

1. ANALISIS SENTIMEN PADA ULASAN APLIKASI TIKTOK TRAGEDI STADION KANJURUHAN MENGGUNAKAN METODE LEXICON (SENTIMENT ANALYSIS ON TIKTOK APPLICATION REVIEWS KANJURUHAN STADIUM TRAGEDY USING THE LEXICON METHOD)
Faruq Amrulloh, Siti Madinah Ladjamuddin

2. PERANCANGAN SISTEM PENDUKUNG KEPUTUSAN PENANGANAN TREND PENJUALAN PADA PT. DORBEAN MUSIC INDONESIA (DESIGN OF A DECISION SUPPORT SYSTEM FOR HANDLING SALES TRENDS AT PT. DORBEAN MUSIC INDONESIA)
Dedy Iskandar

3. GAME EDUKASI KUIS PILIHAN GANDA MENGGUNAKAN GAME ENGINE UNITY 3D (MULTIPLE CHOICE QUIZ EDUCATIONAL GAME USING UNITY 3D GAME ENGINE)
Ade Bayu Saputra, Andi Suprianto, Al-Bahra Bin Ladjamudin

4. PERANCANGAN SISTEM PERSEDIAAN BARANG DALAM PERGUDANGAN PADA PT. PUTRABAKTI PARAMUDA BERBASIS WEB (DESIGN OF A WEB-BASED INVENTORY SYSTEM IN THE WAREHOUSE AT PT. PUTRABAKTI PARAMUDA)
Denny Andriyan

5. ANALISIS HUBUNGAN POPULARITAS SEIYUU DENGAN ANIME MENGGUNAKAN METODE PENGAMBILAN DATA WEB SCRAPING PADA SITUS MYANIMELIST.NET (ANALYSIS OF THE RELATIONSHIP BETWEEN SEIYUU POPULARITY AND ANIME USING THE WEB SCRAPING DATA RETRIEVAL METHOD ON THE MYANIMELIST.NET SITE)
Venerdi Rafid Muhajir, Aryo Nur Utomo

6. PERANCANGAN PEMBELAJARAN E-LEARNING PADA SMK KUSUMA BANGSA TANGERANG (DESIGN OF E-LEARNING INSTRUCTION AT SMK KUSUMA BANGSA TANGERANG)
Sugina

7. RANCANGAN SISTEM FINGERPRINT PRESENSI KARYAWAN BERBASIS WEB (DESIGN OF A WEB-BASED FINGERPRINT EMPLOYEE PRESENCE SYSTEM)
Siti Nurmisti, Luthfi Alkhafid

8. PERANCANGAN SISTEM INFORMASI PERSEDIAAN BAHAN BAKU PADA PT. ANEKA SARIVITA (DESIGN OF A RAW MATERIAL INVENTORY INFORMATION SYSTEM AT PT. ANEKA SARIVITA)
Suhada

9. PEMANFAATAN POLA HETEROGEN TPS UNTUK OPTIMASI STRATIFIELD SAMPLING QUICK COUNT (UTILIZATION OF HETEROGENEOUS TPS PATTERNS FOR OPTIMIZING STRATIFIELD SAMPLING IN QUICK COUNT)
Riadi Marta Dinata

10. PERANCANGAN SISTEM INFORMASI SURAT MASUK DAN SURAT KELUAR PADA SMK PARADUTA BANGSA TANGERANG SELATAN BERBASIS WEB (WEB-BASED DESIGN OF INFORMATION SYSTEM FOR INCOMING LETTER AND OUTGOING LETTER AT PARADUTA BANGSA VOCATIONAL SCHOOL, SOUTH TANGERANG)
Martono



Diterbitkan oleh

**PROGRAM STUDI TEKNIK INFORMATIKA
INSTITUT SAINS DAN TEKNOLOGI NASIONAL**

Jl. Moh.Kahfi II, Bhumi Srengseng Indah, Jagakarsa, Jakarta Selatan 12640

Telp : 021-7874647 Fax : 021-7866955

E-mail : proditi.istn@yahoo.com

ANALISIS HUBUNGAN POPULARITAS SEIYUU DENGAN ANIME MENGGUNAKAN METODE PENGAMBILAN DATA WEB SCRAPING PADA SITUS MYANIMELIST.NET

ANALYSIS OF THE RELATIONSHIP BETWEEN SEIYUU POPULARITY AND ANIME USING THE WEB SCRAPING DATA RETRIEVAL METHOD ON THE MYANIMELIST.NET SITE

Venerdi Rafid Muhajir¹, Aryo Nur Utomo²

Program Studi Teknik Informatika, Fakultas Sains dan Teknologi Informasi
Institut Sains dan Teknologi Nasional Jl. Moh. Kahfi II, Bhumi Srengseng Indah, Jakarta
Selatan 12640 Telp. (021) 7874647,
venerdirafid@gmail.com , aryo.nurutomo@gmail.com

ABSTRAKSI

Saat ini, volume data yang sangat besar dihasilkan setiap hari dengan kecepatan yang belum pernah terjadi sebelumnya dari berbagai sumber, termasuk data di internet. Keberadaan internet membuat jumlah data yang terdapat di website – website semakin bertambah, sehingga dibutuhkan metode yang efisien untuk mengumpulkan banyaknya data yang ada di suatu website, dan web scraping termasuk salah satunya. Pada penelitian ini, peneliti mengimplementasikan metode web scraping pada situs myanimelist.net yang merupakan komunitas dan database seputar anime dan manga terbesar di dunia. Tujuan dari penelitian ini adalah untuk menganalisis hubungan antara seiyuu (pengisi suara) dan anime yang diperankannya melalui analisis korelasi. Metode penelitian ini secara garis besar terdiri dari lima tahapan, meliputi data extracting, data cleansing, data processing, statistic test, dan data visualization. Alat penelitian yang digunakan untuk scraping adalah beautifulsoup dan pandas yang merupakan library python, sedangkan alat untuk proses dan analisis data hasil scraping adalah google colaboratory dan pandas yang merupakan aplikasi pengolahan data. Hasil dari penelitian ini menunjukkan bahwa web scraping dapat diimplementasikan pada situs myanimelist.net dengan efektif. Peneliti dapat mengumpulkan 54298 baris data dari 275 orang dalam waktu 8 hari. Hasil analisis korelasi menunjukkan popularitas seiyuu memiliki hubungan sedang dengan popularitas karakter yang meraka perankan dan jumlah anime yang mereka mainkan..

Kata Kunci : big data, web scraping, anime

ABSTRACT

Today, enormous volumes of data are generated every day at an unprecedented speed from various sources, including data on the internet. The existence of the internet makes the amount of data contained on the website increase, so we need an efficient method to collect a lot of data on a website, and web scraping is one of them. In this study, the researchers implemented the web scraping method on the myanimelist.net site which is the largest community and database of anime and manga in the world. The purpose of this study is to analyze the relationship between the seiyuu (voice actors) and the anime they play through correlation. This research method broadly consists of five stages, including data extracting, data cleansing, data processing, statistical testing, and data visualization. The research tools used for scraping are beautifulsoup and pandas which are python libraries, while the tools for processing and analyzing scraping data are collaborative google and pandas which are data processing applications. The results of this study indicate that web scraping can be implemented on the myanimelist.net site effectively. Researchers were able to collect 54298 rows of data from 275 people within 8 days. The results of the analysis show that the popularity of seiyuu has a moderate relationship with the popularity of the characters they play and the number of anime they play..

Keywords : big data, web scraping, anime..

1. PENDAHULUAN

Saat ini, volume data yang sangat besar dihasilkan setiap hari dengan kecepatan yang belum pernah terjadi sebelumnya dari berbagai sumber (misalnya, kesehatan, pemerintah, jejaring sosial, pemasaran, keuangan). Ini

karena banyaknya tren teknologi, termasuk Internet of Things (Oussous, 2018). Keberadaan internet membuat jumlah data yang terdapat di website – website semakin bertambah. Ini dapat menjadi peluang bagi data engineer untuk mengolah banyaknya data tersebut menjadi statistik yang bermanfaat bagi

perusahaan atau peneliti. Namun dibutuhkan metode yang efisien untuk mengumpulkan banyaknya data yang ada di suatu website, dan web scraping termasuk salah satunya (Satriajati et al., 2021).

Penelitian yang menerapkan teknik web scraping pernah dilakukan oleh Salim Satriajati, Satria Bagus Panuntun, dan Setia Pramana pada tahun 2020 dengan melakukan web scraping pada situs berita yakni detik.com untuk mengumpulkan berita kriminal pada masa pandemi covid-19.

Web scraping sendiri adalah penggunaan alat teknologi untuk ekstraksi otomatis dan pengolahan data dari suatu web untuk tujuan analisis lebih lanjut dari data tersebut (Krotov, 2018; Satriajati et al., 2021). Salah satu alat scraping yang terkenal adalah BeautifulSoup, yaitu sebuah open source milik python yang berfungsi untuk mengurai dokumen HTML dan XML dari website dengan cepat dan sederhana (Hajba, 2018).

Sebagai target pengujian untuk implementasi web scraping, peneliti memilih website myanimelist.net yang merupakan komunitas dan database anime dan manga online paling aktif di dunia. Menurut similiarweb.com pada 1 Desember 2021, myanimelist.net menduduki peringkat 582 dunia dalam website dengan traffic tertinggi dan peringkat 10 dalam kategori Arts & Entertainment. Myanimelist memiliki banyak database seputar anime, mulai dari judul, tanggal rilis, status, pengisi suara hingga rating. Banyaknya data dalam website ini dapat kita ambil dan proses untuk menarik suatu kesimpulan.

Anime sendiri adalah animasi khas Jepang yang memiliki ciri pada gambar yang berwarna-warni yang menampilkan berbagai macam tokoh serta lokasi dari cerita yang ditunjukkan, anime dipengaruhi oleh gaya gambar manga komik khas Jepang (Purnamasari, 2018). Beriring perkembangan waktu, anime menjadi lebih terkenal karena perkembangan grafik yang lebih baik dan cerita yang beragam dari berbagai macam genre, contoh anime yang sedang naik daun sekarang adalah Shingeki no Kyojin / Attack on Titan dengan grafik yang sangat detail dan bergenre action fantasy, akan tetapi perkembangan anime tidak luput dari jasa seorang seiyuu / Seiyū (声優), bahasa jepang dari pengisi suara.

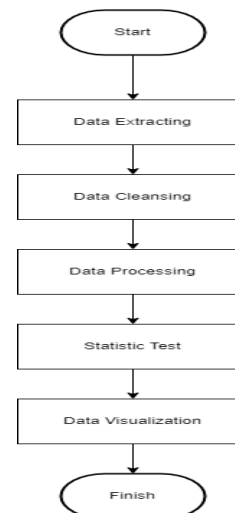
Seiyuu memiliki peran besar dalam industri anime karena beberapa dari mereka dapat menjiwai karakter yang diperankannya dengan baik, hal ini tentu dapat membuat penonton lebih nyaman dalam menikmati anime. Seiyuu berkontribusi ke pasar anime

melalui mempromosikan karakter yang mereka perankan dan serial animenya. Penelitian tentang hubungan antara seiyuu dengan anime pernah dilakukan oleh Ruby Zelka Heinst pada tahun 2017 lewat tesisnya yang berjudul “Seiyū: the art of voice acting the Japanese voice acting industry and its relation to the anime market”. Ruby Zelka Heinst menyimpulkan bahwa beberapa seiyuu memiliki fanbase nya sendiri yang membuat anime yang diperankannya mendapat banyak penonton, karena banyak orang mendasarkan pilihan anime mereka pada seiyuu dan apakah seiyuu favorit mereka muncul di dalamnya atau tidak.

Pada penelitian kali ini, penulis akan mencoba mengimplementasikan metode web-scraping pada myanimelist untuk mengekstrak data seputar seiyuu dan anime yang diperankannya, lalu membuat uji korelasi untuk mencari tahu ada atau tidaknya hubungan antara data tersebut.

2. METODOLOGI PENELITIAN

Penelitian ini secara garis besar terdiri dari 5 tahapan utama, seperti yang dijelaskan pada bagan alur penelitian dan alur proses berikut:



Gambar Flow chart

Gambar Flow chart adalah alur penelitian secara garis besar. Alur penelitian terdiri dari data extracting, data cleansing, data processing, statistic test, dan data visualization.

Data Extracting

Pengumpulan data pada penelitian ini menggunakan metode web scraping, yaitu pengambilan data atau informasi dari web internet secara otomatis sesuai tag html/css yang diinginkan (Krotov, 2018). Proses web scraping dilakukan selama 8 hari, yaitu pada

tanggal 22 Desember 2021 – 29 Desember 2021 sampai people urutan 450. Hasilnya, terdapat 275 nama, 54298 baris dan 9 kolom data yang terpisah menjadi 9 file csv.

Data Cleansing

Tahap ini adalah tahap pembersihan data, dimana data hasil scraping masih perlu dibersihkan karena beberapa alasan. Sebelumnya, peneliti menggunakan google colab dalam proses ini agar lebih mudah dalam memproses dataframe. Proses ini mengurangi jumlah data menjadi 45714 baris dan 208 nama.

Data Processing

Data yang telah bersih kemudian diproses terlebih dahulu sebelum melakukan pengujian. Pada proses ini, data dikelompokkan berdasarkan nama seiyyu dan nama karakter anime dengan mean/rata-rata sebagai acuan pengelompokan data angka. hal ini dilakukan untuk mencegah salah perhitungan rata-rata pada seiyyu yang memainkan karakter yang sama berkali-kali, sehingga memiliki nilai lebih tinggi. Dataframe setelah diproses memiliki 208 baris dan 6 kolom.

Statistic Test

Data yang telah diproses kemudian diuji dengan uji korelasi pearson, dengan rumus berikut:

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{\{n\sum x^2 - (\sum x)^2\} \{n\sum y^2 - (\sum y)^2\}}}$$

Gambar Rumus uji korelasi

Dalam pengujian statistic pada gambar 2 diatas, peneliti mencari nilai Koefisien korelasi yang menunjukkan kuat/tidaknya hubungan linier antar dua variabel. Proses uji korelasi akan dilakukan secara otomatis menggunakan library pandas pada google colab. Pedoman untuk memberikan interpretasi koefisien korelasi adalah sebagai berikut:

- 0,00 – 0,199 = Korelasi sangat rendah
- 0,20 – 0,399 = Korelasi rendah
- 0,40 – 0,599 = Korelasi sedang
- 0,60 – 0,799 = Korelasi kuat
- 0,80 – 1,000 = Korelasi sangat kuat (Sugiyono, 2013)

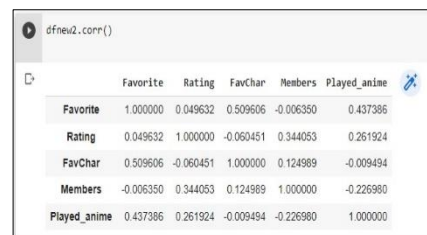
Data Visualization

Data visualization pada penelitian adalah proses menampilkan keseluruhan data yang telah diproses dan diuji dalam bentuk bar chart dan scatter plot berbasis aplikasi web. Pembuatan aplikasi ini menggunakan modul django application dari framework django,

open source milik python untuk pengembangan web. Dalam pembuatan chart dan scatter plot, peneliti menggunakan modul chart.js yang merupakan pustaka javascript yang bersifat open source untuk visualisasi data. Sedangkan dalam pembuatan desain web, peneliti menggunakan modul bootstrap dan w3.css yang juga merupakan open source untuk desain web.

3. HASIL DAN PEMBAHASAN

Peneliti menggunakan google colab untuk melakukan uji korelasi pearson pada dataframe, dimana kolom jumlah favorite seiyyu akan menjadi X dan 4 kolom lainnya sebagai Y.



Gambar Proses uji korelasi menggunakan colab

Gambar Scatter plot dari data jumlah favorite seiyyu dan jumlah favorite karakter

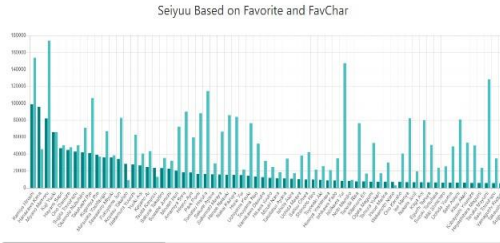
Tabel Hasil analisis korelasi hubungan popularitas seiyyu dengan anime

Tabel hasil uji korelasi			
Variabel X	Variabel Y	Nilai koefisien korelasi	Kekuatan korelasi
Jumlah Favorite Seiyyu	Rata-rata rating	0.049632	sangat rendah
Jumlah Favorite Seiyyu	Rata-rata jumlah favorite karakter	0.509606	sedang
Jumlah Favorite Seiyyu	Rata-rata jumlah member	-0.00635	sangat rendah
Jumlah Favorite Seiyyu	Jumlah anime yang dimainkan	0.437386	sedang

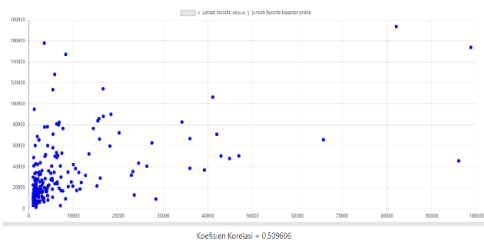
Dari tabel 1 diatas, menurut sugiyono (2013), dapat disimpulkan:

1. Jumlah favorite seiyyu dan rating anime memiliki korelasi sangat rendah
 2. Jumlah favorite seiyyu dan jumlah favorite karakternya memiliki korelasi sedang
 3. Jumlah favorite seiyyu dan jumlah member memiliki korelasi sangat rendah
- Jumlah favorite seiyyu dan jumlah anime yang dimainkan memiliki korelasi sedang.

Peneliti kemudian menampilkan data hasil uji statistik dalam bentuk bar chart dan scatter plot melalui aplikasi web yang telah peneliti buat sebelumnya. Peneliti hanya tinggal memasukkan file SeiyuuData.csv ke dalam database web tersebut.



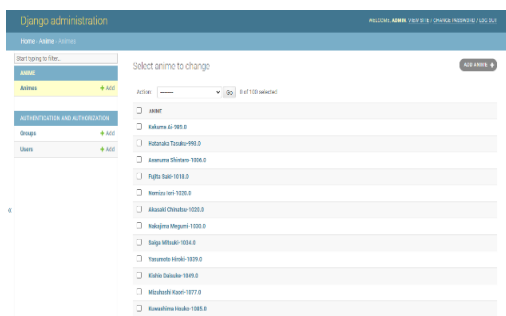
Gambar. Bar Chart dari data jumlah favorite seiyuu dan jumlah favorite karakter



Gambar Scatter plot dari data jumlah favorite seiyuu dan jumlah favorite karakter

Name	Favorite	Rating	FavChar	Members	Played Anime
Kamiya Hitoshi	98732.0	70225.20463320446	133879.1891891892	150941.81384189885	37200
Hayashida Kana	95984.0	68470.70053076184	43050.46258050401	180540.5913763428	43600
Miyano Mamoru	82853.0	70669.88385800865	137371.25	215028.62809117048	31700
Kaji Yuuki	65791.0	69883.2138646022	66825.87320574415	144434.388899068	39400
Hayami Sora	46927.0	69334.982078853	50338.40215053734	227517.61871146984	30700
Ota Daisuke	44844.0	61907.78514569863	47918.44640234948	191097.2192200636	40700
Sigita Tomokazu	42933.0	69596.8321878952	50323.665488427844	209427.14188273175	45400
Okamoto Nobuhiko	42006.0	69890.37146738999	71951.4519883041	232552.87420613825	29500
Takahashi Rie	41106.0	64379.96571342438	10436.98620130988	181889.96932054795	12200
Kaguya Rie	39227.0	69128.60596020709	38955.0064347162	173863.01992788742	49200
Matsuka Yoshinaga	35992.0	67474.63792617836	66859.76331360947	220092.07059030712	28000
Sawashiro Miyuki	34977.0	68733.91539272336	38952.73062730627	171373.1792552588	50100
Fukuyama Jun	34224.0	68894.2238947921	82672.31404958678	175367.68810507673	45400
Kinoshita Takahito	28395.0	69285.22000000006	9187.33333333334	183962.64844444445	57100

Gambar. Tabel keseluruhan data



Gambar 7. Halaman admin

4. SIMPULAN

Berdasarkan hasil analisis dan pembahasan, peneliti mendapatkan kesimpulan dari penelitian analisis hubungan popularitas seiyuu dengan anime menggunakan metode

pengambilan data web scraping pada situs myanimelist.net sebagai berikut:

1. Metode web scraping dapat diimplementasikan pada situs myanimelist.net. Peneliti dapat mengumpulkan 54298 baris data dari 9 halaman dan 275 orang dalam waktu 8 hari pada situs tersebut. Data yang di proses sejumlah 208 people dari total 450 people, yaitu sekitar 46% sedangkan data yang dilewatkan dan dihapus sekitar 56%.
2. Hasil uji korelasi menunjukkan bahwa terdapat 2 variabel yang memiliki hubungan korelasi dengan popularitas seiyuu, yaitu jumlah favorite karakter anime dan jumlah anime yang dimainkan seiyuu.
3. korelasi dari hubungan popularitas seiyuu dengan 2 variabel tersebut menunjukkan nilai sebagai berikut:
 - a. Jumlah favorite seiyuu dan jumlah favorite karakter anime: 0.509606
 - b. Jumlah favorite seiyuu dan jumlah favorite karakter anime: 0.4373

Sehingga dapat disimpulkan bahwa popularitas seiyuu memiliki kekuatan hubungan korelasi sedang pada popularitas karakter anime yang diperankannya dan banyaknya anime yang diperankan.

Saran

Berdasarkan hasil penelitian diatas, maka saran dan masukan yang dapat diberikan adalah sebagai berikut:

1. Bagi pembaca, hasil penelitian ini dapat menambah pengetahuan seputar web-scraping, khususnya yang tertarik dengan teknik pengambilan data berskala besar secara otomatis dari suatu web tanpa harus melakukannya secara manual yang memakan banyak tenaga dan waktu.
2. Bagi animator, hasil penelitian ini menunjukkan bahwa popularitas seorang pengisi suara memiliki hubungan sedang dengan popularitas karakter yang diperankan. Hal ini bisa menjadi acuan untuk memilih pengisi suara yang populer dan berpengalaman untuk mengisi suara karakter-karakter animasinya guna menaikkan popularitasnya.
3. Bagi penelitian selanjutnya, proses penggunaan database dalam aplikasi web masih menggunakan metode manual, yaitu dengan mengimpor data csv menjadi sqlite. Proses ini dapat dikembangkan dengan otomatis, dimana hasil scraping dapat langsung menjadi database.

5. DAFTAR PUSTAKA

- Hajba, G. L. (2018). Using beautiful soup. In Website Scraping with Python (pp. 41–96). Apress, Berkeley, CA. https://doi.org/https://doi.org/10.1007/978-1-4842-3925-4_3
- Heinst, R. Z. (2017). Seiyū: the art of voice acting the Japanese voice acting industry and its relation to the anime market. Doctoral Dissertation, Leiden University.
- Krotov, V. . & T. M. (2018). Research note: scraping financial data from the web using the R language. Journal of Emerging Technologies in Accounting, 15(1), 169-181.
- Purnamasari, Z. (2018). ANALISIS SEMIOTIKA PADA FILM ANIME KIMI NO NA WA . Doctoral Dissertation, PERPUSTAKAAN.
- Neny Rosmawarni, dkk; E-COMMERCE, Editor: Ingrid Yanuar Risca Pratiwi, S.S.T., M.Tr.T., ISBN: 978-623-88989-5-4, Penamuda Media, 2024
- Python Software Foundation. (2021). beautifulsoup4 . PyPI. Pypi.Org. <https://pypi.org/project/beautifulsoup4/>
- Python Software Foundation. (2022). About PythonTM | Python.org. Python.Org. <https://www.python.org/about/>
- Ramadhan, W. (2021, April 2). Ini Dia Sejarah Anime Hingga Sampai ke Indonesia.Kreativv. <https://kreativv.com/sejarah-anime-indonesia/>
- Satriajati, S., Panuntun, S. B., & Pramana,S.(2021). IMPLEMENTASI WEB SCRAPING DALAM PENGUMPULAN BERITA KRIMINAL PADA MASA PANDEMI COVID-19. Seminar Nasional Official Statistics, 2020(1), 300–308. <https://doi.org/10.34123/semnasoffstat.v2020i1.578>
- Sugiyono. (2013). Metode penelitian pendidikan pendekatan kuantitatif, kualitatif dan R&D. Alfabeta
- Sukardi. (2021). Metodologi Penelitian Pendidikan: Kompetensi Dan Praktiknya (Edisi Revisi). Bumi Aksara.
- Tokyo Otaku Mode. (2015). Exclusive Interview with the Founder of MyAnimeList, a Colossal Site with 120 Million Monthly PVs. Tokyo Otaku Mode News. [https://otakumode.com/news/5590fbd763cd06585662ce9d/Exclusive-Interview-](https://otakumode.com/news/5590fbd763cd06585662ce9d/Exclusive-Interview-with-the-Founder-of-MyAnimeList-a-Colossal-Site-with-120-Million-Monthly-PVs)
- Acai Sudirman, dkk; BISNIS INFORMASI, ISBN : 9786235002224 Widina Media Utama, 2024
- Van Baaren, E. (2021). Python History. Python Land . <https://python.land/python-tutorial/python-history/>