepat dan akurat[4].

b. Arsitektur Web Semantik

1). RDF (Resource Description Framework)

Sumber informasi yang didapatkan dan sebuah web baik itu biodata, tempat maupun sebuah dokumen dan data lainnya sehingga penggunaan URI (Iniform Resource Identifier) sebagai alat untuk mengidentifikasi secara unik dalam membedakan sebuah resource. Agar penggunaan sumber daya sebuah web (Web Resource) mampu dideskripsikan secara jelas, maka direkomendasikanlah sebuah bahasa RDF (Resource Description Framework) oleh lembaga Word Wide Web Consortium / W3C sebagai standart dalam mendeskripsikan sebuah web resource [2].

2). OWL (Web Ontology Language)

OWL merupakan salah satu bahasanya yang telah direkomendasi oleh lembaga W3C, dimana bahasa tersebut merupakan bahasa ontologi yang paling ekspresif digunakan dalam sebuah aplikasi semantik web yang digunakan untuk merepresentasi sebuah arti dari sebuah terminologi-terminologi pada sebuah pengetahuan dimana relasi pada terminologi tersebut tidak bisa dipisah dari doumen dan aplikasi web. Penggunaan OWL dimaksudkan untuk melakukan sebuah representasi makna dari sebuah kosa kata serta sebuah relasi antara kata dengan kata yang lain sehingga dapat menghasilkan sebuah informasi yang eksplisit. Pembuatan OWL dibentuk dari penggabungan RDF dan RDFS dan ditambahkan vocabulary agar sebuah classes, properties serta individuals dapat didefinisikan. OWL memberikan pilihan sub bahasa yang dapat digunakan sesuai dengan kebutuhan, diaman pilihan tersebut adalah OWL Lite, OWL DL, serta OWL FULL[5].

3). SPARQL (SPARQL Protocol and RDF Query Language)

Selain RDF dan OWL W3C juga membuat sebuah standart tentang penggunaan query dalam memperoleh data dari web resource, dimana standart tersebut dinamakan SPARQL (SPARQL Protocol and RDF Query Language). SPARQL Query memiliki triple pattern yaitu Subjek, predikat serta objek diman dari masing-masing kriteria tersebut dpat dijadikan sebuah variabel dalam penggunaan SPARQL. Penggunaan Query SPARQL berdasarkan kepada proses mencocokkan sebuah pola triple pada RDF, ada beberapa sintax yang dapat digunakan pada query SPRQL diantaranya adalah :

a) PREFIX Statemen PREFIX

Adalah sebuah cara menunjukkan informasi yang ada dalam sebuah resource dimana carai ini diwakili oleh URI (Uniform Resource Identifier), pada intinya PREFIX ini digunakan untuk membuat simple sebuah sumber data.

b) SELECT Statemen SELECT

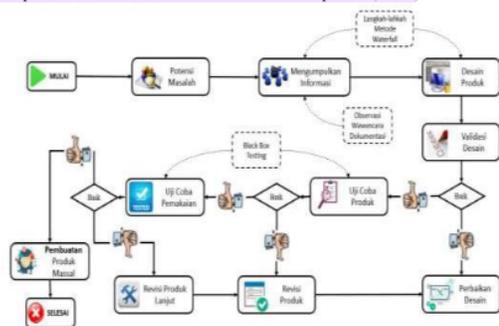
Pendefinisian sebuah variabel dimana hasil nya akan dikembalikan sebagai hasil proses query. Setiap variabel diawali dengan notasi "??".

c) WHERE

Penggunaan WHERE digunakan untuk mendefinisikan sejumlah pattern yang wajib dimiliki dari setiap proses query. Semua pola yang mewakili sebuah kalimat RDF harus sesuai dengan RDF Triple yang terdiri dari sebuah subjek, predikat, dan objek. RDF triple tersebut dapat direpresentasikan oleh URI atau sebuah variable dan nilai literal.

3. Metodologi Penelitian

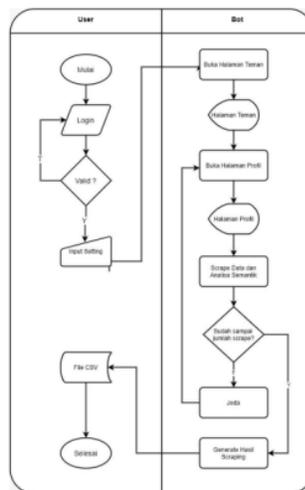
Metode penelitian yang digunakan dalam penelitian ini adalah Research and Development (R&D)



Gambar 1. Bagan Penelitian R&D
[Sumber: google.com]

- a. **Potensi dan Masalah**
Untuk mendapatkan sebuah permasalahan yang terjadi perlu dilakukannya sebuah penelitian awal. Hal ini dilakukan untuk mendapatkan informasi dan pengumpulan data serta mencari referensi yang tepat untuk mendukung dari sebuah penelitian
- b. **Mengumpulkan Informasi**
Proses pengumpulan data dilakukan dengan cara melakukan uji coba pencarian data tanpa menggunakan metode semantik. Peran ketua pengusul pada tahap ini adalah mencari sumber referensi tentang penggunaan semantik dan Webdriver Selenium. Sedangkan untuk anggota pengusul melakukan beberapa uji coba pencarian tanpa menggunakan metode semantik dan Webdriver Selenium
- c. **Desain Produk**
Proses ini dilakukan setelah proses pengumpulan data dilakukan, hal ini dilakukan untuk mempermudah dalam proses pengembangan aplikasi, pada proses ini ketua pengusul melakukan desain sistem, sedangkan anggota pengusul memberikan informasi terkait hasil uji coba pencarian untuk dijadikan acuan.
- d. **Validasi Desain**
Untuk mengetahui kesesuaian sebuah desain sistem maka perlu dilakukan validasi desain dengan aplikasi yang telah dihasilkan. ketua pengusul mengemukakan hasil rancangan desain dan aplikasi sedangkan anggota pengusul memberikan masukan sesuai dengan hasil. Langkah ini dilakukan untuk mevalidasi desain dan aplikasi agar siap diterapkan.
- e. **Perbaikan Desain**
Proses perbaikan desain dilakukan untuk menghasilkan desain yang lebih sempurna, proses ini dilakukan mengacu kepada hasil validasi desain yang telah dilakukan.
- f. **Uji Coba Produk**
Setelah desain dan aplikasi selesai dibuat maka perlu ada proses uji coba produk untuk mengetahui kesesuaian aplikasi dengan kebutuhan, pada proses ini ketua pengusul melakukan uji coba pencarian dengan menggunakan metode semantik, sedangkan anggota pengusul melakukan dokumentasi dari hasil percobaan.
- g. **Revisi Produk**
Revisi produk dilakukan apabila dalam tahapan uji coba mengalami ketidaksesuaian dengan sistem, revisi ini dilakukan dengan mengacu kepada hasil dokumentasi uji coba.
- h. **Uji Coba Pemakaian**
Proses pengujian ini dilakukan dengan sistem black box testing. Menurut Nidhra dan Dondeti black box testing juga disebut functional testing, sebuah teknik pengujian fungsional yang merancang test case berdasarkan informasi dari spesifikasi [6]. Pada proses ini ketua pengusul membuat draf kuisisioner untuk diberikan kepada masyarakat, sedangkan anggota pengusul membantu ketua dalam penyebaran kuisisioner
- i. **Revisi Produk Lanjut**
Pada proses ini dilakukan proses perbaikan yang mengacu pada hasil kuisisioner yang telah direkap oleh anggota pengusul, yang kemudian akan dijadikan acuan oleh ketua pengusul dalam melakukan perbaikan sistem dan aplikasi.
- j. **Pembuatan Produk Masal**
Proses ini adalah proses akhir yang dilakukan dalam proses pengembangan aplikasi agar bisa digunakan oleh masyarakat

4. Hasil Uji Coba Dan Pembahasan



Gambar 2. Flowchart Sistem

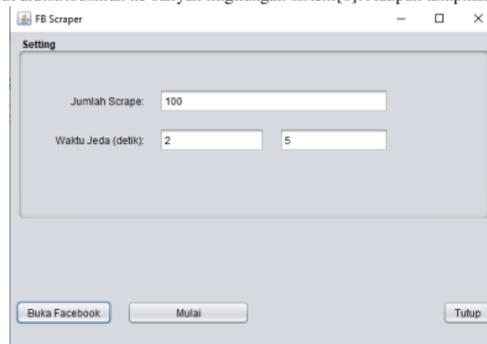
Dalam proses ekstraksi data profil pengguna Facebook, Kami menggunakan analisa semantik untuk mengenali kategori profil pengguna. Hal ini dilakukan karena data profil yang ditampilkan di halaman facebook tidak memiliki label kategori seperti data pendidikan. Semantik dapat menyamakan kata-kata atau kalimat yang memiliki makna yang sama meskipun kata-kata atau kalimat tersebut berbeda[7].



Gambar 3. Halaman Pendidikan Pengguna Facebook
[Sumber: facebook.com]

Pada gambar di atas facebook hanya menampilkan riwayat pendidikan pengguna, tapi tidak disebutkan jenjang pendidikan pada setiap riwayat pendidikan. Dengan semantik, kita bisa mengetahui jenjang setiap pendidikan hanya dari nama instansi pendidikan tersebut. Analisa semantik mengenali makna dari kata/kalimat berdasarkan pola. Pola dapat diketahui dengan menganalisa struktur HTML pada halaman Facebook. Contoh, jika pada jenjang pendidikan terdapat kata SMA, MA, SMK dan kata yang berkaitan dengan 3 kata tersebut seperti "Sekolah Menengah Atas", maka Riwayat pendidikan tersebut akan diberi masukkan ke jenjang pendidikan menengah atas.

Bot dikembangkan menggunakan bahasa pemrograman Java dan plugin Selenium Webdriver. Plugin ini bisa digunakan untuk membuat otomatisasi browser yang tangguh, terukur dan dapat didistribusikan ke banyak lingkungan sistem[8]. Adapun tampilan dari Bot adalah seperti gambar 3.



Gambar 4. Tampilan Bot

Pada Bot terdapat 2 setting yang harus di isi, pertama jumlah scrape dan waktu jeda. Waktu jeda ini bersifat acak antara waktu minimum dan maksimum. Hal ini digunakan untuk mengelabui bot detector milik Facebook dengan membuat bot ini berperilaku seolah-olah manusia. Setelah itu, pengguna user bisa menekan tombol "Buka Facebook" untuk membuka halaman login Facebook dan User akan diminta login secara manual. Selanjutnya user dapat menekan tombol "Mulai" untuk memulai proses scraping. Adapun tahapan proses scraping adalah sebagai berikut :

1. Bot akan membuka halaman daftar teman



Gambar 5. Halaman Daftar Teman Facebook

2. Pada halaman daftar teman, Bot akan membuka halaman profil setiap teman facebook satu-persatu



Gambar 6. Halaman “Tentang” Facebook

3. Bot akan mengambil data profil teman facebook dan mengelompokkan data tersebut menggunakan metode Analisa Semantik (Regular Expression)

Adapun hasil dari scraping bot ini adalah berupa file CSV yang berisi data profil pengguna Facebook yang sudah dikelompokkan ke dalam kolom-kolom.

Tabel 1 : Data Hasil Scraping

Nama User	Jenis Kelamin	Riwayat Pekerjaan	Dasar	Menengah Pertama	Menengah Atas
Mahmud	Laki-laki	Guru di SMP 1 Probolinggo, Programmer di PT Jaya Makmur	-	SMP NJ	SMK NJ
Siti	Perempuan	Admin di Toko ABC, Owner TOKO XYZ	SD Negeri 1 Kraksaan	MTs NJ	MA Nurul Jadid
Rudianto	Laki-laki	Pelajar	-	MTs Miftahul Huda	SMK 2 Probolinggo
M. Fajar	Laki-laki	Setia Mencari Cinta	-	SMP NJ	SMA Nurul Jadid

Tabel 1 : Data Hasil Scraping (Lanjutan)

Perguruan Tinggi	No Telp	Alamat	Kepercayaan	Bahasa
UNUJA	085234124865	Probolinggo	Islam	Jawa, Indonesia
UNUJA	085234125462	Surabaya	Islam	Madura, Indonesia
UNUJA	085264532569	Kraksaan	Islam moderat	Madura, Indonesia
-	082315648952	paiton	Islam	Indonesia

5. Kesimpulan

Berdasarkan hasil uji coba pada internet dengan kecepatan 20Mbps, Bot ini dapat mendapatkan 100 data profil pengguna Facebook dalam waktu ± 20 menit. Bot ini juga bisa dijalankan secara paralel untuk mendapatkan hasil *scraping* yang lebih cepat dan singkat. Selain itu, kecepatan internet dan spesifikasi PC yang digunakan akan berpengaruh terhadap kecepatan Bot. Dari hasil uji coba tersebut dapat disimpulkan bahwa Bot ini bisa mendapatkan data profil pengguna Facebook dalam jumlah banyak dengan waktu yang relatif singkat. Selain itu, Bot ini juga bisa mengelompokkan data tersebut berdasarkan kategori profil seperti pendidikan.

6. Daftar Pustaka

- [1] S. Rizal, P. Studi, T. Informatika, and U. Yudharta, “Penerapan Algoritma Naïve Bayes Untuk Prediksi Penerimaan Siswa Baru Di Smk Al-Amien Wonorejo,” *Explor.IT J. Keilmuan dan Apl. Tek. Inform.*, vol. 10, no. 1, pp. 14–21, 2018. doi: 10.35891/explorit.v10i1.1671.
- [2] J. Suciadi, “Studi Analisis Metode-Metode Parsing Dan Interpretasi Semantik Pada Natural Language Processing,” *J. Inform.*, vol. 2, no. 1, pp. 22–22, 2001. doi: 10.9744/informatika.2.1.pp.13-22.

- 8
- [3] Y. Li, R. Krishnamurthy, S. Raghavan, S. Vaithyanathan, and H. V. Jagadish, "Regular expression learning for information extraction," *EMNLP 2008 - 2008 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf. A Meet. SIGDAT, a Spec. Interes. Gr. ACL.*, no. October, pp. 21–30, 2008, doi: 10.3115/1613715.1613719.
 - [4] M. Unik and M. Ramli, "Penerapan Metode Semantik Search Dalam Sistem Informasi Pencarian Kerja Praktek Dan Skripsi Berbasis Web," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 3, no. 2, pp. 83–88, 2018, doi: 10.36341/rabit.v3i2.480.
 - [5] Y. F. Badron, F. Agus, and H. R. Hatta, "STUDI TENTANG PEMODELAN ONTOLOGI WEB SEMANTIK DAN PROSPEK PENERAPAN PADA BIBLIOGRAFI ARTIKEL JURNAL ILMIAH," *Pros. Semin. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 1, pp. 2338–5081, 2017, [Online]. Available: <http://journal.student.uny.ac.id/ojs/ojs/index.php/pgsd/article/viewFile/135/130>.
 - [6] S. Nidhra, "Black Box and White Box Testing Techniques - A Literature Review," *Int. J. Embed. Syst. Appl.*, vol. 2, no. 2, pp. 29–50, 2012, doi: 10.5121/ijesa.2012.2204.
 - [7] N. Riza Akbar, R. Faisal, and H. Budi, "Pengembangan Data Uji Sistem Komputasi Kemiripan Teks Secara Semantik Berbahasa Indonesia," *Semin. Inform. Apl. Polinema*, 2016.
 - [8] Anonymous, "Selenium WebDriver. Internet: <https://www.selenium.dev>, 03 September 2020".

Cek Similarity Jurnal Zainal Arifin, M.kom

ORIGINALITY REPORT

24%

SIMILARITY INDEX

25%

INTERNET SOURCES

5%

PUBLICATIONS

6%

STUDENT PAPERS

PRIMARY SOURCES

1

njca.co.id

Internet Source

6%

2

e-journals.unmul.ac.id

Internet Source

3%

3

id.123dok.com

Internet Source

1%

4

jurnal.fikom.umi.ac.id

Internet Source

1%

5

repositori.uin-alauddin.ac.id

Internet Source

1%

6

search.unikom.ac.id

Internet Source

1%

7

jurnal.stiki-indonesia.ac.id

Internet Source

1%

8

Chaofan Tu, Menglin Cui. "Learning Regular Expressions for Interpretable Medical Text Classification Using a Pool-based Simulated Annealing Approach", 2020 IEEE Congress on

1%

Evolutionary Computation (CEC), 2020

Publication

9	jurnal.polibatam.ac.id Internet Source	1%
10	dppm.uui.ac.id Internet Source	1%
11	dev.kinetik.umm.ac.id Internet Source	1%
12	Submitted to UIN Sunan Ampel Surabaya Student Paper	1%
13	jurnal.univrab.ac.id Internet Source	1%
14	ejournal.uin-suska.ac.id Internet Source	1%
15	eprints.umg.ac.id Internet Source	<1%
16	if.yudharta.ac.id Internet Source	<1%
17	jurnal.uisu.ac.id Internet Source	<1%
18	www.lab.biblioteconomia.unir.br Internet Source	<1%
19	boris.unibe.ch Internet Source	<1%

20	id.m.wikipedia.org Internet Source	<1%
21	repository.ubaya.ac.id Internet Source	<1%
22	jurnal.unswagati.ac.id Internet Source	<1%
23	docplayer.fi Internet Source	<1%
24	jurnal.unej.ac.id Internet Source	<1%
25	puslit2.petra.ac.id Internet Source	<1%
26	jurnalmahasiswa.unesa.ac.id Internet Source	<1%
27	researchonline.jcu.edu.au Internet Source	<1%

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off

Cek Similarity Jurnal Zainal Arifin, M.kom

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6
