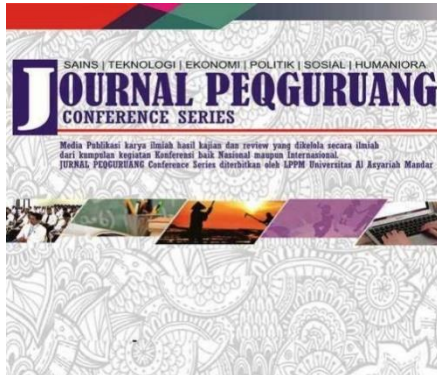


Graphical abstract



WEB SCRAPING DAN TEXT MINING UNTUK PENGELOMPOKAN INFORMASI PERGURUAN TINGGI DI SULAWESI BARAT

^{1*}Muh. Rijal Nur, ²Muh. Rusli Said, ³Idhan Zaldy
¹Sistem Informasi, ²Universitas Al Asyariah Mandar

muhrijaljanuari@gmail.com

Abstract

The goal to be achieved in this study is to design an information system for grouping higher education information, namely a system that can be a bridge between universities and prospective students. The development of this application uses web scraping and text mining methods in collecting university data, where data in the form of titles and article links are taken using web scraping and then calculating the amount of data using text mining. Based on the results of research conducted from the beginning of the application of the information system to the testing process, it can be concluded that this system can help prospective students / students in finding higher education information based on predetermined categories such as campus facilities, accreditation, study program faculties, tuition fees, alumni can be seen from the website portal.

Keywords: Website, College, *Text mining*, *Web scraping*

Abstrak

Tujuan yang akan di capai dalam studi ini yaitu merancang sebuah sistem informasi pengelompokan informasi Perguruan Tinggi, yaitu sistem yang dapat menjadi jembatan antara perguruan tinggi dan calon mahasiswa/mahasiswi. Pengembangan aplikasi ini menggunakan metode *web scraping* dan *text mining* dalam pengumpulan data Perguruan Tinggi, di mana data berupa judul dan link artikel di ambil menggunakan *web scraping* kemudian menghitung jumlah data dengan *text mining*. Berdasarkan hasil penelitian yang di lakukan dari awal penerapan sistem informasi hingga proses pengujian, maka dapat di simpulkan bahwa sistem ini dapat membantu calon mahasiswa/mahasiswi dalam mencari informasi perguruan tinggi berdasarkan kategori yang telah di tentukan seperti fasilitas kampus, akreditasi, fakultas prodi, biaya kuliah, alumni dapat di lihat dari portal *website*.

Kata kunci: . *Website*, Perguruan Tinggi, *Text mining*, *Web scraping*

Article history

DOI: [10.35329/jp.v5i2.3457](https://doi.org/10.35329/jp.v5i2.3457)

Received : 01/08/2022 | Received in revised form : 01/08/2022 | Accepted : 30/11/2023

1. PENDAHULUAN

Saat ini banyak bermunculan perguruan tinggi baru di Indonesia, baik perguruan tinggi swasta (PTS) maupun perguruan tinggi negeri (PTN). Jika dibandingkan dengan jumlah PTS di Indonesia, mencapai 3.154 PTS. (Tunggal and Budi 2020). Dalam hal perkembangan ilmu pengetahuan dan teknologi, sistem informasi memiliki pengaruh yang besar dalam berbagai bidang. Adanya sistem informasi memudahkan dalam pengelolaan data dan menghemat waktu, ruang dan biaya (Muh. Amin, Idhan Zaldy and Ul Khairat 2021)

Perguruan tinggi adalah satuan pendidikan yang menyelenggarakan pendidikan tinggi, jenjang pendidikan lanjutan dari pendidikan sebelumnya. Mereka yang ingin melanjutkan pendidikan ke perguruan tinggi harus melewati jenjang pendidikan sebelumnya seperti TK, SD, SMP, dan SMA. (Hasanah 2019).

Seiring dengan merambahnya internet di masyarakat, berbagai informasi menjadi mungkin untuk diperoleh dengan mudah dalam waktu yang relatif singkat. Banyaknya sumber informasi yang ada memberikan manfaat, di satu sisi, karena satu sumber dapat melengkapi informasi yang disajikan oleh sumber lainnya. Namun, di sisi lain, jumlah dan variasi informasi yang besar menyebabkan informasi yang berlebihan. Kelebihan informasi adalah ketika jumlah informasi yang diterima orang membuat sulit untuk diproses. Karena informasi yang berlebihan, informasi yang berbeda dari berbagai sumber harus digabungkan menjadi informasi yang utuh, lengkap, akurat dan bermanfaat. (Priyanto and Ma'arif 2018). Ada begitu banyak artikel akademik di 1. Internet sehingga ketika kita mencari artikel akademik di mesin pencari, dibutuhkan waktu yang sangat lama bagi seseorang untuk menemukan artikel yang sesuai. (Veronica Ambassador Flores, Putri Agung Permatasari and Lie Jasa 2020)

Oleh karena itu diperlukan adanya teknologi informasi yang bertujuan untuk menelusuri halaman-halaman yang memberikan informasi tentang Perguruan Tinggi Negeri dan Swasta yang berstatus Universitas, Institut dan Sekolah tinggi di Sulawesi Barat, kemudian mengumpulkan semua informasi pada halaman tersebut secara otomatis dengan menggunakan teknologi web scraping. Informasi yang telah terkumpul selanjutnya akan diolah untuk mengetahui jumlahnya menggunakan teknologi *text mining* kemudian akan dibuat sebuah *website* sistem informasi untuk disajikan kepada calon mahasiswa. (Ma'arif 2016).

Web scraping adalah teknik yang digunakan untuk secara otomatis mengambil data atau informasi dari situs web. Informasi ini dapat berupa teks, link, video, audio, atau dokumen (A. Rahmatulloh and R. Gunawan 2020) *Text mining* adalah proses mengekstrak data dalam bentuk teks. Sumber data biasanya dari dokumen, dan tujuannya adalah untuk menemukan kata-kata yang dapat menggambarkan isi dokumen sehingga dilakukan analisis hubungan antar dokumen. (Ervita Kusuma Putri, and Tedy Setiadi 2014)

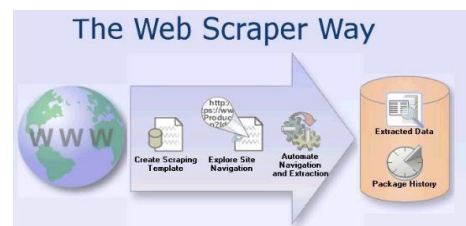
Berdasarkan uraian di atas maka di lakukan sebuah penelitian untuk membuat sistem informasi

dalam membantu calon mahasiswa/mahasiswi mencari informasi mengenai perguruan tinggi negeri dan swasta dengan status Universitas yang ada di Sulawesi Barat dengan judul “*Web Scraping Dan Text Mining Untuk Pengelompokan Informasi Perguruan Tinggi Di Sulawesi Barat*”.

2. METODE PENELITIAN

Metode Penelitian ini menggunakan penerapan teknik *web scraping* dan *text mining*, untuk pengambilan data kategori informasi perguruan tinggi yang ada di internet berupa link url dan judul artikel dengan teknik *web scraping*, dan kemudian menghitung jumlah data menggunakan *text mining*.

Web Scraping



Gambar 2.1 Ilustrasi cara kerja web scraping

Web scraping memiliki sejumlah langkah sebagai berikut (Josi, Andretti Abdillah and Suryayusra 2014):

Membuat template *scraping* adalah proses pemantauan dokumen HTML dari sebuah situs web dari mana informasi diambil atau di *scraping*. Caranya adalah dengan membuat tag HTML untuk melampirkan informasi yang ingin Anda ambil. Proses selanjutnya adalah memeriksa navigasi *website* yang terkena *scraping*. Proses selanjutnya adalah mengotomatiskan informasi yang diperoleh dari *website* yang di tentukan. Ini biasa disebut proses ekstraksi informasi. Setelah informasi berhasil diekstraksi, proses selanjutnya adalah menyimpan informasi ke database.

Text Mining

Penambangan teks atau *teks mining* adalah teknik / pendekatan algoritmik komputasi untuk mengekstraksi pengetahuan tersembunyi baru dari teks. Tujuan utama dari *text mining* adalah untuk mengekstrak informasi dari data dalam bentuk bahasa teks tertentu. Dibandingkan dengan data yang disimpan dalam database teks (data dalam format data teks), ini adalah data yang tidak terstruktur dan sulit untuk merancang algoritma untuk menyelesaikannya. (Muhammad Mulajati 2017). *Text mining* merupakan bagian dari ilmu temu kembali informasi yang mengolah data seperti teks yang cenderung tidak terstruktur. Secara umum, fase utama dari *text mining* terdiri dari tiga bagian utama: *text preprocessing*, *feature selection*, dan *text analysis*. (Priyanto and Ma'arif 2018).

Bahan Penelitian

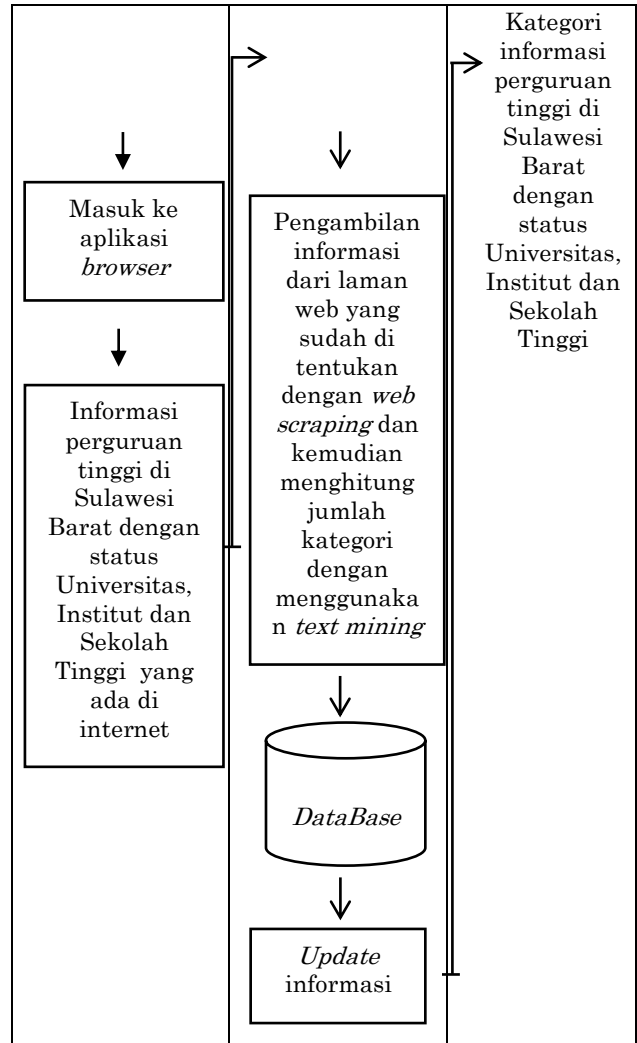
Bahan penelitian yang di gunakan untuk pembuatan aplikasi sistem informasi pengelompokan informasi perguruan tinggi yaitu data perguruan tinggi yang ada di internet.

Tempat Dan Waktu Penelitian

Tempat pelaksanaan penelitian di provinsi Sulawesi Barat. Waktu yang di butuhkan selama melakukan penelitian adalah mulai dari bulan Maret-2022 sampai Mei- 2022.

Teknik Pengumpulan Data

1. Observasi
2. Studi Pustaka



Gambar 2.2 Kerangka Sistem

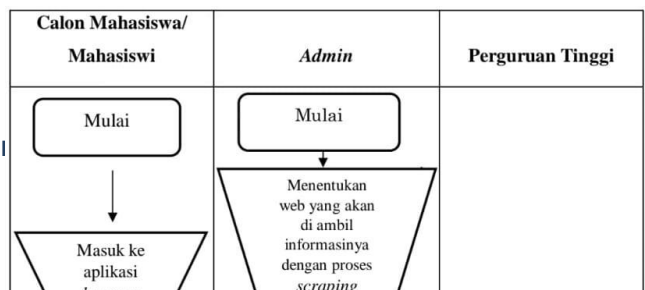
Berdasarkan kerangka sistem yang di buat langkah pertama yang di lakukan oleh *admin login* ke aplikasi *browser*/mesin pencari, menentukan *website* perguruan tinggi dengan status Universitas, Institut dan Sekolah Tinggi di Sulawesi Barat yang akan di ambil informasinya dengan *web scraping* dan kemudian menghitung jumlah kategori informasi menggunakan *text mining* berdasarkan informasi : a) Fasilitas, b) Akreditasi, c) Fakultas dan Prodi, d) Biaya Kuliah, e) Alumni. Kemudian di proses menjadi *database*, *Update* informasi perguruan tinggi dengan status Universitas, Institut dan Sekolah Tinggi di Sulawesi Barat. Calon mahasiswa/mahasiswi kemudian dapat melihat informasi perguruan tinggi dengan status Universitas berdasarkan kategori yang di pilih.

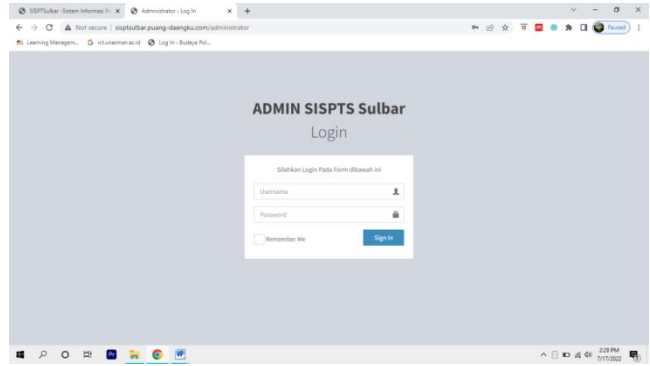
Kerangka Sistem

Input	Proses	Output
Star		



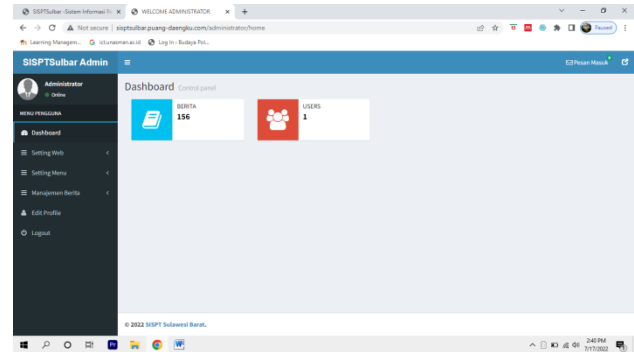
Analisis Sistem Yang Di Usulkan





Gambar 1. Halaman Login Admin

2. Halaman *dashboard*
Jika admin berhasil login maka admin dapat melihat tampilan menu *dashboard* seperti data informasi perguruan tinggi di Sulawesi Barat dan data user.



Gambar 2. Halaman *Dashboard*

Gambar 2.3 Analisis Sistem Yang Di Usulkan

3. HASIL DAN PEMBAHASAN

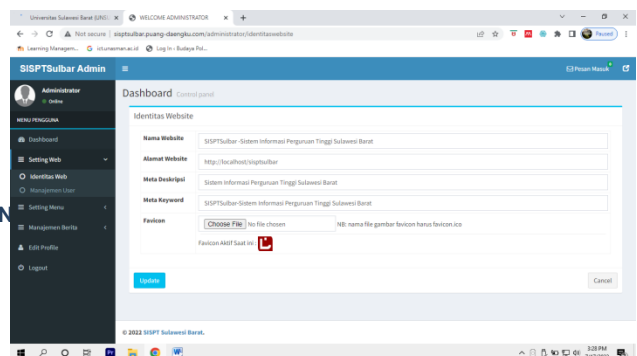
Hasil penelitian yang dirancang yaitu *web scraping* dan *text mining* untuk pengumpulan data pengelompokan informasi perguruan tinggi di Sulawesi Barat berbasis *website*.

Pada aplikasi ini memiliki 1 user yaitu *admin*, *website* pengelompokan informasi perguruan tinggi di Sulawesi Barat berisi informasi mengenai perguruan tinggi yang di kelompokkan ke dalam lima kategori informasi yang nantinya dapat memudahkan calon mahasiswa/mahasiswi dalam mencari informasi seputar perguruan tinggi yang ada di Sulawesi Barat, Program dibuat sesuai dengan desain yang telah dirancang agar sesuai dengan pengguna sistem. Pada tahap ini desain sistem *website* telah diterjemahkan kedalam Kode menggunakan bahasa pemrograman yang telah ditentukan.

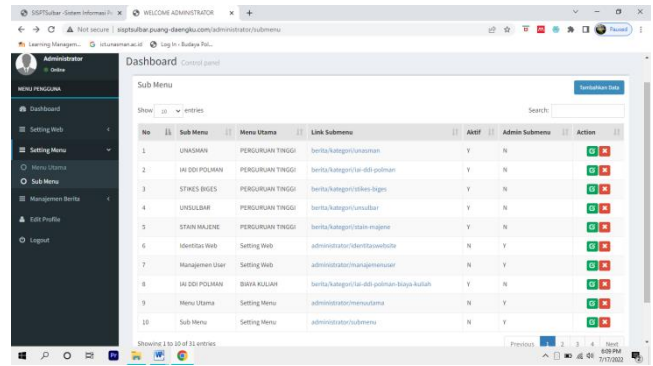
3. Halaman Identitas Web
Pada bagian sub menu identitas web halaman ini merupakan tampilan setelah admin memilih menu setting web, admin dapat melihat informasi identitas website sistem informasi pengelompokan informasi perguruan tinggi seperti nama *website*, meta *keyword* dan yang lainnya serta dapat mengeditnya.

Hasil Perancangan Sistem

1. Halaman *Login Admin*
Pada bagian halaman form *login admin* ini di gunakan khusus *admin* untuk dapat mengakses data informasi yang ada pada sistem.



menambahkan sub menu pada bagian menu utama yang di inginkan.

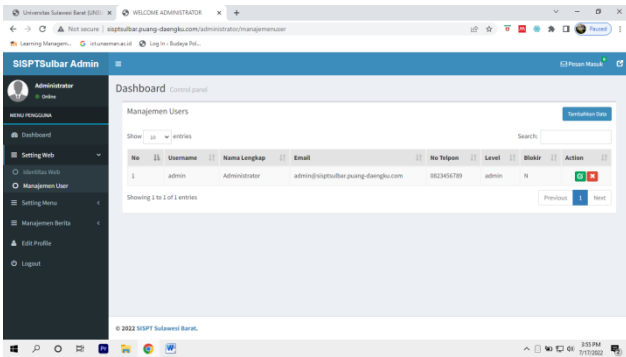


Gambar 6. Halaman Sub Menu

Gambar 3. Halaman Identitas Web

4. Halaman Manajemen User

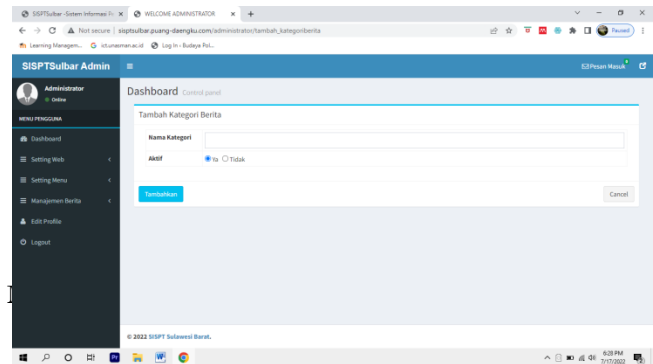
Halaman Menejemen User merupakan sub menu setelah admin memilih menu setting web, pada bagian ini admin dapat melihat data user serta melakukan tambah data menejemen user



Gambar 4. Halaman Manajemen User

7. Halaman Kategori Berita

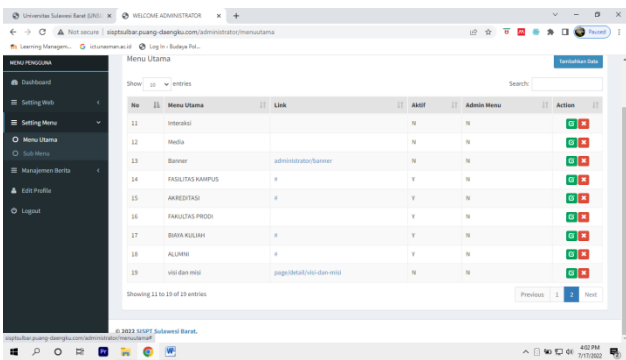
Halaman kategori berita merupakan sub menu setelah admin memilih menu manajemen berita, admin dapat menambahkan kategori berita berdasarkan kategori informasi dari perguruan tinggi dan dapat pula menghapus kategori berita yang telah di tambahkan.



Gambar 7. Halaman Kategori Berita

5. Halaman Menu Utama

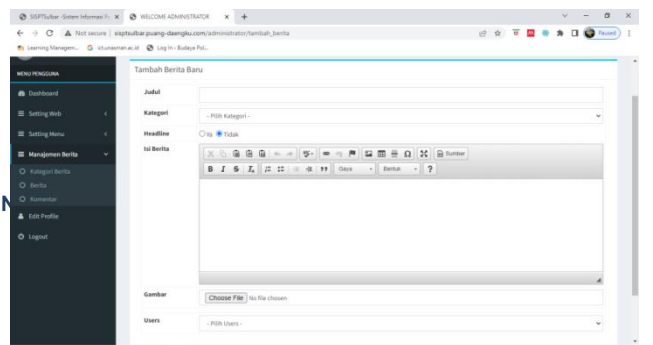
Halaman menu utama merupakan sub menu setelah admin memilih menu setting menu, pada bagian halaman ini admin dapat melakukan penambahan atau pengurangan menu utama pada aplikasi sistem informasi pengelompokan informasi perguruan tinggi.



Gambar 5. Halaman Menu Utama

8. Halaman Berita

Halaman berita merupakan sub menu setelah admin memilih manu manajemen berita, pada manejemen berita admin dapat menambahkan judul berita, pilihan kategori berita, isi berita dan informasi yang lainnya yang terdapat dalam halaman berita serta dapat menghapus berita yang telah di tambahkan



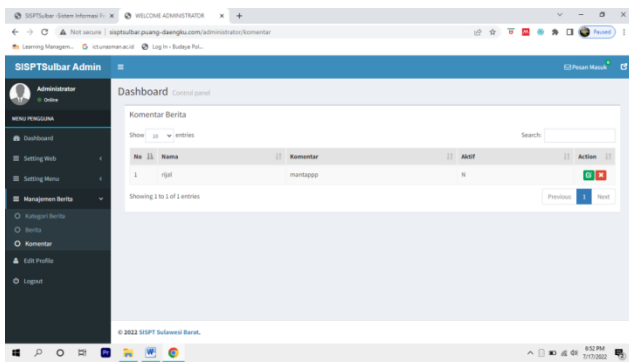
6. Halaman Sub Menu

Halaman sub menu merupakan bagian sub menu setelah admin memilih menu setting menu, pada bagian ini admin dapat menambahkan atau mengurangi sub menu dan mangatur ingin

Gambar 8. Halaman Berita

9. Halaman Komentar

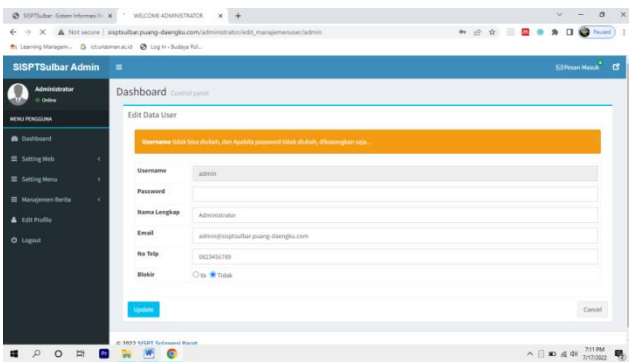
Halaman komentar merupakan bagian sub menu setelah *admin* memilih manu manajemen berita, pada bagian ini *admin* dapat melihat komentar dari pengguna *website* tentang artikel perguruan tinggi yang di publikasikan.



Gambar 9. Halaman Komentar

10. Halaman Edit Profile

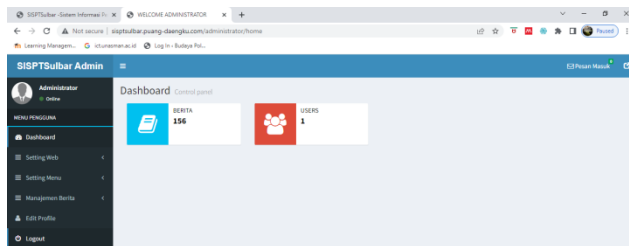
Pada halaman ini admin dapat menambahkan *user* serta melakukan tambahan data menejemen *user*.



Gambar 10. Halaman Edit Profile

11. Halaman Logout

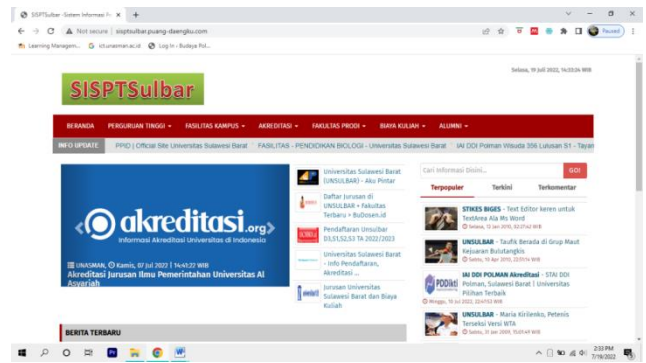
Pada halaman ini admin dapat keluar dari halaman administrator.



Gambar 11. Halaman Logout

12. Halaman Dashboard Pengguna

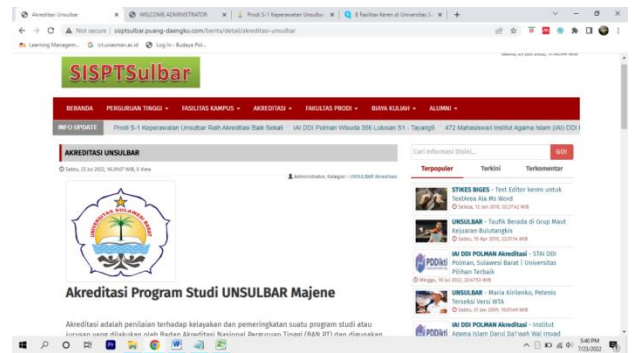
Halaman ini adalah tampilan informasi perguruan tinggi di mana pengguna aplikasi dapat memilih informasi perguruan.



Gambar 12. Halaman Dashboard Pengguna

13. Halaman Kategori Informasi

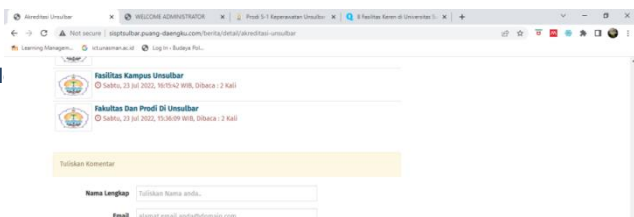
Halaman ini adalah tampilan di mana pengguna aplikasi dapat melihat informasi Perguruan Tinggi.



Gambar 13. Halaman Kategori Informasi

14. Halaman Komentar

Halaman ini adalah tampilan di mana pengguna aplikasi dapat melihat kategori informasi berdasarkan menu kategori informasi yang di pilih.



Gambar 14. Halaman Komentar

Uji Coba Sistem Dan Program

Sistem Pengujian *BlackBox* Dan *Whitebox*

no	Pengujian	Hasil yang di harapkan	Pengujian	Ket
1.	From Login Admin :			
	Button login	Input username dan <i>password</i> , jika berhasil maka akan di arahkan kehalaman selanjutnya yaitu dashboard	Berhasil login ke halaman adminisrator	Valid
2.	From Menu Utama Admin			
	Menu dashboard	Jika menu dashboard di klik maka akan tampil data informasi perguruan tinggi di Sulawesi Barat dan data user.	Sesuai yang diharapkan	Valid
	Menu setting web	Jika menu setting web di klik maka akan tampil informasi indentitas website dan data user serta dapat melakukan tambah data	Sesuai yang diharapkan	Valid

		menejemen <i>user</i>		
	Menu setting menu	Jika menu setting menu di klik maka akan tampil informasi menu utama dan sub menu serta admin dapat melakukan penambahan	Sesuai yang diharapkan	Valid
	Menu manajemen berita	Jika menu manajemen berita di klik maka akan tampil informasi kategori berita, dan komentar, admin juga dapat melakukan penambahan, pengurangan serta mengedit informasi	Sesuai yang diharapkan	Valid
	Menu edit profile	Jika menu edit profile di klik maka akan tampil informasi manajemen user admin juga dapat melakukan tambahan data menejemen <i>user</i> .	Sesuai yang diharapkan	Valid
	Menu <i>logout</i>	Pada halaman ini admin dapat keluar dari halaman administrator	Sesuai yang diharapkan	Valid
3.	From Menu Utama Pengguna			

Menu beranda	Jika menu beranda di klik maka akan muncul informasi perguruan tinggi.	Sesuai yang diharapkan	Valid
Menu perguruan tinggi	Jika menu perguruan tinggi di klik maka akan muncul informasi nama perguruan tinggi dan informasinya.	Sesuai yang diharapkan	Valid

Tabel 1. Pengujian *BlackBox*

no	Flowgraph	Independent path	Region	Cyclomatic Complexity
1.	Form Halaman Admin	8	8	8
2.	Form Halaman Donatur	4	4	4
3.	Jumlah	12	12	12

Tabel 2. Hasil Pengujian *WhiteBox*

4. SIMPULAN

Dari hasil penelitian yang dilakukan, maka dapat diambil kesimpulan bahwa:

1. Dengan adanya sistem informasi pengelompokan informasi perguruan tinggi menggunakan metode *web scraping* dan *text mining* berbasis *website* dapat memudahkan pengguna mendapatkan berbagai kategori informasi mengenai perguruan tinggi sehingga memudahkan calon mahasiswa/mahasiswi dalam mencari informasi yang di butuhkan.
2. Sistem mampu memenuhi kebutuhan calon mahasiswa/mahasiswi yang berfungsi sebagai sistem informasi perguruan tinggi dan memiliki kategori informasi yang telah di tentukan.

DAFTAR PUSTAKA

Tunggal and Budi (2020). Pengambilan Keputusan Strategis Pemasaran Di Perguruan Tinggi Dengan Menggunakan *Analytics Hierarchy Process* (AHP).

Muh. Amin, Idhan Zaldy and Ul Khairat (2021). Sistem Informasi Rumah Sakit Bersalin Dan Klinik Spesialis Mifta Berbasis Website.

Hasanah (2019). Implementasi Nilai Toleransi Terhadap Mahasiswa Lintas Keyakinan Pada Perguruan Tinggi Di Yogyakarta.

Priyanto and Ma'arif (2018). Implementasi *Web Scrapping* Dan *Text Mining* Untuk Akuisisi Dan Kategorisasi Informasi Dari Internet (Studi Kasus: Tutorial Hidroponik).

Veronica Ambassador Flores, Putri Agung Permatasari and Lie Jasa (2020). Penerapan Web Scraping Sebagai Media Pencarian dan Menyimpan Artikel Ilmiah Secara Otomatis Berdasarkan Keyword.

Ma'arif (2016). Integrasi Laman Web Tentang Pariwisata Daerah Istimewa Yogyakarta Memanfaatkan Teknologi *Web Scraping* Dan *Text Mining*.

A. Rahmatulloh and R. Gunawan (2020). Web Scraping with HTML DOM Method for Data Collection of Scientific Articles from Google Scholer.

Ervita Kusuma Putri, and Tedy Setiadi (2014). Penerapan Text Mining Pada Sistem Klasifikasi Email Spam Menggunakan Naive Bayes.

Josi, Andretti Abdillah and Suryayusra (2014). Penerapan Teknik *Web Scraping* Pada Mesin Pencari Artikel Ilmiah.

Muhammad Mulajati (2017). Implementasi Teknik Web Scraping Dan Klasifikasi Sentimen Menggunakan Metode Naive Bayes Classifier Dan Asosiasi Teks.