Telematika: Jurnal Informatika dan Teknologi Informasi
Vol. 18, No. 2, Juni 2021, pp. 222-230
ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v18i2.5506

Automated Website Monitoring System Using Web Scraping and Raspberry Pi

Sistem Pemantauan Situs Web Otomatis Menggunakan Web Scraping dan Raspberry Pi

Putra Prima Arhandi[1], Irsyad Arief Mashudi[2], Fuad Adi Nugroho[3]

[1,2,3] Teknik Informatika, Politeknik Negeri Malang, Indonesia

[1]putraprima@polinema.ac.id, [2]irsyad.arif@polinema.ac.id, [3]fhadinugroho@gmail.com

Informasi Artikel
Received: March 2021
Revised: May2021
Accepted: June 2021
Published: August 2021

Abstract

***Purpose**: Create a system to monitor website availability automatically using web scraping and raspberry pi*
***Design/methodology/approach**: This system successfully checks website availability using various ISPs with an accuracy of more than 90%.*
***Findings/result**: This system successfully checks website availability using various ISPs with an accuracy of more than 90%.*
***Originality/value/state of the art**: The contribution of this research is to create systems and agents that collaborate automatically to check website availability.*

Keywords:automation,website monitoring, website availability
Kata kunci: otomasi, pemantauan situs web, ketersediaan situs web

Abstrak

Tujuan: Membuat sebuah sistem untuk melakukan pemantauan ketersediaan situs web secara otomatis menggunakan web scraping dan raspberyy pi
Perancangan/metode/pendekatan: Pada penelitian ini dibuat sebuah sistem utama sebagai pusat data dan beberapa agent menggunakan raspberry pi. Sistem utama dibangun menggunakan codeigniter dan web scraping di raspberry pi dilakukan menggunakan node js serta REST API untuk komunikasi antara agent dan sistem utama.
Hasil: Sistem ini berhasil melakukan pengecekan ketersediaan situs web menggunakan berbagai ISP dengan keakuratan lebih dari 90%.
Keaslian/ *state of the art*: Kontribusi penelitian ini adalah membuat sistem dan agen yang berkolaborasi secara otomatis untuk mengecek ketersediaan situs web.

1. Pendahuluan

Sebuah sistem yang terintegrasi sudah menjadi kebutuhan pokok yang harus dimiliki sebuah perguruan tinggi [1]. Sistem ini membantu perguruan tinggi dalam menjalankan fungsinya.

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v18i2.5506

Umumnya, pengguna sistem ini menggunakan antarmuka web untuk berinteraksi dengan sistem. Seiring berjalannya waktu, sistem yang ada biasanya akan bertambah kompleks seiring dengan bertambahnya kebutuhan yang harus dipenuhi sistem. Padahal, sistem harus terus diperbarui agar dapat memenuhi kebutuhan civitas akademika [2].

Bertambahnya kebutuhan yang harus dipenuhi sistem juga akan menambah antarmuka baru. Antarmuka yang berbentuk web ini umumnya harus tersedia 24 jam. Secara periodik, antarmuka-antarmuka ini harus diperiksa ketersediaannya. Karena jumlahnya yang banyak, pengecekan secara manual sangat tidak disarankan.

Pemeriksaan ketersediaan web antarmuka sistem juga harus memperhatikan ISP. Seringkali sebuah antarmuka berjalan pada satu ISP tapi tidak berjalan pada ISP lain. Hal ini semakin membuat proses pemeriksaan ketersediaan web antarmuka semakin panjang. Apalagi pemeriksaan ini harus dilakukan secara periodik dengan interval yang singkat. Karena itu diperlukan sebuah sistem yang dapat memeriksa web-web antarmuka ini secara otomatis dengan menggunakan beberapa ISP.

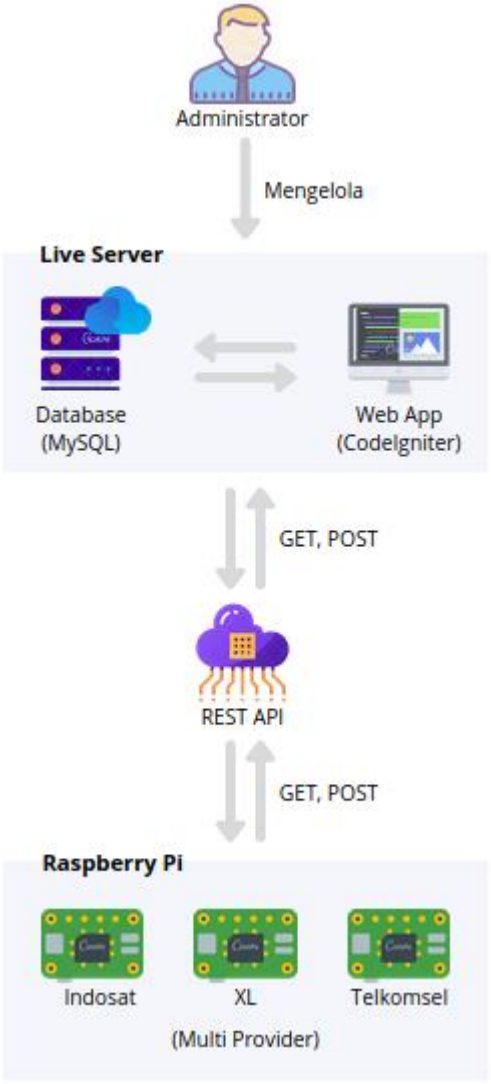
Banyak metode yang dilakukan oleh peneliti untuk melakukan pemeriksaan atau mengekstrak data dari sebuah website metode yang dominan digunakan adalah web scraping dan web crawlling. Web crawling dapat dilakukan pada website public yang berbentuk social media seperti twitter, youtube dan facebook [3]. Web scraping dapat digunakan untuk melakukan ekstraksi Document Object Model (DOM) dan mengkombinasikannya dengan JSON dalam proses ekstraksi gambar [4], selain itu web scraping juga dapat digunakan pada proses pencarian artikel ilmiah [5] dan ekstraksi data pada korpus yang berbeda bahasa [6]. Web scraping dan crawling juga dapat di kombinasikan untuk melakukan pengumpulan dataset pada proses data mining [7].Selain pada halaman web, web scraping juga dapat dilakukan terhadap response sebuah API [8].

Raspberry pi merupakan salah satu solusi untuk melakukan automasi pada berbagai bidang, raspberry pi dapat digunakan pada sistem identifikasi plat nomor kendaraan [9]. Dengan menambahkan berbagai sensor pada Raspberry Pi dapat digunakan untuk melakukan automasi yang berbasis sensor [10] [11] [12]. Karena ukuran yang kecil dan mudah dipindahkan raspberry pi juga dapat digunakan dalam bentuk agen baik yang bekerja sendiri sendiri atau bekerja bersama salah satu contohnya dalap sistem segmentasi gambar medis dan sistem e health [13] [14].

Rakhmawati dkk telah membuat sistem perankingan website untuk Institut Teknologi Sepuluh Nopember. Sistem ini memantau semua situs dibawah domain ITS. Sayangnya, sistem ini hanya menggunakan satu ISP dalam memeriksa ketersediaan situs. Padahal bisa saja sebuah web antarmuka dapat berjalan dengan lancar pada satu ISP namun tidak bisa diakses menggunakan ISP yang lain. Di sisi lain, sistem ini tidak hanya memeriksa ketersediaan web saja, tapi juga kecepatan akses, broken link, dan aspek-aspek lain [15].

Pada penelitian ini dibuat sebuah sistem untuk memeriksa ketersediaan website antarmuka sebuah sistem terintegrasi. Sistem yang dibuat diperuntukkan memeriksa ketersediaan web antarmuka pada sistem terintegrasi di Politeknik Negeri Malang (POLINEMA). Sistem ini dapat

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v18i2.5506



Gambar 1 Arsitektur Sistem

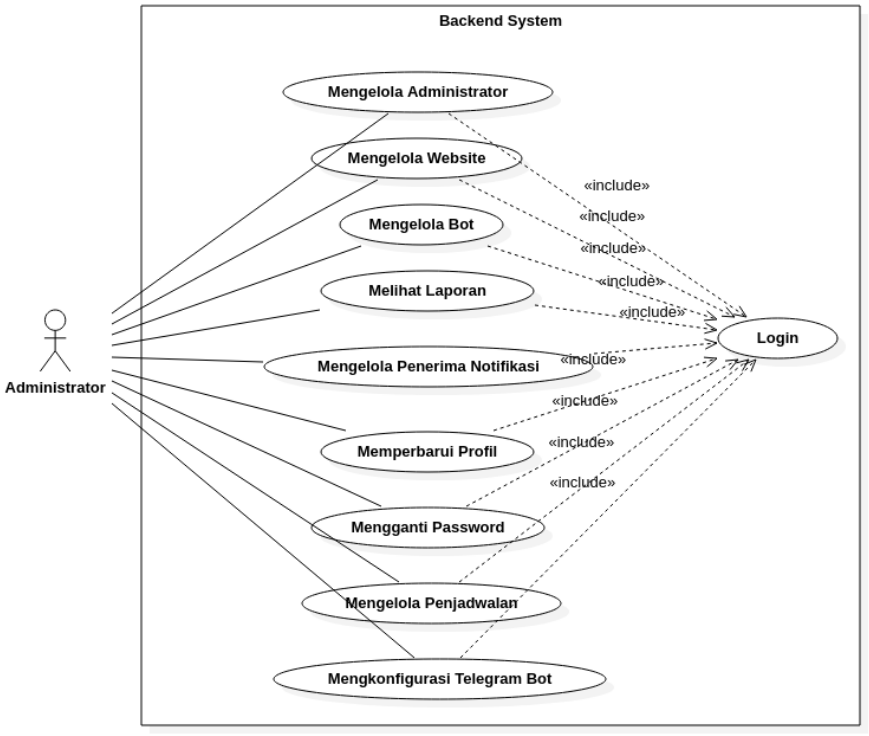
memeriksa ketersediaan web antarmuka melalui banyak provider internet dengan menggunakan raspberry pi sebagai agen.

2. Perancangan

Penelitian ini dilakukan berdasarkan kasus nyata yang terjadi pada Pusat Komputer (PusKom) POLINEMA. PusKom ini adalah unit kerja pada Polinema yang bertanggung jawab atas ketersediaan situs-situs pada Polinema. Situs-situs tersebut harus dipantau dan dipastikan

ketersediaannya agar dapat dipakai semua civitas akademik Polinema. Penelitian ini menggunakan data url situs yang ada di bawah PusKom.

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v18i2.5506



Gambar 2 Use Case Diagram Sistem

Gambar 3 Langkah Langkah Melakukan Web Scraping

Penelitian ini akan membuat sebuah sistem yang dapat memantau ketersediaan situs di bawah manajemen PusKom secara otomatis. Sistem yang dibuat terdiri dari dua subsistem : satu sistem utama berupa web admin dan sistem bot yang berupa scrapper website yang berjumlah banyak. Sistem utama diinstal pada sebuah web server sedangkan sistem bot diinstal pada Rasberry Pi. Masing-masing bot mengakses internet menggunakan provider yang berbeda-beda.

Untuk memastikan bahwa situs yang diperiksa sedang berjalan, sistem akan mengambil data berupa screenshot, response code, dan waktu loading. Data-data tersebut akan didapatkan dengan cara scraping. Para peneliti juga sepakat bahwa cara ini terbukti handal untuk mengekstak data dari sebuah situs [4]–[7], [16]. Pertama, user akan memerintahkan untuk memeriksa ketersediaan situs melalui web admin. Kemudian, web admin akan menyalurkan perintah tersebut ke bot-bot yang ada. Masing-masing bot mengakses situs yang dituju menggunakan provider pada bot masing-masing. Kemudian, masing- masing bot akan men-scrap situs untuk mendapatkan response code, Screenshot, dan waltu loading. Masing-masing

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v18i2.5506

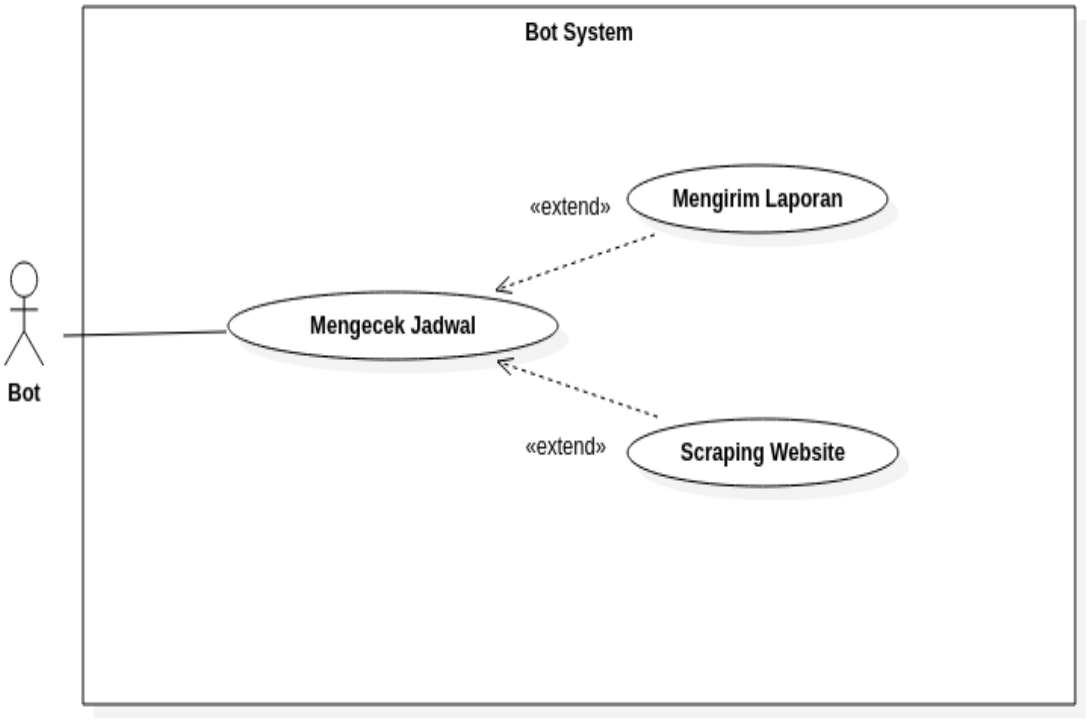
bot akan mengembalikan data ini untuk ditampilkan pada user pada web admin. Arsitektur sistem dapat dilihat pada **Gambar 1**.

Pengguna sistem berinteraksi dengan sistem melalui antarmuka pada sistem utama. Sistem utama ini adalah sistem berupa web admin. Sistem utama ini memiliki fungsi utama sebagai antarmuka untuk mengkonfigurasi sistem secara keseluruhan. Konfigurasi dapat dilakukan pada sistem utama maupun bot-bot dibawah sistem utama.

Pada sistem utama, pengguna dapat mengkonfigurasi situs-situs yang akan diperiksa ketersediaannya. Umumnya, situs-situs ini berupa halaman web yang ada di bawah Puskom Polinema. Situs-situs ini secara berkala akan diperiksa ketersediaannya sesuai jadwal yang dapat diatur pada sistem utama ini. Jika ketika diperiksa situs tidak tersedia, sistem utama dapat mengirimkan notifikasi kepada pengguna yang ditunjuk. Segala hal yang berhubungan dengan administrasi sistem juga diatur pada sistem utama.

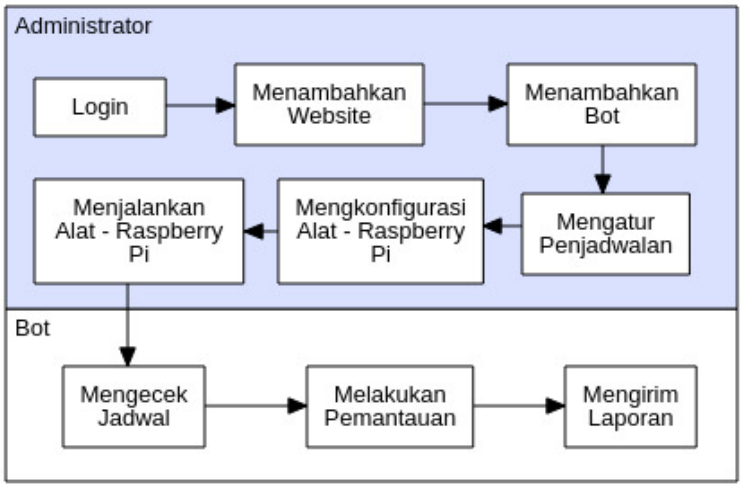
Pengguna juga bisa menambah/mengurangi sistem bot melalui sistem utama. Setiap bot mengakses internet melalui provider yang berbeda. Bot-bot ini berupa Raspberry Pi.

Konfigurasi bot-bot ini dilakukan melalui sistem utama. Diagram Use Case sistem utama dapat dilihat pada **Gambar 2**.



Gambar 4 Diagram Use Case Bot

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v18i2.5506



Gambar 5 Cara Kerja Sistem

Sesuai jadwal yang diatur pada sistem utama, tiap bot akan memeriksa ketersediaan situs. Karena tiap bot menggunakan provider yang berbeda dalam mengakses internet, maka hasil pemeriksaan ini bisa berbeda-beda untuk setiap bot. Secara default, bot akan meminta jadwal pemeriksaan situs pada sistem utama setiap dua menit.

Bot akan memeriksa ketersediaan situs dengan cara scraping seperti pada **Gambar 3**. Bot akan mengambil metadata dari website yang dituju. Metadata ini dapat mewakili ketersediaan situs[17]. Pertama bot akan menghapus screenshot dari pemantauan sebelumnya jika ada. Lalu bot akan meminta data url pada server. Server akan mengirim data url yang diminta. Setelah bot mendapatkan data url, proses yang dilakukan selanjutnya adalah mengambil data kode respon halaman (response code) dan tangkapan layar (screenshot) dengan cara mengekstrak data dari url. Data ini nantinya akan disimpan ke penyimpanan sementara pada sistem bot sampai jadwal pengiriman data ke sistem utama. Diagram usecase sistem bot dapat dilihat pada **Gambar 4**.

Setiap bot dikonfigurasi untuk mengirimkan data hasil scraping ke sistem utama. Ketika waktu pengiriman ini tiba, maka bot akan mengirimkan data tersebut ke sistem utama. Kemudian, sistem utama akan menyimpan data ini dalam database.

Setelah berhasil dikirim, data scraping pada bot akan dihapus untuk menghemat tempat penyimpanan. Secara keseluruhan, cara kerja sistem dapat dilihat pada **Gambar 5**.

Tabel 1 Pengujian Fungsionalitas Sistem

Bot	Jumlah Data	Sukses	Gagal	Persentase
Indihome	195	188	7	96,4
Indosat	195	186	9	95,4
XL	195	188	7	96,4

Tabel 2 Pengujian Kebergunaan Sistem

Kode	Pertanyaan	Nilai				
		STS	**TS**	**RR**	**S**	**SS**
Learnability						
P1	Sistem mudah dipelajari dan digunakan	-	-	-	2	
P2	Menu-menu yang ada mudah dipahami	-	-	-	2	
P3	Jenis huruf yang digunakan mudah dibaca	-	-	-	2	
P4	Penggunaan bahasa mudah dimengerti	-	-	-	2	
P5	Simbol, icon dan gambar mudah dipahami	-	-	-	2	
Efficiency						
P6	Halaman dapat ditampilkan dengan cepat	-	-	-	1	1
P7	Filter dan pencarian dapat ditampilkan dengan tepat	-	-	-	2	
Memorability						
P8	Menu dan halaman mudah diingat	-	-	-	1	1
P9	Saya dapat menggunakan sistem dengan mudah tanpa instruksi tertulis	-	-	-	1	1
Error						
P10	Saya tidak menemukan under reconstruction dari sistem ini	-	-	-	1	1
P11	Saya tidak menemukan link error	-	-	-	2	
Satisfaction						
P12	Sistem bekerja sesuai dengan harapan saya	-	-	-	2	

3. Hasil dan Pembahasan

Sistem diimplementasikan dengan menggunakan url situs dibawah Puskom Polinema untuk ujicoba. Pengujian dilakukan dengan melihat tingkat keberhasilan alat dalam melakukan pemantauan ketersediaan situs web. Sistem diuji menggunakan tiga bot yang memakai provider Indihome, Indosat, dan XL dalam mengakses internet. Ujicoba dilakukan selama dua minggu dengan hasil seperti pada **Tabel 1**.

Berdasarkan hasil pengujian metode di atas menunjukkan bahwa tidak menutup kemungkinan persentase keberhasilan dapat berubah secara dinamis. Ada 2 faktor yang dapat mempengaruhi tingkat keberhasilan pemantauan ini. Pertama adalah jaringan koneksi internet yang digunakan apakah stabil atau tidak. Kedua, kondisi server website yang dilakukan pemantauan apakah dapat diakses atau tidak. Namun menurut data di atas didapatkan rata-rata persentase keberhasilan Indihome sebesar 96.41%, Indosat sebesar 95.38% dan XL sebesar 96.41%. Dapat disimpulkan, metode ini memiliki persentase keberhasilan diatas 90%.

Selain pengujian kebutuhan fungsional, sistem juga diuji kebergunaannya. Kebergunaan penting agar sistem dapat digunakan dengan mudah oleh pengguna . Pengujian kebergunaan dilakukan dengan memberi kuesioner pada tim PusKom Polinema selaku pengguna utama sistem. Hasil pengujian menunjukkan bahwa sebagian besar fitur sistem sudah sesuai dengan kebutuhan. Hasil kuesioner dapat dilihat pada **Tabel 2**.

4. Kesimpulan

Pada penelitian ini telah dibuat sebuah sistem yang dapat memeriksa ketersediaan sebuah sistem terintegrasi. Sistem ini memiliki kelebihan yaitu dapat memeriksa ketersediaan menggunakan banyak provider internet sekaligus. Saat diuji, sistem dapat memeriksa ketersediaan dengan presentase keberhasilan lebih dari 90%. Sistem juga telah diuji kebergunaannya. Hasil pengujian kebergunaan menunjukkan bahwa sistem sudah sesuai dengan kebutuhan. Untuk kedepannya, sistem dapat ditingkatkan kehandalannya atau mengganti perangkat Raspberry Pi dengan perangkat sejenis yang biayanya lebih murah dan diuji pengaruhnya.

Daftar Pustaka

[1] M. Manzoor, W. Hussain, A. Ahmed, and M. J. Iqbal, "The importance of Higher Education Website and its Usability," *Int. J. Basic Appl. Sci.*, vol. 1, no. 2, pp. 150–163, Apr. 2012, doi: 10.14419/ijbas.v1i2.73.

[2] N. Kesswani and S. Kumar, "Accessibility analysis of websites of educational institutions," *Perspect. Sci.*, vol. 8, pp. 210–212, Sep. 2016, doi: 10.1016/j.pisc.2016.04.031.

[3] A. N. Abadi and N. A. Rakhmawati, "Rancang Bangun Perangkat Lunak Benchmarking Sosial Media Pemerintah Daerah Indonesia," *J. Tek. ITS*, vol. 6, no. 2, pp. A302-307, Sep. 2017, doi: 10.12962/j23373539.v6i2.23260.

[4] I. A. Ahmad Sabri, M. Man, W. A. W. Abu Bakar, and A. N. Mohd Rose, "Web Data Extraction Approach for Deep Web using WEIDJ," *Procedia Comput. Sci.*, vol. 163, pp. 417–426, 2019, doi: 10.1016/j.procs.2019.12.124.

[5] A. Josi and L. A. Abdillah, "PENERAPAN TEKNIK WEB SCRAPING PADA MESIN PENCARI ARTIKEL ILMIAH," p. 6.

[6] V. Mitra and H. Sujaini, "Rancang Bangun Aplikasi Web Scraping untuk Korpus Paralel Indonesia - Inggris dengan Metode HTML DOM," vol. 5, no. 1, p. 6, 2017.

[7] L. P. Kaspa, V. N. S. S. Akella, Z. Chen, and Y. Shi, "Towards Extended Data Mining: An Examination of Technical Aspects," *Procedia Comput. Sci.*, vol. 139, pp. 49–55, 2018, doi: 10.1016/j.procs.2018.10.216.

[8] L. C. Dewi, Meiliana, and A. Chandra, "Social Media Web Scraping using Social Media Developers API and Regex," *Procedia Comput. Sci.*, vol. 157, pp. 444–449, 2019, doi: 10.1016/j.procs.2019.08.237.

[9] A. O. Agbeyangi, O. A. Alashiri, and A. E. Otunuga, "Automatic Identification of Vehicle Plate Number using Raspberry Pi," in *2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS)*, Ayobo, Ipaja, Lagos, Nigeria, Mar. 2020, pp. 1–4. doi: 10.1109/ICMCECS47690.2020.246983.

[10] Noorinder, J. Singh, and E. Sidhu, "Raspberry pi based smart fire management system employing sensor based automatic water sprinkler," in *2017 International Conference on Power and Embedded Drive Control (ICPEDC)*, Chennai, India, Mar. 2017, pp. 102–106. doi: 10.1109/ICPEDC.2017.8081068.

[11] C. N. Cabaccan, F. R. G. Cruz, and I. C. Agulto, "Wireless sensor network for agricultural environment using raspberry pi based sensor nodes," in *2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and*

Control, Environment and Management (HNICEM), Manila, Dec. 2017, pp. 1–5. doi: 10.1109/HNICEM.2017.8269427.

[12] A. Sreejithlal, M. N. Syam, T. M. Letha, K. P. M. Madhusoodanan, and A. Shooja, "Pressure Sensor Test System Using Raspberry Pi," in *2018 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, Thiruvananthapuram, India, Dec. 2018, pp. 182–185. doi: 10.1109/RAICS.2018.8635050.

[13] H. Bensag, M. Youssfi, and O. Bouattane, "Embedded agent for medical image segmentation," in *2015 27th International Conference on Microelectronics (ICM)*, Casablanca, Morocco, Dec. 2015, pp. 190–193. doi: 10.1109/ICM.2015.7438020.

[14] C. G. Toader, "Multi-Agent Based E-Health System," in *2017 21st International Conference on Control Systems and Computer Science (CSCS)*, Bucharest, Romania, May 2017, pp. 696–700. doi: 10.1109/CSCS.2017.107.

[15] N. A. Rakhmawati, V. Ferlyando, F. Samopa, and H. M. Astuti, "A performance evaluation for assessing registered websites," *Procedia Comput. Sci.*, vol. 124, pp. 714–720, 2017, doi: 10.1016/j.procs.2017.12.209.

[16] U. Baskaran and K. Ramanujam, "Automated scraping of structured data records from health discussion forums using semantic analysis," *Inform. Med. Unlocked*, vol. 10, pp. 149–158, 2018, doi: 10.1016/j.imu.2018.01.003.

[17] V. Draxl, "Web Scraping Data Extraction from websites," p. 38.