

# Analisis Data Hasil Web Scraping untuk Menentukan Kualitas Jurnal Ilmiah

Lucky Metha Purnomo<sup>#1</sup>, Mewati Ayub<sup>\*2</sup>

#Program Studi Teknik Informatika Universitas Kristen Maranatha  
Jln. Surya Sumantri. No. 65 Bandung  
\*1472065@maranatha.ac.id  
\*mewati.ayub@it.maranatha.edu

**Abstract** — Portal of Science Journal has many journals or scientific papers with many subjects. In other words, there are many searchable journals and it takes time to search and manage. Web Scraping is a process of taking some documents from a website to other needs. The result of web scraping is managed as csv file. For this study, web scraping data was taken from SINTA and SCIMAGOJR portals. Data file from web scraping is given to an application written in Python. The application is equipped with features to display and search journal data.

**Keywords**— journal, portal, web scraping

## I. PENDAHULUAN

Perkembangan teknologi saat ini sangat pesat, hampir semua segi kehidupan tidak terlepas dari penggunaan teknologi. Terdapat banyak sekali portal pengindeks hasil karya publikasi ilmiah baik itu berupa jurnal, artikel yang diterbitkan dalam prosiding dari seminar ilmiah, maupun artikel ilmiah lainnya. Sebut saja yang paling banyak digunakan yaitu Google Scholar, lalu portal lainnya seperti *Science and Technology Index* (SINTA), Portal Garuda, *Indonesia Science and Technology Index* (InaSTI), *Indonesian Publication Index* (IPI), dan yang lainnya.

Dalam pendidikan tingkat tinggi di Universitas, dosen berperan sebagai tenaga pendidik. Di samping itu, dosen juga harus melakukan penelitian sesuai bidang ilmunya dan hasil penelitian harus dipublikasikan. Tentu saja dosen memberikan kontribusi yang besar dalam meningkatkan indeks publikasi ilmiah untuk Indonesia [1]. Apalagi dalam masa pandemi seperti saat ini, pencarian publikasi ilmiah *online* menjadi lebih tinggi karena hampir semua dosen melakukan studi secara daring yang tidak memungkinkan untuk mencari jurnal di perpustakaan.

Ketika dosen melakukan publikasi hasil penelitian, kadangkala terdapat kesulitan dalam mencari jurnal yang mempunyai peringkat baik. Mengingat begitu banyaknya publikasi ilmiah yang ada di portal pengindeks, tentu saja dosen memerlukan bantuan untuk memperoleh informasi peringkat jurnal sesuai bidang ilmu dari publikasi ilmiah tersebut, juga pada saat dosen akan mempublikasikan hasil penelitiannya.

Oleh karena hal tersebut, diperlukan program yang dapat membantu dosen menentukan tingkatan publikasi ilmiah yang ada di dalam portal, ataupun memilah hasil publikasi ilmiah berdasarkan bidang ilmu. Program yang dikembangkan dalam makalah ini didasarkan pada teknik *web scraping* untuk mengekstrak data publikasi ilmiah yang akan disimpan dalam bentuk file untuk dianalisis lebih lanjut. Teknik web scraping telah banyak digunakan oleh peneliti lain untuk berbagai keperluan, seperti analisis topik penelitian artikel ilmiah [2], analisis deskripsi produk [3], analisis situs wikipedia [4], dan untuk marketplace [5].

Tujuan penelitian yang dibahas dalam makalah ini adalah untuk mengembangkan aplikasi yang dapat menampilkan informasi jurnal ilmiah dengan memilah bidang studi dan tingkatan jurnal dari *portal* pengindeks SINTA dan SCImagoJR. Pengambilan data dari portal pengindeks dilakukan dengan cara *web scraping*.

## II. LANDASAN TEORI

### A. Publikasi Ilmiah

Publikasi ilmiah merupakan luaran dari suatu penelitian yang telah selesai dilaksanakan. Publikasi ilmiah dapat berupa artikel di dalam jurnal ilmiah atau prosiding, atau berupa buku. Dengan melakukan publikasi ilmiah, hasil temuan dari suatu penelitian dapat diketahui oleh masyarakat sehingga mendapatkan pengakuan dari sejawat dalam bidang yang sama. Publikasi ilmiah memiliki peran penting untuk menunjukkan kemajuan suatu negara dalam ilmu pengetahuan dan teknologi,

sehingga pemerintah meminta perguruan tinggi untuk mewajibkan calon lulusannya melakukan publikasi ilmiah di akhir masa studinya [1].

#### B. Jurnal

Jurnal ilmiah sering pula disebut jurnal akademik, atau '*scientific journal*' atau '*academic journal*'. Jurnal akademik dapat dinyatakan sebagai kumpulan artikel ilmiah yang dipublikasikan secara reguler dalam rangka mendiseminasi hasil penelitian. Mutu jurnal ilmiah dinilai secara berkala oleh lembaga penilai untuk menunjukkan kualitas artikel yang dipublikasikan [1] [2].

#### C. Science and Technology Index(SINTA)

Kementerian Riset dan Teknologi atau biasa disingkat Kemenristekbrin mengembangkan sebuah aplikasi berbasis web bernama SINTA mulai tahun 2016. SINTA berisi pengukuran kinerja Ilmu Pengetahuan dan Teknologi yang meliputi kinerja peneliti, jurnal, institusi, dan penulis artikel ilmiah. SINTA juga menjadi alat pengindeks internasional sebagai arsip jurnal, buku, artikel, dan karya ilmiah lainnya dari institusi ilmiah di Indonesia [6]. Kehadiran SINTA diharapkan dapat memotivasi para peneliti maupun dosen untuk lebih berinovasi dalam pengembangan ilmu pengetahuan dan teknologi di Indonesia..

#### D. Scimago Journal and Country Rank

*Scimago* merupakan portal publik yang mencakup jurnal dan indikator ilmiah yang dikembangkan dari informasi yang terdapat dalam pangkalan data *Scopus* dari penerbit *Elsevier*. Indikator *SJR* dapat digunakan untuk menilai mutu dan menganalisis domain ilmiah. Peringkat jurnal dapat dibandingkan atau dianalisis secara terpisah, selain itu peringkat negara juga dapat dibandingkan atau dianalisis secara terpisah berdasarkan publikasi ilmiah dari setiap negara[7].

#### E. Web Scraping

Dalam dunia *online*, diperlukan data yang berasal dari sumber-sumber yang tersedia di dunia virtual. Data yang berasal dari suatu situs web dapat diambil dan disimpan di dalam sebuah *spreadsheet*, baik itu menggunakan *Microsoft Excel*, *Google Sheet* atau aplikasi sejenisnya. Proses pengambilan data dari suatu situs web inilah yang disebut sebagai *web scraping* [8],[9],[10].

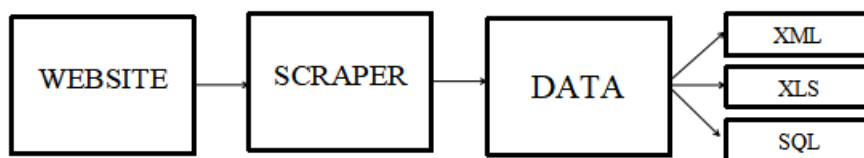
#### F. Python

Bahasa pemrograman Python pertama kali diciptakan oleh *Guidovan Rossum* pada awal tahun 1990 di negeri Belanda sebagai pengganti bahasa pemrograman yang disebut ABC. Walaupun *Guido* adalah orang yang pertama kali menciptakan bahasa pemrograman ini, tetapi bahasa pemrograman Python yang digunakan sekarang merupakan kontribusi dari berbagai sumber. Bahasa pemrograman Python merupakan bahasa pemrograman yang dapat dikembangkan oleh siapa saja karena bersifat *Open Source* [11]. Kelebihan dari Python adalah library yang tersedia cukup lengkap, terutama untuk menganalisis data teks.

### III. ANALISIS, RANCANGAN DAN IMPLEMENTASI

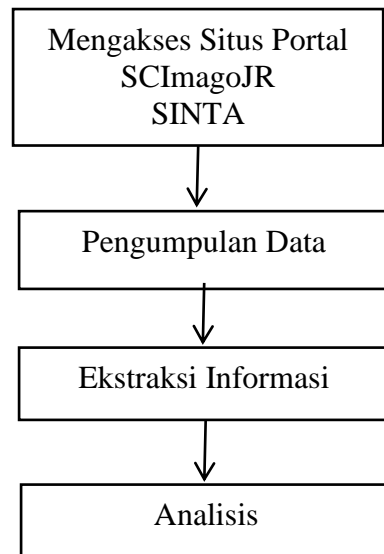
#### A. Alur Kerja Web Scraping

Proses *scraping* dimulai dari pemindaian *website*. Dalam proses ini, kita dapat melihat data dari suatu *website* dan melakukan *scraping* dengan *tools scraper* yang akan digunakan. Hasil *scraping* berupa data yang dapat berformat .xml, .xls, .csv atau pun bisa dimasukkan ke dalam *database*, seperti pada Gambar 1.



Gambar 1 Alur Kerja *Web Scraping*

Proses *scraping* pada portal SINTA atau Scimago dimulai dari membuka situs <https://sinta.ristekbrin.go.id/>, kemudian mengumpulkan data jurnal dengan memakai perangkat *scraper* seperti pada Gambar 2. Setelah melakukan proses *scraping* dengan *tools scraping*, maka data tersebut dapat diolah sesuai keperluan [2][3][4].



Gambar 2 Alur Pengambilan Data pada SINTA atau Scimago [2]

Proses *scraping* pada situs [www.scimagojr.org](http://www.scimagojr.org) dilakukan dengan cara yang sama seperti pada gambar 2, hanya saja di situs web yang berbeda.

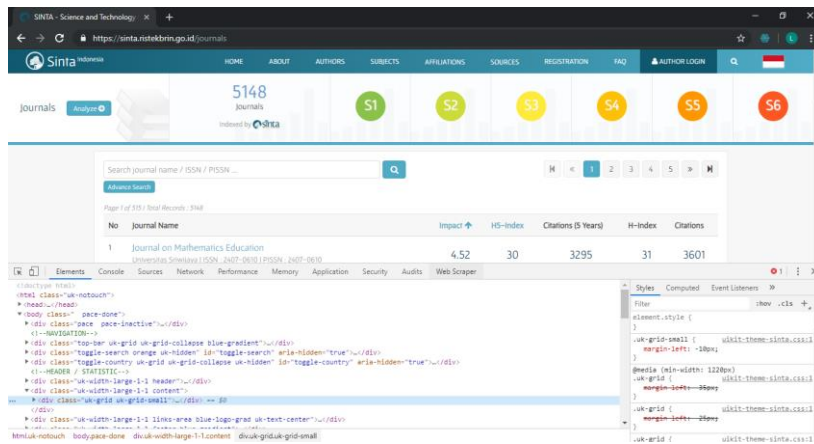
#### B. Alur Kerja Web Scraper

Gambar 3 adalah proses membuka website dari [sinta.ristekbrin.go.id](http://sinta.ristekbrin.go.id), lalu memilih menu journal untuk mengumpulkan data jurnal.



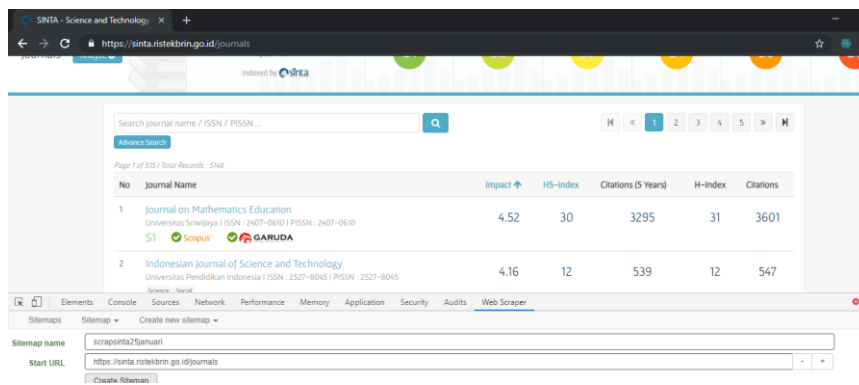
Gambar 3 Situs web SINTA

Gambar 4 adalah proses *scraping* dengan *tools* yang diekstensikan ke dalam *Google Chrome*. *Tools* tersebut dapat digunakan dengan melakukan *inspect element* pada *Google Chrome*.



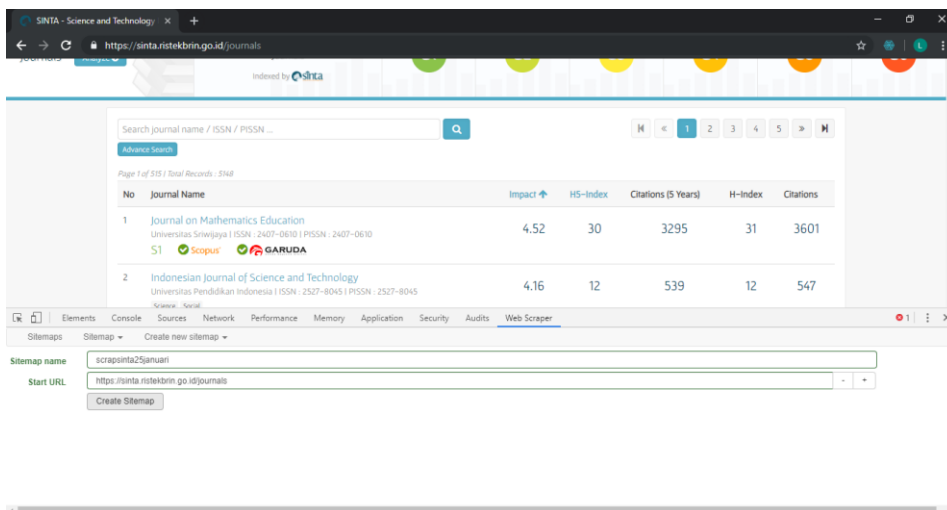
Gambar 4 Proses *scraping* pada web SINTA

Gambar 5 adalah proses ekstraksi informasi dengan menentukan *selector* mana saja yang diperlukan untuk melakukan pengumpulan data dari situs *portal*. Dimulai dari memilih *link-link* jurnal yang terindeks oleh *portal*, lalu memilih *detail* dari jurnal tersebut seperti judul, penerbit, dan data lainnya.



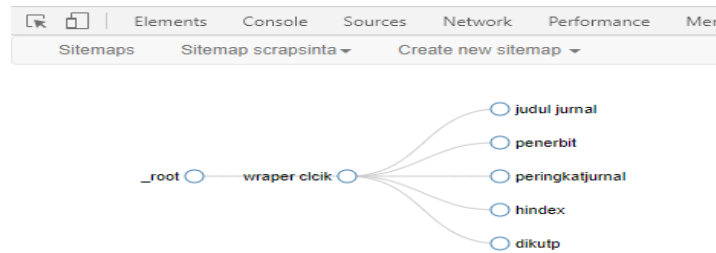
Gambar 5 Proses Ekstraksi data

Gambar 6 adalah proses yang masih berhubungan dengan gambar 5 yaitu proses pembuatan *sitemap* yang berfungsi untuk menggambarkan *graph* dari proses *scraping*.



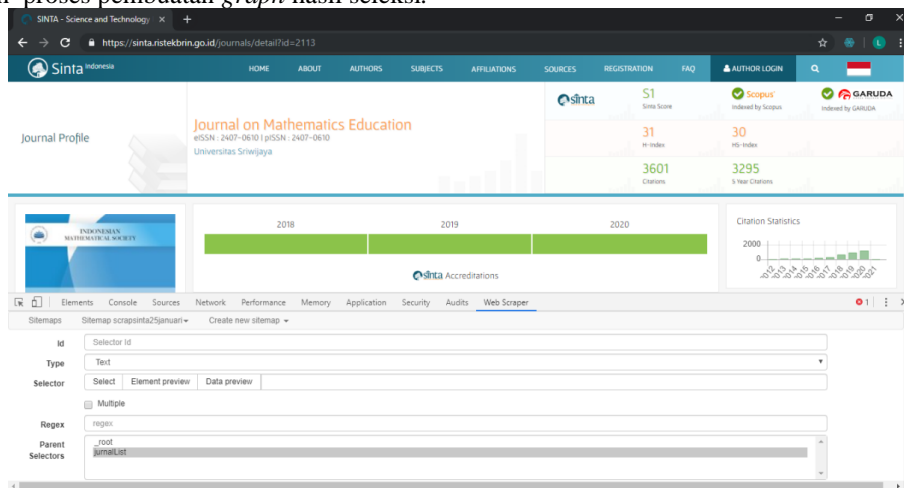
Gambar 6 Membuat *sitemap*

Gambar 7 adalah proses yang masih berhubungan dengan gambar 6 yaitu proses pembuatan *selector link* yang akan dimasukkan ke dalam *graph* dari proses *scraping*.



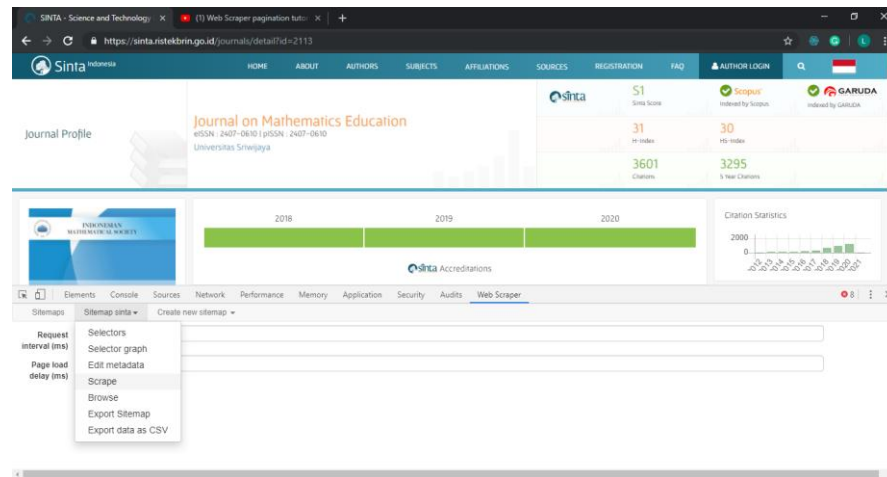
Gambar 7 Pembuatan *selector link*

Gambar 8 adalah proses pembuatan *graph* hasil seleksi.



Gambar 8 Pembuatan *graph* hasil seleksi

Pada Gambar 9, setelah ditentukan *selector* maka akan dilakukan *scraping*. Durasi dari proses *scrape* ini tergantung dari berapa banyak halaman dan jurnal yang akan di-*scrape*. Setelah proses *scrape* selesai, maka akan diperoleh file *.csv* yang berisi data hasil *scraping*.



Gambar 9 Proses *scraping*

### C. Tabel Hasil Analisis

Tabel I dan Tabel II adalah tabel yang menunjukkan hasil *scraping* data yang selanjutnya akan ditampilkan oleh program aplikasi. Pada Tabel I terdapat kolom yang berisi judul jurnal, penerbit, peringkat dari SINTA, berapa kali jurnal tersebut dikutip oleh pengguna, dan H-index dari jurnal tersebut. Pada Tabel II terdapat kolom judul jurnal, *SJR* atau nilai *SCIMAGO Journal Rank* atau peringkat dari *SCIMAGO*.

TABEL I  
CONTOH HASIL DATA SCRAPING SINTA

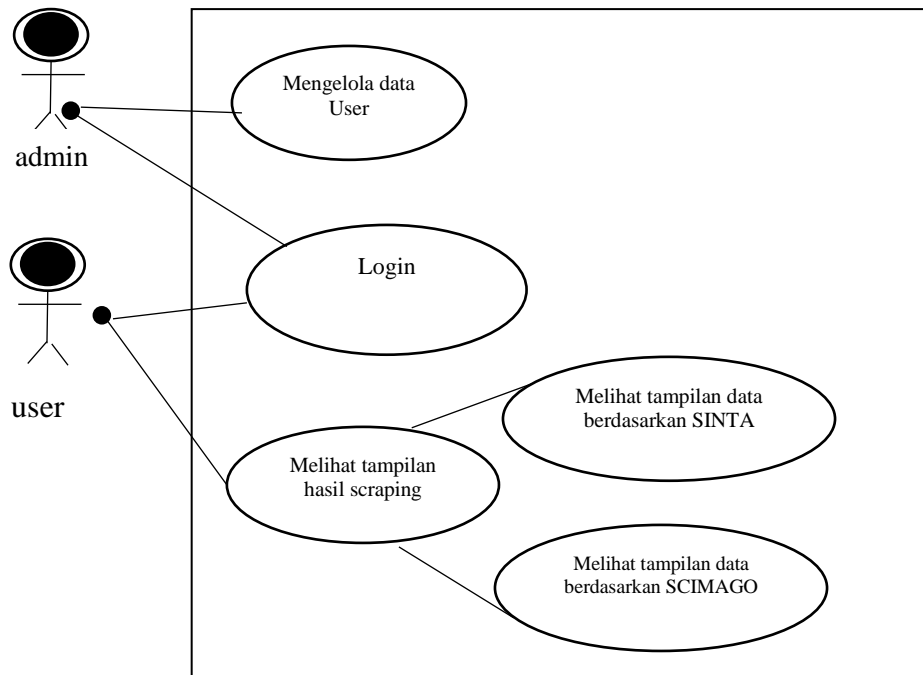
NO	Judul Jurnal	Penerbit	Peringkat Jurnal	Dikutip	H-Index
1	International Journal Of Electrical and Computer Engineering	Universitas Pendidikan Indonesia	S1	7093	30
2	International Journal on Advanced Science, Engineering and Information Technology (IJASEIT)	Indonesia Society Of Journals	S1	3525	21
3	Informatika Pertanian	Universitas Sriwijaya	S2	1812	15
4	CESS (Journal of Computer Engineering, System and Science)	Universitas Brawijaya	S3	258	12
5	SIMETRIS : JURNAL TEKNIK MESIN, ELEKTRO DAN ILMU KOMPUTER	Sekolah Tinggi Teknik Bandung	S3	112	11

TABEL II  
CONTOH HASIL DATA SCRAPING SCIMAGO

NO	Judul Jurnal	Penerbit	SJR	Quartile	H-Index
1	International Journal Of Electrical and Computer Engineering	Universitas Pendidikan Indonesia	14,367	Q1	30

<b>NO</b>	<b>Judul Jurnal</b>	<b>Penerbit</b>	<b>SJR</b>	<b>Quartile</b>	<b>H-Index</b>
2	International Journal on Advanced Science, Engineering and Information Technology (IJASEIT)	Indonesia Society Of Journals	11,221	Q1	21
3	Informatika Pertanian	Universitas Sriwijaya	8,541	Q3	15
4	CESS (Journal of Computer Engineering, System and Science)	Universitas Brawijaya	8,199	Q3	12
5	SIMETRIS : JURNAL TEKNIK MESIN, ELEKTRO DAN ILMU KOMPUTER	Sekolah Tinggi Teknik Bandung	8,021	Q3	11

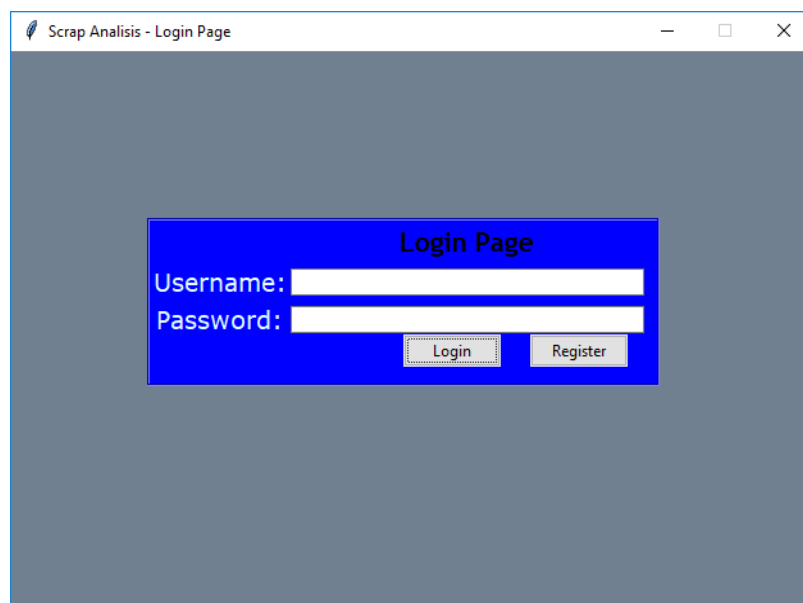
Gambar 10 adalah *use case* diagram untuk program aplikasi yang melakukan analisis terhadap data hasil *scraping* [12]. Admin dapat melakukan Login dan dapat mengelola data user. User dapat melakukan Login dan dapat melihat atau mencari data hasil *scraping* dari portal SINTA dan SCIMAGOJR.



Gambar 10 Diagram *Use case* aplikasi

#### D. Implementasi Program

Gambar 11 adalah halaman *login page* untuk aplikasi. Login harus dilakukan setelah mengisi *username* dan *password*. Jika belum terdaftar, user dapat memilih Register untuk mendaftarkan sebagai user baru. Kode program dapat dilihat pada kode program 1.



Gambar 11 Gambar tampilan login



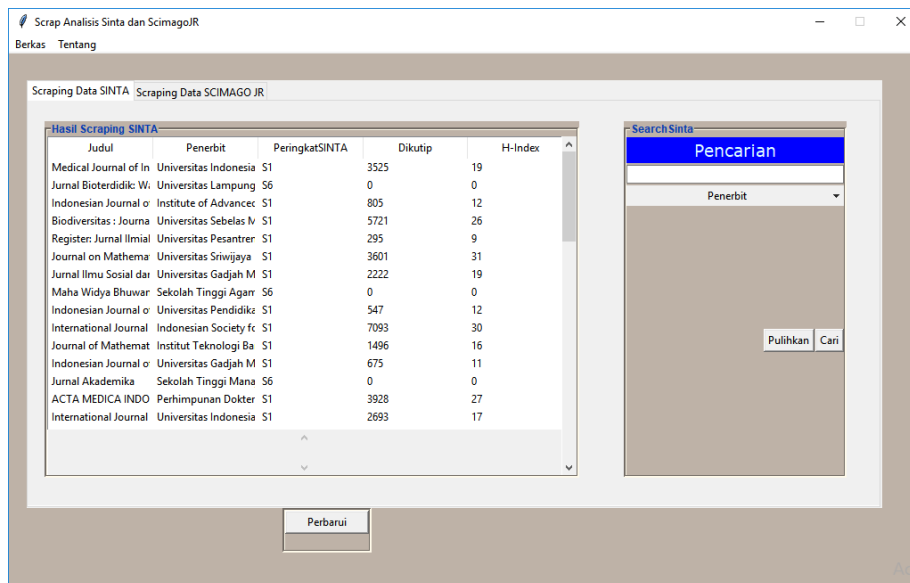
```

def getlogin():
    Database()
    if entry_user.get() == "" or entry_pw.get() == "":
        tk.messagebox.showerror("Information",
            "Username dan password tidak boleh kosong ")
    else:
        cursor.execute("SELECT * FROM `user` WHERE
            `username` = ? AND `password` = ?",
            (entry_user.get(), entry_pw.get()))
        if entry_user.get() == "admin" or entry_pw.get()
            == "admin":
            top.destroy()
            MyAppAdmin()
        elif (entry_user.get() != "admin" or
            entry_pw.get() != "admin")and
            cursor.fetchone() is not None:
            top.destroy()
            MyApp()
        elif (entry_user.get() != "admin" or
            entry_pw.get() != "admin")and cursor.fetchone()
            is None:
            tk.messagebox.showerror("Information", "The
                Username or Password you have entered are
                incorrect ")
        cursor.close()

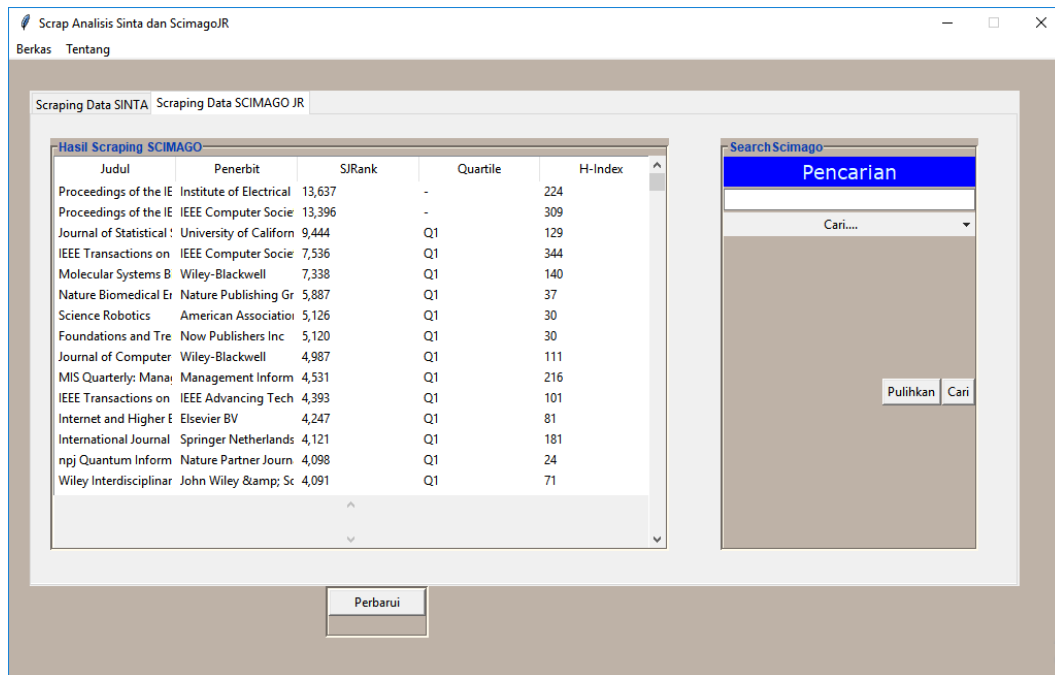
```

Kode Program 1 Kode program untuk login

Gambar 12 dan gambar 13 adalah halaman awal *role* User untuk aplikasi. Terdapat 2 buah *tab* yang akan menampilkan data hasil *scraping*. Data dari portal SINTA di *tab* 1, dan dari portal SCIMAGOJR di *tab*2. User juga dapat melakukan pencarian berdasarkan *ranking* yang diinginkan. Kode program untuk menampilkan data *tree view* untuk user dapat dilihat pada Kode Program 2, sedangkan kode untuk mengambil data dari *database* pada kode program 3.



Gambar 12 Tampilan awal untuk role User pada *tab* SINTA



Gambar 13 Tampilan Awal untuk *role* User pada *tab* SCIMAGO

```

tv11 = ttk.Treeview(frame1)
    column_list_account =
["Judul", "Penerbit", "PeringkatSINTA", "Dikutip", "H-Index"]
    tv11['columns'] = column_list_account
    tv11["show"] = "headings"
    for column in column_list_account:
        tv11.heading(column, text=column)
        tv11.column(column, width=50)
    tv11.place(relheight=1, relwidth=0.995)
    treescrollY = tk.Scrollbar(frame1)
    treescrollY.configure(command=tv11.yview)
    tv11.configure(yscrollcommand=treescrollY.set)
    treescrollY.pack(side="right", fill="y")

```

Kode Program 2 Kode program untuk tampilan *tree view* user

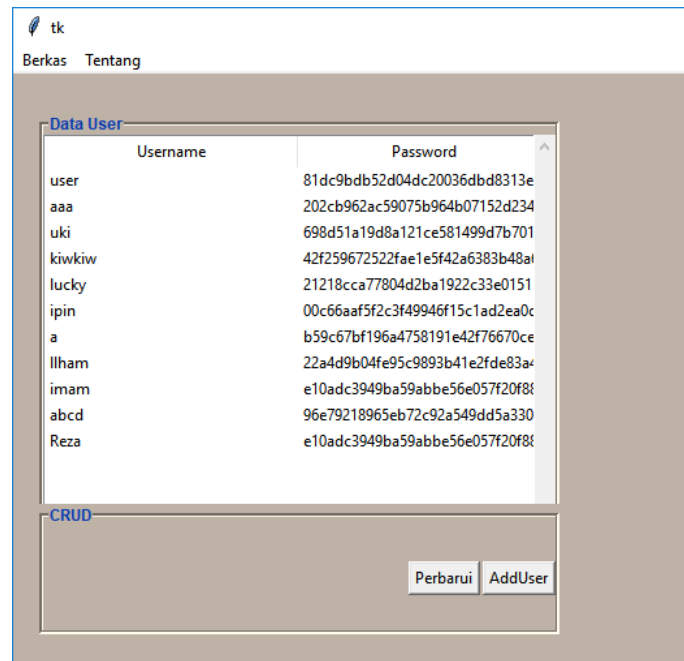
```

cursor.execute("SELECT
judul,instansi,peringkat,dikutip,hindex FROM
`sinta`)
    conn.commit()
    fetch = cursor.fetchall()
    for data in fetch:
        tv11.insert('', 'end',
values=(data[0], data[1],data[2],data[3],data[4]))

```

Kode Program 3 Kode program untuk mengambil data dari *database*

Gambar 14 adalah halaman awal Admin untuk aplikasi. Admin dapat melihat data *user* dan *password* yang sudah terenkripsi dengan md5 untuk menjamin keamanan *data*. Admin juga dapat menambahkan *user* baru.



Gambar 14 Tampilan Halaman Awal Admin

#### IV. KESIMPULAN

Dalam penelitian ini telah dikembangkan program aplikasi yang dapat menampilkan data jurnal hasil *scraping* yang berupa file *.csv* menjadi bentuk *tree view*. Program juga dilengkapi fitur pencarian yang dapat digunakan *user* untuk mengklasifikasikan dan mencari jurnal yang diperlukan. Keterbatasan program aplikasi, hanya dapat menampilkan jurnal untuk bidang studi *computer science*. Untuk antarmuka aplikasi, program menggunakan *GUI Python Tkinter*, sehingga antarmuka masih bersifat statis.

#### DAFTAR PUSTAKA

- [1] Lukman, S. S. Ahmadi, W. Manalu, D. S. Hidayat, Pedoman Publikasi Ilmiah 2019, Kementerian Riset, Teknologi, dan Pendidikan Tinggi, 2019.
- [2] Y. Sahria, Implementasi Teknik Web Scraping pada Jurnal SINTA Untuk Analisis Topik Penelitian Kesehatan Indonesia. *The 11th University Research Colloquium 2020 Universitas Aisyiah Yogyakarta*, 2020.
- [3] D. F. Setiawan, Tristiyanto, A. Hijriani. Aplikasi Web Scraping Deskripsi Produk, *Jurnal TEKNOINFO*, Vol. 14, No. 1, 41-47, 2020.
- [4] M.S. Utomo, Web Scraping pada Situs Wikipedia menggunakan Metode Ekspresi Regular. *Dinamik-Jurnal Teknologi Informasi*, vol.18, 2013.
- [5] A. Maulana, A. Susanto, D. Purwanti. Rancang Bangun Web Scraping Pada Marketplace di Indonesia, *Journal of Information System*, Vol. 4, No. 1, 2019.
- [6] (2020) Website SINTA. [Online]. Tersedia : <https://sinta.ristekbrin.go.id/about>
- [7] (2020) Website SJR. [Online]. Tersedia : <https://www.scimagojr.com/aboutus.php>
- [8] The Computer Advisor. (2014) Web site scraper the most effective tool for web data extraction. [Online]. Tersedia : <http://www.thecomputeradvisor.net/web-sitescraper-the-most-effective-tool-for-web-dataextraction/>
- [9] M. Turland, *PHP architect's Guide to Web Scraping with PHP. Introduction-Web Scraping*. Marco Tabini & Associates, Canada. 2010.
- [10] F. Lindenberg. (2020) Getting Data from the Web. [Online]. Tersedia : [http://data.journalismhandbook.org/1.0/en/getting\\_data\\_3.html](http://data.journalismhandbook.org/1.0/en/getting_data_3.html)
- [11] D.Amos. (2020) Python GUI Programming With Tkinter. [Online]. Tersedia : <https://realpython.com/python-gui-tkinter/>
- [12] A. Nugroho, *Rekayasa Perangkat Lunak Menggunakan UML*, Yogyakarta : ANDI, 2009.