 ISSN (Online) 2828-786X

IMPLEMENTASI TEKNIK WEB SCRAPING UNTUK MENAMPILKAN DATA TIM ENGLISH PREMIER LEAGUE

IMPLEMENTATION OF WEB SCRAPING TECHNIQUE TO SHOWING ENGLISH PREMIER LEAGUE TEAM DATA

Anisa Rahma Salsabila[1)], Muhammad Daffa[2)], Muhammad Kandias Happy Maulana[3)], Eka Dyar Wahyuni[4)]
E-mail : [1)]anisarsalsabila@gmail.com , [2)]muhammaddaffa868@gmail.com ,
[3)]muhammadkandias.hm@gmail.com , [4)]ekadyar.w@gmail.com

[1, 2, 3,4]Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur

Abstrak

Perkembangan teknologi membawa dampak langsung terhadap kehidupan manusia, baik sisi positif maupun negatif. Terdapat beberapa cara dalam mengintegrasikan sebuah sistem yang akan dibangun, salah satunya dengan menerapkan integrasi level *user interface.* Integrasi level *user interface* dapat dilakukan dengan teknik *web scraping*. Pada penelitian ini akan diterapkan integrasi level *user interface* menggunakan teknik *web scraping* dengan bantuan aplikasi *web scraper* yang merupakan ekstensi dari Google Chrome dalam menghasilkan file CSV. Data penelitian ini akan diambil dari *website English Premier League.* Data CSV yang dihasilkan melalui *web scraping* berupa data tim *English Premier League*, dimana nantinya akan diproses menggunakan ETL (*Extraction, Transformation, Loading*) pada aplikasi *Pentaho Data Integration*. Luaran dari proses yang telah dijelaskan sebelumnya akan dilakukan visualisasi dalam bentuk halaman website.
Kata kunci: *implementasi, web scraping, data, tim english premier league*

Abstract

Technology developments have a direct impact on human life, either positive or negative sides. There are several ways to integrate the system that will be built, one of them using the user interface integration level. User interface integration level can be implemented with web scraping technique. In this research will be implemented with web scraping technique using the help of web scraper application which is Google Chrome extension to produce CSV file. The research data will be taken from English Premier League website. This CSV data generated through web scraping is an English Premier League team data, which will be processed using ETL (Extraction, Transformation, Loading) on Pentaho Data Integration application. The output of the previously process described will be visualized in the form of a website page.
***Keywords*:** *implementation, web scraping, data, english premier league team*

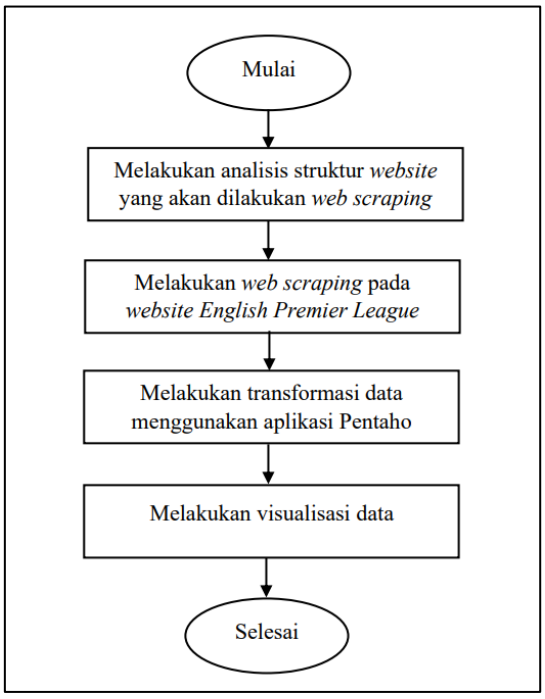
1. PENDAHULUAN

Teknologi yang semakin berkembang menyebabkan manusia harus dapat beradaptasi dalam berbagai aspek kehidupan. Dengan adanya penerapan teknologi, pekerjaan manusia akan sangat terbantu. Akan tetapi, pada penerapannya teknologi justru menimbulkan masalah bagi manusia. Salah satu permasalahan yang terjadi yaitu keterbatasan dalam mengakses data sehingga tidak semua platform memperbolehkan untuk mengambil data yang dibutuhkan.
Terdapat beberapa teknik atau cara dalam mengintegrasikan sebuah sistem, dimana bisa atau tidaknya teknik tersebut digunakan atau diimplementasikan tergantung pada ketersediaan data.

 ISSN (Online) 2828-786X

Dalam melakukan pengintegrasian sistem terdapat beberapa tingkatan atau level, diantaranya adalah level data, aplikasi (API), metode, dan *user interface*[1]. Pada integrasi level data terjadi proses pemisahan atau perpindahan data antar aplikasi dengan tujuan untuk membagi data yang sama ke dalam beberapa aplikasi yang berbeda, sedangkan integrasi level aplikasi berpusat pada fungsionalitas *sharing* dalam logika bisnis. Kemudian, pada integrasi level metode memiliki fungsi sebagai proses representasi data dari metode bisnis melalui *interface*, sedangkan integrasi level *user interface* memungkinkan dalam mengakses fungsi melalui *user interface* secara umum. Pada penelitian ini akan berfokus pada integrasi level *user interface* untuk mengambil beberapa data yang terdapat pada situs *website English Premier League* menggunakan teknik *web scraping*. *Web scraping* merupakan metode untuk mengumpulkan informasi yang tidak dapat dilakukan dengan pencarian manual melalui *search engine* Google Search[2]. *Web scraping* merupakan suatu teknik dalam mendapatkan informasi dari *website* secara praktis tanpa harus menggali informasi secara manual. Teknik *web scraping* ini bisa dilakukan dengan bantuan aplikasi *web scraper* dari Google Chrome. Tujuan dari *web scraper* adalah untuk mengambil informasi tertentu yang selanjutnya dapat diunduh dalam bentuk CSV atau XLSX file[3].
Dalam proses integrasi level *user interface*, nantinya juga akan menerapkan integrasi level data menggunakan *Pentaho Data Integration* untuk memindahkan data dari CSV ke dalam tabel pada *database* yang telah dibuat sebelumnya. *Pentaho Data Integration-Kettle* (PDI) adalah suatu aplikasi yang dapat mengintegrasikan sebuah data, dimana Pentaho dapat mengolah data yang siap untuk diolah pada *data warehouse* sehingga nantinya data dapat disajikan secara tepat waktu dan akurat[4]. *Software* Pentaho dapat menjalankan proses ETL (*Extraction, Transformation, Loading*). Selain digunakan untuk mengintegrasikan data, Pentaho dapat digunakan untuk membersihkan data serta memuat sebuah data dari file ke *database* atau sebaliknya dalam jumlah skala besar[5]. Data-data yang telah dilakukan pengintegrasian sistem, nantinya akan dilakukan visualisasi dalam bentuk tampilan *website.*

2. METODOLOGI



Gambar 1. Alur Penelitian

Alur penelitian ini seperti dijelaskan pada Gambar 1, dimana penelitian ini terdiri dari tiga bagian utama. Sebelum melakukan bagian pertama, peneliti melakukan analisis struktur *website* yang

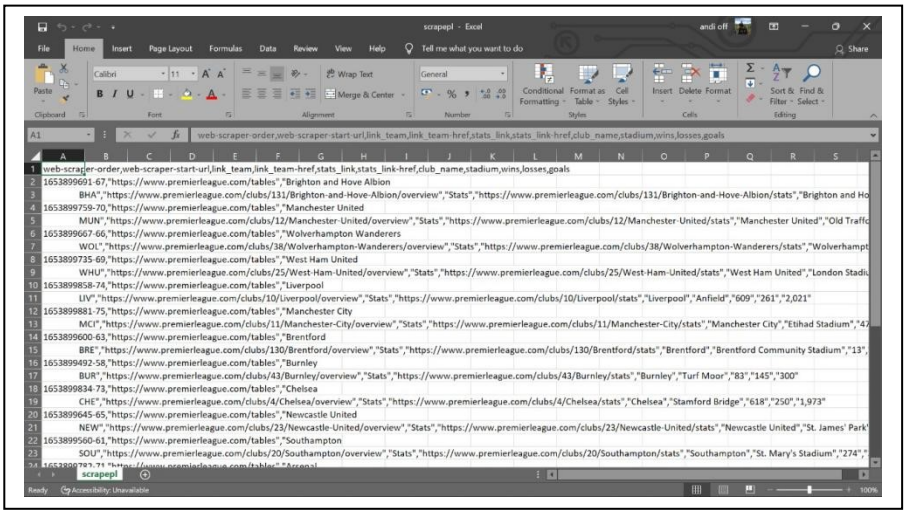
 ISSN (Online) 2828-786X

akan dilakukan *web scraping.* Hal ini perlu dilakukan karena peneliti harus mengetahui struktur *website* yang akan diambil datanya dan terkadang terdapat *website* yang tidak bisa untuk dilakukan *scraping*. Bagian pertama dimulai dengan mengimplementasikan teknik *web scraping* yang merupakan jenis integrasi pada level *user interface*. *Output* dari proses *web scraping* ini adalah berupa data CSV yang didapatkan dari analisis terhadap halaman *website* dalam bahasa *markup* seperti HTML atau XHTM[6]. Data yang diambil didapatkan dari situs resmi *English Premier League* (https://www.premierleague.com/) berupa nama tim, nama stadion, jumlah kemenangan, jumlah kekalahan, dan total gol. Aplikasi yang digunakan untuk melakukan teknik *web scraping* adalah *web scraper* yang merupakan ekstensi dari Google Chrome.
Data tim yang dihasilkan dari teknik *web scraping* akan diproses pada bagian kedua yaitu proses tranformasi data. Proses ini termasuk ke dalam jenis integrasi level data yang bertujuan untuk membuat sumber data yang sama antara data di *database* dan data pada CSV[7]. Dalam melakukan integrasi level data tersebut digunakan aplikasi *Pentaho Data Integration* untuk melakukan proses ETL *(Extract Transform Load)*[8]. Langkah pertama yang dilakukan adalah mengkonfigurasi *CSV Input* untuk mengambil data hasil dari *web scraping.* Kemudian *output* dari proses ini adalah berupa tabel dalam *database* yang telah dibuat sebelumnya melalui MySQL.
Bagian terakhir pada penelitian ini yaitu melakukan visualisasi data. Hasil transformasi data setelah dilakukan proses integrasi level data nantinya akan ditampilkan dalam bentuk halaman *website*. *Website* untuk visualisasi teknik *web scraping* dibuat menggunakan bahasa pemrograman PHP dan terhubung dengan database MySQL yang biasa digunakan dalam membuat *website*[9]. Konfigurasi koneksi ke *database* tersimpan di dalam file koneksi.php. Sedangkan, tampilan yang digunakan yaitu berupa tabel dari *template bootstrap.*

3. HASIL DAN PEMBAHASAN

3.1 Hasil *Web Scraping*

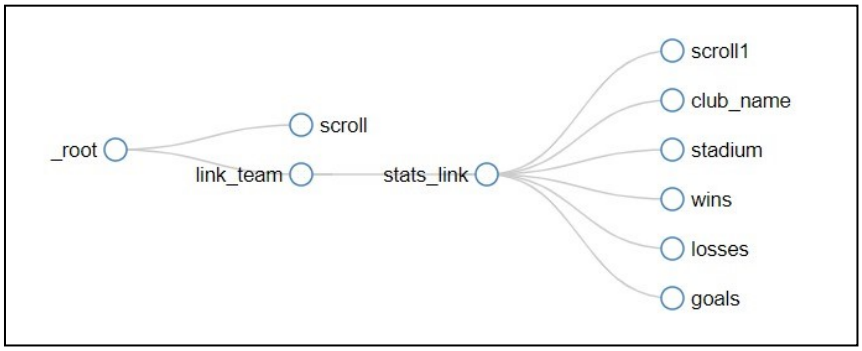
Data yang diperlukan terdapat pada halaman *tables* di situs *website* resmi *English Premier League* (https://www.premierleague.com/tables). Hasil *scraping* memberikan data yang ditunjukkan pada Gambar 2, dimana data tersebut didapatkan dari otomatisasi pengambilan data pada setiap tim yang berlaga pada kasta tertinggi sepak bola Inggris tersebut. Data yang telah diambil melalui proses *scraping* kemudian dilakukan ekspor data berupa file CSV.



Gambar 2. Data CSV Tim *English Premier League*

Data CSV diatas didapatkan dari hasil dan skema yang digambarkan pada panel *Selector Graph. Selector Graph* berbentuk seperti pohon yang memiliki cabang – cabang pada tiap sub-nya seperti tampak pada Gambar 3.

 ISSN (Online) 2828-786X



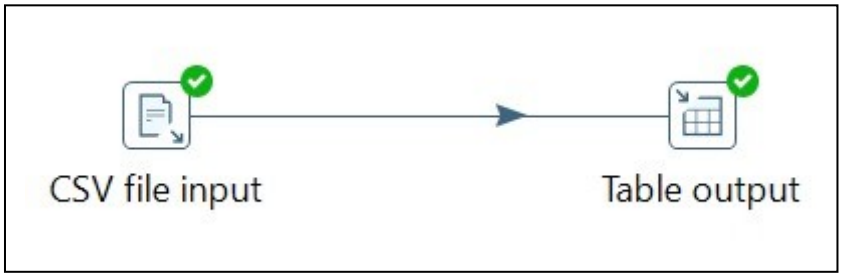
Gambar 3. *Selector Graph*

3.2 Hasil Proses ETL (*Extract Transform Load*) Menggunakan *Pentaho Data Integration*

Sebelum dilakukan proses ETL yaitu membuat *database* pada MySQL terlebih dahulu yang nantinya digunakan untuk menyimpan data hasil dari transformasi ETL yang bersumber dari data CSV.

Gambar 4. Struktur Tabel MySQL

Setelah *database* dibuat, proses ETL dapat dilakukan. Proses ETL ini terdiri dari satu *input* yaitu menggunakan *CSV File Input* dengan memasukkan data hasil *scraping* dan satu *output* berupa *Table Output* yang merujuk pada *database* yang telah dibuat sebelumnya. Hasil akhir dari proses transformasi ini yaitu data *scraping* yang telah masuk pada *database.* Dimana alur transformasi data ditunjukkan pada Gambar 5.

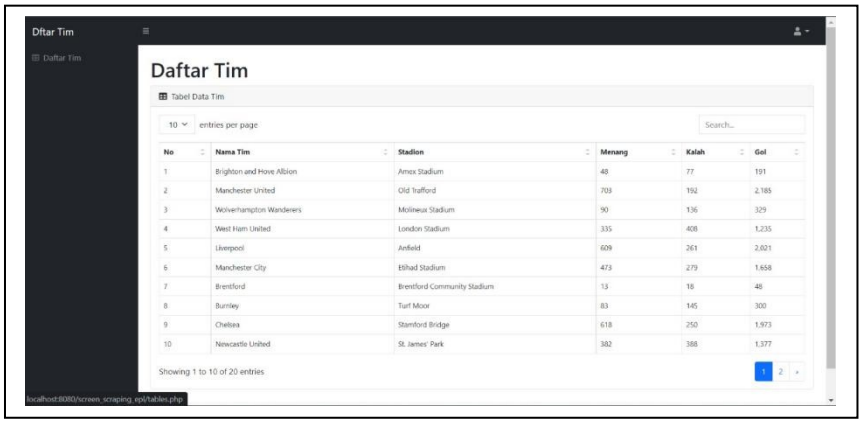


Gambar 5. Proses Transformasi ETL

3.3 Hasil Visualisasi

Data yang telah terdapat pada *database* selanjutnya dilakukan visualisasi dalam bentuk halaman *website* menggunakan *template bootstrap*. Data kemudian dikelompokkan dalam tabel berdasarkan nama tim, stadion, jumlah menang, jumlah kalah, dan total gol seperti tampak pada Gambar 6.

 ISSN (Online) 2828-786X



Gambar 6. Tampilan Website Data Tim *English Premier League*

4. KESIMPULAN DAN SARAN

Teknik *web scraping* dapat dengan mudah digunakan oleh semua orang tanpa harus menguasai bahasa pemrograman. *Web scraping* ini tidak akan mengganggu atau mengubah *database* dari situs *website* yang diambil datanya. Akan tetapi, *web scraping* memiliki batasan yaitu hanya dapat diterapkan pada sebagian situs *website*. Teknik ini juga tidak bisa digunakan untuk mengambil data yang diperlukan proses *login* didalamnya. Pada penelitian ini telah berhasil melakukan teknik *web scraping* untuk menampilkan tim sepak bola pada *website English Premier League.*
Dengan kemudahan akses data menggunakan teknik *web scraping*, maka sebaiknya teknik ini digunakan secara bertanggung jawab. Teknik ini pada dasarnya dilakukan dengan cara *copy paste* data yang diinginkan secara otomatis. Hal ini menyangkut hak cipta terhadap karya orang lain. Oleh karena itu, dalam penggunaannya teknik ini menjadi alternatif terakhir dengan mempertimbangkan hak cipta dan hak intelektual dari sumbernya.

5. DAFTAR RUJUKAN

[1] Utomo, W. U., 2013. *Integrasi Data Dengan Web Service.* 2nd ed. Salatiga: Fakultas Teknologi Informasi UKSW Press.
[2] Ayani, D. D., Pratiwi, H. S., Muhardi, H., Implementasi *Web Scraping* untuk Pengambilan Data pada Situs *Marketplace*. *Jurnal Sistem dan Teknologi Informasi,* 7 (4), pp. 257-262.
[3] Djufri, M., 2020. PENERAPAN TEKNIK WEB SCRAPING UNTUK PENGGALIAN POTENSI PAJAK(Studi Kasus pada Online Market Place Tokopedia, Shopee dan Bukalapak), *Jurnal BPPK,* 13 (2), pp. 65-75.
[4] Nisa, K., Sugiarto, D., Siswanto, T., 2021. Perancangan Data Warehouse Harga Pangan Di Wilayah Perumda Pasar Jaya, *Jurnal Sistem Informasi dan Telematika (Telekomunikasi, Multimedia dan Informatika),* 12 (1), pp. 47-55.
[5] Pratama, P. A. E., Widhiasih, N. P. N. D., 2020. PERANCANGAN DATA WAREHOUSE UNTUK PREDIKSI PENJUALAN PRODUK PADA ORBA EXPRESS MENGGUNAKAN PENTAHO, *(JUSS) Jurnal Sains dan Sistem Informasi,* 3 (2), pp. 43-48
[6] Setiawan, D. F., Tristiyato, Hijriani, A., 2020, APLIKASI WEB SCRAPING DESKRIPSI PRODUK. *Jurnal TEKNOINFO,* 14 (1), pp. 41-47.
[7] Utomo, D. W., Alsyah, G. R., Prabowo, D. A., 2018, INTEGRASI SISTEM INFORMASI AKADEMIK DAN E-LEARNING BERBASIS WEB. *Jurnal Sistem Informasi dan Bisnis Cerdas (SIBC),* 11 (1), pp. 18-24.
[8] Fatmasari, Kunang, Y. N., Purnamasari, S. D., Web Scraping Techniques to Collect Weather Data in South Sumatera, *INTERNATIONAL CONFERENCE ON ELECTRICAL ENGINEERING AND COMPUTER SCIENCE (ICECOS),* 2-4 October 2018, IEEE.

 ISSN (Online) 2828-786X

[9] Iwan Purnama & Ronal Watrianthos, 2018. *Sistem Informasi Kursus PHP dan MySQL,* Ponorogo: Uwais Inspirasi Indonesia.