

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi digital telah mendorong pertumbuhan pesat sektor e-commerce global. Berdasarkan laporan Statista (2024), nilai pasar e-commerce dunia diproyeksikan mencapai lebih dari 6,3 triliun USD pada tahun 2025, dengan peningkatan signifikan pada aktivitas perdagangan lintas negara, termasuk impor produk melalui platform daring. Kondisi ini membuka peluang besar bagi pelaku usaha impor di Indonesia untuk memperoleh produk langsung dari produsen luar negeri.

Salah satu platform yang banyak dimanfaatkan oleh pelaku usaha impor Indonesia adalah 1688.com, sebuah platform Business to Business (B2B) asal Tiongkok yang menyediakan jutaan data produk dari berbagai produsen. Penggunaan platform ini dapat diidentifikasi dari kemiripan foto katalog produk yang dijual di marketplace lokal Indonesia seperti Shopee dan Tokopedia dengan gambar yang terdapat pada halaman produk di 1688.com. Hal ini mengindikasikan bahwa sebagian besar pelaku usaha impor Indonesia menggunakan 1688.com sebagai sumber produk utama dan secara langsung memanfaatkan aset visual dari platform tersebut untuk keperluan penjualan lokal.

Meskipun menawarkan harga kompetitif dan variasi produk yang luas, pemanfaatan data produk dari 1688.com masih menghadapi kendala utama, yaitu seluruh antarmuka dan informasi produk disajikan dalam Bahasa Mandarin. Tantangan terbesar terletak pada konten visual produk, di mana spesifikasi teknis, fitur keunggulan, dan informasi penting lainnya sering disajikan dalam bentuk teks yang tertanam langsung di dalam gambar produk (in-image text). Teks yang tertanam dalam gambar tersebut tidak dapat dikenali atau diterjemahkan menggunakan ekstensi penerjemah konvensional seperti Google Translate Extension, karena ekstensi tersebut hanya bekerja pada teks HTML yang dapat dibaca secara struktural, bukan pada teks yang menjadi bagian dari elemen visual gambar.

Hambatan bahasa pada gambar produk ini secara langsung berdampak pada proses analisis produk (product analysis) yang menjadi tahap krusial sebelum pelaku usaha impor mengambil keputusan pembelian. Dalam praktik pengadaan produk impor, pelaku usaha perlu memahami secara menyeluruh keunggulan, spesifikasi teknis, dan diferensiasi produk yang umumnya disajikan dalam bentuk infografis pada gambar katalog 1688.com. Ketidakmampuan membaca informasi tersebut secara langsung menyebabkan pelaku usaha impor tidak dapat melakukan evaluasi kualitas produk secara efisien, sehingga berisiko mengambil keputusan pengadaan yang kurang tepat. Kondisi ini menjadi justifikasi utama kebutuhan akan sistem yang mampu menerjemahkan teks pada gambar produk secara otomatis dan

langsung di halaman 1688.com, sehingga proses product analysis dapat dilakukan dengan lebih cepat, akurat, dan efisien.

Akibat keterbatasan tersebut, pelaku usaha impor di Indonesia masih melakukan proses pencarian dan pemahaman informasi produk secara manual. Proses ini umumnya dilakukan dengan membuka halaman produk pada platform 1688.com satu per satu, kemudian mengunduh gambar produk dan menerjemahkan teks berbahasa Mandarin yang terdapat di dalam gambar menggunakan layanan penerjemah gambar yang hanya mampu memproses satu gambar dalam satu waktu. Proses yang berulang tersebut menyebabkan pelaku usaha membutuhkan waktu yang cukup lama untuk memahami spesifikasi dan informasi produk, terutama ketika harus menganalisis puluhan hingga ratusan produk sekaligus. Selain itu, konten pada platform 1688.com dimuat secara dinamis menggunakan teknologi Asynchronous JavaScript and XML (AJAX) serta arsitektur Single Page Application (SPA), sehingga metode pengambilan data berbasis HTTP request konvensional tidak efektif diterapkan pada platform tersebut.

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan perancangan ekstensi Chrome untuk melakukan web scraping data produk dari platform 1688.com yang dilengkapi dengan Optical Character Recognition (OCR) dan mekanisme image text swapping. Sistem dirancang untuk mengambil informasi produk secara otomatis, mengenali teks berbahasa Mandarin yang terdapat pada gambar produk menggunakan OCR, kemudian menerjemahkannya ke dalam bahasa Indonesia dan mengganti teks hasil terjemahan secara langsung pada tampilan gambar di halaman 1688.com. Dengan demikian, pelaku usaha impor dapat memahami informasi produk secara langsung tanpa harus berpindah ke aplikasi penerjemah maupun mengunduh gambar terlebih dahulu, sehingga proses analisis dan evaluasi produk menjadi lebih efisien sebelum keputusan pembelian dilakukan.

Untuk meningkatkan efisiensi proses scraping dan image text swapping, penelitian ini menerapkan Firefly Algorithm (FA) sebagai mekanisme optimasi pada pipeline scheduler. Algoritma ini digunakan untuk menentukan konfigurasi jumlah tugas yang dapat dijalankan secara bersamaan (*concurrency limit*) pada setiap tahapan pemrosesan, yang meliputi ekstraksi URL gambar dari *Document Object Model (DOM)* halaman 1688, pengunduhan gambar, *Optical Character Recognition (OCR)*, penerjemahan teks, dan *image text swapping*. Selama proses berlangsung, konfigurasi *scheduler* dapat disesuaikan secara adaptif berdasarkan kondisi performa sistem. Setiap konfigurasi dievaluasi menggunakan metrik kinerja seperti waktu penyelesaian proses (*makespan*), *throughput*, dan keseimbangan beban antar tahapan (*pipeline balance*), sehingga diperoleh konfigurasi scheduler yang mampu meningkatkan efisiensi proses scraping dan pemrosesan gambar. Selain itu, sistem juga mengintegrasikan Gemini API untuk menghasilkan

ringkasan deskripsi produk berbahasa Indonesia secara otomatis berdasarkan hasil ekstraksi OCR. Integrasi antara *web scraping*, *OCR*, *image text swapping*, *Firefly Algorithm*, dan kecerdasan buatan diharapkan mampu membantu pelaku usaha impor memahami informasi produk pada platform 1688.com secara lebih cepat, akurat, dan efisien.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana merancang ekstensi Chrome yang mampu melakukan v data foto produk dari platform 1688 secara otomatis?
2. Bagaimana mengimplementasikan image text swapping berbasis OCR untuk melakukan penerjemahan dan penggantian teks berbahasa Mandarin pada gambar produk secara langsung di halaman website 1688?
3. Bagaimana menerapkan Firefly Algorithm untuk mengoptimalkan penjadwalan proses scraping dan image text swapping pada sistem yang dikembangkan?

1.3 Batasan Masalah

Agar penelitian lebih terarah, batasan masalah ditetapkan sebagai berikut:

1. Platform yang digunakan sebagai sumber data adalah 1688.com.
2. Proses scraping dibatasi pada pengambilan informasi produk yang meliputi nama produk, harga, gambar produk, informasi toko, URL produk, serta data pendukung yang tersedia pada halaman produk.
3. OCR hanya diterapkan pada gambar produk yang memiliki teks berbahasa Mandarin atau Inggris.
4. Penerjemahan teks dilakukan menggunakan Google Translate API dengan bahasa Indonesia sebagai bahasa utama keluaran sistem.
5. Firefly Algorithm hanya diterapkan sebagai mekanisme optimasi Pipeline Scheduler dan tidak digunakan untuk meningkatkan akurasi OCR, hasil translasi, maupun kualitas image text swapping.
6. Sistem dikembangkan dalam bentuk Chrome Extension berbasis Manifest V3 serta Web Dashboard.
7. Sistem hanya dapat dijalankan pada browser Google Chrome yang mendukung Chrome Extension Manifest V3.
8. Aspek keamanan dan autentikasi akun pengguna tidak menjadi fokus utama penelitian.
9. Web Dashboard hanya digunakan untuk menampilkan, mencari, dan mengelola hasil scraping serta hasil pengolahan gambar.

10. OCR hanya diterapkan pada gambar produk dan tidak mencakup teks yang berada di luar area gambar.
11. Gemini AI hanya digunakan untuk menghasilkan ringkasan deskripsi produk dan tidak digunakan pada proses scraping, OCR, maupun penerjemahan gambar.
12. Pengambilan data dilakukan menggunakan DOM Scraping terhadap halaman yang telah selesai di render oleh browser.
13. User Acceptance Testing (UAT) dibatasi pada evaluasi penerimaan pengguna terhadap fitur-fitur utama sistem dan tidak mencakup pengujian keamanan, kompatibilitas lintas browser, maupun performa jangka panjang.

1.4 Tujuan

Tujuan dari penelitian ini adalah sebagai berikut:

1. Merancang dan membangun ekstensi Chrome yang mampu melakukan web scraping data produk dari platform 1688 secara otomatis.
2. Mengimplementasikan fitur image text swapping berbasis OCR untuk melakukan penerjemahan dan penggantian teks Mandarin pada gambar produk secara langsung di website 1688.
3. Menerapkan Firefly Algorithm untuk mengoptimalkan penjadwalan proses scraping dan image text swapping pada sistem yang dikembangkan.

1.5 Manfaat Penelitian

Bagi Peneliti

1. Meningkatkan pemahaman dalam penerapan Firefly Algorithm sebagai mekanisme optimasi penjadwalan (*scheduler*) pada sistem berbasis Chrome Extension yang terintegrasi dengan teknologi OCR dan kecerdasan buatan.
2. Menjadi pengalaman praktis dalam pengembangan perangkat lunak berbasis Chrome Extension yang mengintegrasikan Optical Character Recognition (OCR), Firefly Algorithm, Image Text Swapping, serta Large Language Model (LLM) Gemini.

Bagi Akademisi

1. Menjadi referensi bagi penelitian selanjutnya mengenai optimasi proses web scraping, OCR multibahasa, serta pengembangan sistem berbasis Chrome Extension yang terintegrasi dengan kecerdasan buatan.
2. Memberikan kontribusi bagi pengembangan penelitian di bidang rekayasa perangkat lunak (*software engineering*), kecerdasan buatan (*artificial intelligence*), dan pengolahan citra digital (*image processing*).

Bagi Praktisi dan Industri

1. Membantu pelaku usaha atau importir dalam mempercepat proses riset penerjemahan dan pembuatan kebutuhan katalog produk multibahasa.

1.6 Waktu dan Tempat Penelitian

1.6.1 Waktu Penelitian

Penelitian ini akan dilaksanakan selama 10 bulan yaitu dari bulan September 2025 sampai bulan Juni 2026.

1.6.2 Tempat Penelitian

Penulis akan melakukan penelitian pada Fakultas Ilmu Komputer UPN “Veteran” Jakarta yang terletak di Jalan Rumah Sakit Fatmawati Raya, Pondok Labu, Cilandak, Jakarta Selatan, Jakarta.

1.7 Sistematika Penulisan

Penulisan laporan penelitian ini disusun dengan sistematika sebagai berikut:

BAB I PENDAHULUAN

Menjelaskan latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat, waktu, tempat, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Berisi teori-teori terkait seperti Web Scraping, OCR, Firefly Algorithm, Chrome Extension Development, serta penelitian terdahulu yang relevan.

BAB III METODOLOGI PENELITIAN

Menjelaskan tahapan penelitian, perancangan sistem, arsitektur teknis, algoritma yang digunakan, serta alat dan bahan penelitian.

BAB IV HASIL DAN PEMBAHASAN

Menampilkan implementasi sistem, hasil pengujian, serta analisis performa dan efektivitas optimasi.

BAB V PENUTUP

Menyajikan kesimpulan dan saran pengembangan lebih lanjut.