

Perancangan Sistem Web Scraping untuk informasi kursus di Coursera dengan rekomendasi model dan metrik untuk pencarian kursus = Design of Web Scraping System for course information on Coursera with models and metrics recommendation for course search

Hadina Widyasari, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920565301&lokasi=lokal>

Abstrak

Skripsi ini membahas penelitian dan perancangan sebuah sistem web scraping untuk informasi kursus di Coursera dengan rekomendasi model dan metrik untuk pencarian kursus. Telah banyak tersedia layanan kursus online yang bisa dicari dan diakses sesuai dengan ketersediaan dan kebutuhan, seperti Coursera. Namun, ketika pengguna hendak mencari kursus dengan memasukkan teks ke dalam section penelusuran kursus Coursera, hasil pengurutan penelusuran bisa saja kurang sesuai dengan apa yang sebenarnya diharapkan atau dituju pengguna dengan keyword. Tujuan penelitian dan perancangan sistem ini adalah agar pengguna yang ingin mencari kursus di Coursera berdasarkan keyword dapat menggunakan sistem ini sebagai cara alternatif untuk memperoleh dan melihat gambaran urutan hasil penelusuran kursus-kursus yang tersedia di Coursera dengan akurasi kecocokan yang lebih tinggi sehingga diharapkan dapat menemukan kursus-kursus yang lebih sesuai dengan yang dicari melalui keyword atau teks. Penelitian dan perancangan sistem ini secara garis besar terdiri dari program scraper, database, pengujian model dan metrik, dan program pencarian kursus menggunakan perhitungan kemiripan antara teks dalam database dengan teks input dari user. Hasilnya, jika dibandingkan dengan metode yang digunakan Coursera untuk menampilkan hasil pengurutan kursus dengan pilihan sorting sudah berdasarkan “Best Match”, ternyata program pencarian kursus menggunakan soft cosine similarity dan cosine similarity yang dikombinasikan dengan all-MiniLM-L6-v2 telah menunjukkan peningkatan akurasi sebesar 33.33%.

.....This thesis discusses the research and design of a web scraping system for course information on Coursera with models and metrics recommendation for course search. Many online course services are available that can be searched and accessed according to availability and needs, such as Coursera. However, when users want to search for courses by entering text into the Coursera course search section, the search sorting results may not match what users actually expect or aim for with the keyword. The aim of this research and system design is for users who want to search for courses on Coursera based on keywords to use this system as an alternative way to obtain and see an overview of the order of search results for courses available on Coursera with higher accuracy of match so that it is expected to find more suitable courses through the keyword or text. This research and system design consists of a scraper program, a database, model and metric testing, and a course search program using similarity calculations between text in the database and user input text. As a result, compared to the method used by Coursera to display course sorting results with the sorting option already based on “Best Match”, it turns out that the course search program using soft cosine similarity and cosine similarity combined with all-MiniLM-L6-v2 has shown an accuracy increase of 33.33%.