

## WEB SCRAPING SENYAWA HERBAL DI INDONESIA MENGUNAKAN SELENIUM PYTHON

Nur Fadhilla Rosia Afrianti<sup>1</sup>, Afif Badawi<sup>2</sup>, Muslim<sup>3</sup>

Universitas Panca Budi, Medan

e-mail: <sup>1</sup>nurfadhillar8@gmail.com, <sup>2</sup>afifbadawi@dosen.pancabudi.ac.id,

<sup>3</sup>muslim@dosen.pancabudi.ac.id

**Abstract:** *Indonesia is one of the countries with the highest number of herbal plant species in the world. This wealth of biodiversity shows that the development of herbal compounds has the potential to develop the economy in Indonesia. However, in reality, data and research related to herbal compounds are still difficult to find. In this study, a web scraper will be built that aims to collect data on compounds and herbal plants in Indonesia from various databases. The scraper is built using Selenium Python on the KNApSAcK website. The automatic scraping process with Selenium produces 1341 herbal plants in Indonesia and 954 parts of herbal plants used for the composition of various herbal medicines in Indonesia.*

**Keywords:** *Indonesia herbal plants, Web scraping, Selenium, Herbal medicine*

**Abstrak:** Indonesia merupakan salah satu negara dengan jumlah spesies tanaman herbal tertinggi di dunia. Kekayaan biodiversitas ini menunjukkan bahwa pengembangan senyawa herbal memiliki potensi untuk mengembangkan perekonomian di Indonesia. Namun pada kenyataannya data dan penelitian terkait senyawa herbal masih sulit ditemukan. pada penelitian ini akan dibangun suatu web scraper yang bertujuan untuk mengumpulkan data senyawa dan tanaman herbal yang ada di Indonesia dari berbagai basis data. Scraper dibangun menggunakan Selenium Python pada situs KNApSAcK. Proses scraping otomatis dengan Selenium menghasilkan 1341 tanaman herbal yang ada di Indonesia dan 954 bagian tanaman herbal yang digunakan untuk komposisi berbagai jamu di Indonesia.

**Kata kunci:** Tanaman herbal Indonesia, Web scraping, Selenium, Jamu

### PENDAHULUAN

Indonesia merupakan salah satu negara dengan jumlah spesies tanaman herbal tertinggi di dunia (Salim & Munadi, 2017). Familiaritas masyarakat Indonesia terhadap tanaman herbal juga tinggi dikarenakan tanaman tersebut telah sering dikonsumsi masyarakat secara turun temurun (Elfahmi et al., 2014). Riset Kesehatan Dasar pada tahun 2010 menunjukkan bahwa lebih dari 50% masyarakat telah mengkonsumsi obat herbal (Purwaningsih, 2013). Kekayaan biodiversitas ini menunjukkan bahwa pengembangan senyawa herbal memiliki potensi untuk mengembangkan

perekonomian di Indonesia (Elfahmi et al., 2014).

Senyawa herbal dianggap dapat memberikan khasiat yang lebih baik, efek samping yang lebih sedikit dan harga yang lebih murah dibandingkan dengan senyawa konvensional (Ekor, 2013). Penelitian (Zuhud & Siswoyo, 2001) mengidentifikasi sekitar 1845 tanaman hutan di Indonesia memiliki potensi untuk menjadi tanaman obat. BPOM juga telah mencatat sekitar 283 spesies tanaman dapat digunakan sebagai sarana pengobatan alternatif.

Namun pada kenyataannya data dan penelitian terkait senyawa herbal masih sulit ditemukan (Noor et al., 2022). Oleh

karena itu pada penelitian ini akan dibangun suatu web scraper yang bertujuan untuk mengumpulkan data senyawa dan tanaman herbal yang ada di Indonesia dari berbagai basis data. Scraper dibangun menggunakan Selenium Python. Selenium merupakan framework open source yang dapat menjalankan otomatisasi penjelajahan berbagai situs dari berbagai browser dan platform. Salah satu keuntungan Selenium ada pada penggunaan webdriver yang mudah dibaca dan diatur (Paul & Tommy, 2018).

## METODE

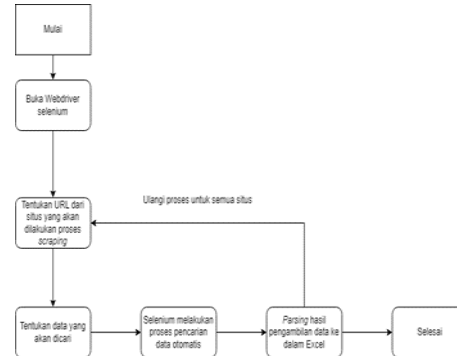
Adapun langkah pengerjaan pada penelitian ini, yaitu:

1. Identifikasi masalah  
Adapun tujuan utama penelitian adalah melakukan web scraping menggunakan Selenium python untuk mengumpulkan data senyawa herbal dari berbagai situs.
2. Studi literatur  
Untuk membantu penelitian ini, penulis melakukan studi literatur sumber sumber yang berkaitan dengan penelitian, baik dalam bentuk buku, jurnal maupun karya ilmiah lain yang relevan.
3. Pemilihan situs untuk pengambilan data  
Situs yang akan digunakan untuk proses web scraping adalah KNAPsAcK (Afendi et al., 2012).
4. Web scraping dengan Selenium  
Selenium dibangun dengan menggunakan bahasa Python dengan IDE menggunakan Jupyter Notebook dan webdriver menggunakan Google Chrome.

## Selenium

Pemrograman Selenium menggunakan ide Jupyter Notebook dan Google Chrome Webdriver. Penggunaan Webdriver dapat membantu mengamati proses scraping yang sedang berjalan sehingga dapat dengan mudah mendeteksi error. Adapun langkah proses scraping

dengan selenium dapat dilihat pada Gambar 1.



**Gambar 1** Proses scraping dengan Selenium

URL yang digunakan untuk proses scraping adalah [http://www.knapsackfamily.com/knapsack\\_core/top.php](http://www.knapsackfamily.com/knapsack_core/top.php) (KNAPsAcK). Proses pencarian otomatis dilakukan dengan mencari semua senyawa herbal yang ada di Indonesia pada situs. Hasil pencarian akan berupa web page yang berbentuk HTML dan/atau XML.

Kemudian web page yang sudah didapat akan di-parsing menggunakan library BeautifulSoup. Data senyawa herbal yang disimpan dalam bentuk tabel pada situs penyedia data. Proses parsing bertujuan untuk mendeteksi tabel data tersebut dikarenakan pada tampilan situs terdapat berbagai elemen HTML lain seperti <div>, <footer> dan lain lain. Proses kemudian diulang sampai semua web page data telah selesai diproses. Data kemudian di-parsing menjadi bentuk Excel.

## HASIL DAN PEMBAHASAN

Pencarian semua tanaman herbal di Indonesia Pencarian semua tanaman herbal di Indonesia dilakukan pada situs KnapSacK World ([http://www.knapsackfamily.com/KNAPsAcK\\_World/top.php](http://www.knapsackfamily.com/KNAPsAcK_World/top.php)). Situs menampilkan data spesies tanaman dari berbagai negara dengan total 229 negara berbeda dan 76.100 spesies tanaman. Gambar 2

menunjukkan tampilan awal halaman situs.



**Gambar 2 tampilan awal situs knapsack world**

Pencarian spesies tanaman di Indonesia dapat dilakukan dengan mengklik simbol negara Indonesia pada situs. Pencarian dilakukan dengan Selenium dengan webdriver google chrome. Kemudian Selenium melakukan klik otomatis pada situs dengan mencari XPATH simbol negara Indonesia. XPATH negara Indonesia: “/html/body/table[2]/tbody/tr[168]/td[8]/a/img”. Hasil klik otomatis Selenium menghasilkan jendela baru pada webdriver yang dirincikan pada Gambar 3.



**Gambar 3 Tampilan jendela baru pada webdriver**

Pada halaman terdapat berbagai informasi mengenai spesies tanaman di Indonesia, baik tanaman obat ataupun tanaman yang dapat dikonsumsi. Terdapat kolom nama spesies, klasifikasi, nama Family, nama lokal, fungsi tanaman, dan referensi. Tabel pada situs kemudian akan di-parsing menggunakan BeautifulSoup dengan mencari nama ID tabel, yang dalam hal ini ID tabel adalah: “tablekit-table-1”. Kemudian tabel hasil dirapikan dengan menghilangkan kolom dengan

aksara jepang dan baris yang berisi NaN pada baris awal. Hasil parsing dapat dilihat pada Gambar 4

	Species Name	UpperClassification(Region ; Classification)	FamilyName	CommonName	Purpose
1	Abelmoschus manihot	plants(vegetables)	Malvaceae	NaN	medicinal
2	Abelmoschus manihot (L.) Medic.	plants	Malvaceae	NaN	medicinal
3	Abelmoschus moschatus (L.) Medic.	plants	Malvaceae	NaN	medicinal
4	Abelmoschus moschatus Medic.	plants	Malvaceae	NaN	medicinal
5	Abrus fruticulosus Wight & Arn.	plants	Fabaceae	NaN	medicinal
...	...	...	...	...	...
2123	Zizyphus jujuba	plants	Rhamnaceae	NaN	edible
2124	Zizyphus napeca	plants	Rhamnaceae	NaN	edible
2125	Zizyphus rotundifolia	plants	Rhamnaceae	NaN	edible
2126	Zizyphus rugosa	plants	Rhamnaceae	NaN	edible
2127	Zizyphus xylopyrus	plants	Rhamnaceae	NaN	edible

**Gambar 4 Hasil parsing halaman situs dengan BeautifulSoup.**

Hasil pengambilan data menunjukkan bahwa terdapat 1341 tanaman yang memiliki tujuan medicinal atau obat, dan 786 tanaman yang memiliki tujuan edible atau dapat dimakan. Dari data tersebut, terdapat 36 tanaman yang tergolong dapat dimakan dan merupakan tanaman obat, diantaranya bawang merah (*Allium cepa*), belimbing (*Averrhoa carambola*), dan manggis (*Garcinia mangostana*).

Pencarian tanaman herbal untuk jamu Pencarian tanaman herbal yang digunakan untuk pembuatan jamu dilakukan pada situs KnapSack jamu (<http://www.knapsackfamily.com/jamu/top.php>). Tampilan awal situs dapat dilihat pada Gambar 5.



**Gambar 5 Tampilan awal situs KnapSack jamu.**

Pencarian dimulai dengan mengklik tombol Herb List yang akan menampilkan seluruh tanaman herbal yang menjadi bahan jamu. Kemudian pilih satu per satu tanaman herbal dan tekan tombol search untuk mencari jamu yang berhubungan dengan tanaman tersebut. Terdapat 1133 tanaman herbal yang diproses menjadi jamu pada situs KNApSack. Contoh hasil pencarian dapat dilihat pada Gambar 6.



**Gambar 6 Tampilan hasil pencarian jamu untuk tanaman herbal waron (Abelmoschus moschatus Folium)**

Pada menu hasil terdapat informasi dari tanaman herbal seperti nama latin, nama lokal di Indonesia, nama lokal dalam Bahasa Inggris/cina, dan bagian tanaman yang dipakai untuk jamu. Selain itu juga terdapat informasi mengenai jamu yang menggunakan tanaman tersebut sebagai komposisinya.

Proses scraping dilakukan dengan Selenium dengan mengotomatisasi pemilihan tanaman herbal satu per satu dan melakukan pencarian jamu untuk tiap senyawa herbal. Adapun langkah otomatisasi Selenium:

1. Cari XPATH dari tombol herbal list dan tekan otomatis dengan webdriver selenium. XPATH tombol herbal list: `“/html/body/table/tbody/tr[1]/td/input”`
2. Cari XPATH dari masing masing tanaman herbal untuk dilakukan klik otomatis. XPATH: `/html/body/table[2]/tbody/tr[{x}]/td[2]/input`, dimana x merupakan index masing masing tanaman herbal

dengan index 3-1135.

3. Klik tombol Ok dengan XPATH `“/html/body/input[1]”` dan search dengan XPATH `“/html/body/table[2]/tbody/tr[3]/td/form/input”`
4. Parsing hasil pencarian ke dalam tabel dataframe python dan gabungkan dengan hasil pencarian sebelumnya.
5. Ulangi langkah sampai semua pencarian tanaman herbal selesai

Herb Name	Herb Local Name(Indonesia)	Herb Local Name(English/Chinese)	Science Name	Position of Plants	effect	List Jamu
Abelmoschus moschatus Folium	Waron	MUSKALLODI, MUSK OKRA	Abelmoschus moschatus Medik	Leaf	Halal	[PT. Bio-Life MedLab - Gudang] [PT. Bio-Life MedLab - Pro - Rien]
Abelmoschus moschatus Oleum	Waron	MUSKALLODI, MUSK OKRA	Abelmoschus moschatus Medik	Oil	Halal	[PT. Septa Sari - Minyak Sangkil Palung]
Abus precatorius Folium	Saga	JEOURTY	Abus precatorius L.	Leaf	Halal	[CV. Alam Aulia - Jamu Formula 23 (Waron)] [CV. Alam Aulia - Jamu Formula 23 (Waron)]
Acacia sebalana Radix	Wata	Acacia sebalana DC	Acacia sebalana DC	Root	Halal	[PT. Kanner - Hebatidri Saji jaha] [PT. Kanner - Hebatidri Saji jaha]
Acathopanax serrifolius Cortex	Notopneng	TEHCH	Acathopanax serrifolius	Bark	Halal	[FA. Berkah Jaya - Obat Candi Galin] [PT. D. ...]
Zingiber officinale Rhizoma	Jahe	GINGER	Zingiber officinale Rosc	Rhizome	Halal	[41 Mancur - Jamu Antor - Antor Madu] [41 Mancur - Jamu Antor - Antor Madu]
Zingiber purpurum Rhizoma	Bangle	Zingiber purpurum Rosc	Zingiber purpurum Rosc	Rhizome	Halal	[41 Mancur - Jamu Antor - Antor Madu] [41 Mancur - Jamu Antor - Antor Madu]
Zingiber zerumbet Rhizoma	Lempuyang Gajah	Qiu Jiang (Ginger)	Zingiber zerumbet Sm	Rhizome	Halal	[CV. Prima Sari - Primar] [CV. Sari Bumi Pa. ...]
Zizophus jujuba Fructus	Chinese Date	JUJUBE, COMMON JUJUBE, CHINESE DATE	Zizophus jujuba Mill	Fruit	Halal	[CV. Natural Hills - Hono Sho-Sako-To Gran. ...]
Zizophus spina-christi Semen	Waj Jubate	SYRIAN CHRIST-THORN	Zizophus spina-christi (L.) Desf	Seed	Halal	[PT. Cahaya Baru - SEHAT JANTUNG PL] [PT. ...]

**Gambar 7 Hasil scraping tanaman herbal untuk jamu**

Terdapat informasi seperti nama latin tanaman, nama lokal, bagian tanaman dan Jamu apa saja yang berhubungan dengan tanaman herbal tersebut. Proses scraping menghasilkan total 954 data tanaman herbal yang digunakan untuk berbagai jamu. Beberapa tanaman herbal juga memiliki bagian yang berbeda yang dapat dijadikan jamu, seperti jahe (*Zingiber officinale*) yang minyak dan bagian rhizoma dapat dijadikan komposisi untuk beberapa jamu yang berbeda.

## SIMPULAN

Proses web scraping otomatis dengan Selenium pada situs KNApSack telah berhasil dilakukan untuk mencari informasi tanaman herbal di Indonesia dan tanaman herbal yang digunakan untuk

pembuatan jamu. Proses scraping menghasilkan total 1341 tanaman herbal yang terdapat di Indonesia dengan kategori tanaman obat atau tanaman yang dapat dikonsumsi. Selain itu terdapat 954 bagian tanaman yang berbeda yang sudah dijadikan jamu oleh berbagai perusahaan di Indonesia. Proses otomatisasi web scraping membantu pengerjaan dalam pengumpulan data herbal di Indonesia yang selanjutnya dapat digunakan untuk kepentingan lain seperti penelitian lanjutan dan lain lain.

#### DAFTAR PUSTAKA

- Afendi, F. M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., Ikeda, S., Takahashi, H., Altaf-Ul-Amin, M., Darusman, L. K., Saito, K., & Kanaya, S. (2012). KNApSAcK family databases: integrated metabolite-plant species databases for multifaceted plant research. *Plant & Cell Physiology*, 53(2).  
<https://doi.org/10.1093/PCP/PCR165>
- Ekor, M. (2013). The growing use of herbal medicines: issues relating to adverse reactions and challenges in monitoring safety. *Frontiers in Pharmacology*, 4.  
<https://doi.org/10.3389/FPHAR.2013.00177>
- Elfahmi, Woerdenbag, H. J., & Kayser, O. (2014). Jamu: Indonesian traditional herbal medicine towards rational phytopharmacological use. *Journal of Herbal Medicine*, 4(2), 51–73.  
<https://doi.org/10.1016/J.HERMED.2014.01.002>
- Noor, F., Tahir ul Qamar, M., Ashfaq, U. A., Albutti, A., Alwashmi, A. S. S., & Aljasir, M. A. (2022). Network Pharmacology Approach for Medicinal Plants: Review and Assessment. *Pharmaceuticals (Basel, Switzerland)*, 15(5), 572.  
<https://doi.org/10.3390/PH15050572>
- Paul, N., & Tommy, R. (2018). An Approach of Automated Testing on Web Based Platform Using Machine Learning and Selenium. *Proceedings of the International Conference on Inventive Research in Computing Applications, ICIRCA 2018*, 851–856.  
<https://doi.org/10.1109/ICIRCA.2018.8597297>
- Purwaningsih, E. (2013). Jamu, Obat Tradisional Asli Indonesia: Pasang Surut Pemanfaatannya di Indonesia. *EJournal Kedokteran Indonesia*, 1(2).  
<http://journal.ui.ac.id/index.php/eJKI/article/view/2065>
- Salim, Z., & Munadi, E. (2017). Info komoditi tanaman obat. *Badan Pengkajian dan Pengembangan Perdagangan Kementerian Perdagangan Republik Indonesia*.
- Zuhud, E., & Siswoyo. (2001). *Rancangan Strategi Konservasi Tumbuhan Obat Indonesia*. Kerjasama Badan Pengendali Dampak Lingkungan dengan Fakultas Kehutanan, IPB.