

ABSTRAK

Sepatu merupakan salah satu kebutuhan sandang manusia. Seperti halnya pakaian, penggunaan sepatu juga harus melihat tempat kegiatan, dan kategori (model/jenis sepatu). Untuk membeli sepatu terjadi perubahan aktivitas dan gaya hidup manusia dalam pembelian sepatu. Teknologi Internet menggeser pola belanja masyarakat di Indonesia, sebelumnya belanja atau membeli sepatu di toko secara langsung, sekarang sudah dimudahkan dengan adanya *e-commerce* yang dapat membuat masyarakat berbelanja secara online. Tidak semua sepatu memiliki gerai atau toko online sendiri, sehingga mereka menjualnya lewat platform Shopee, Lazada Indonesia, Bukalapak, Tokopedia dsb.

Perkembangan *e-commerce* yang semakin pesat menyebabkan manusia untuk meminta atau mendapatkan servis yang lebih. Seperti halnya mencari sebuah sepatu lewat shopee dengan bukalapak tentunya berbeda harga dan *output* yang ditampilkan. Maka dari itu dibutuhkan sebuah *web service* yang mengumpulkan data-data dari web *e-commerce*. Teknik yang digunakan untuk mengumpulkan data-data dari web adalah *crawling*. *Crawling* akan mengambil data sesuai dengan URL dan kata kunci. Web yang akan di *crawling* yaitu Shopee, Bukalapak, dan Lazada Indonesia. Algoritma yang dipakai untuk proses penelusuran pada studi kasus *crawling* yaitu Algoritma *Depth First Crawling*. Algoritma ini akan melakukan penelusuran data secara mendalam pada satu link website sampai berhasil menemukan data sesuai kata kunci kemudian menelusuri pada link website berikutnya. Pada proses *crawling* memanfaatkan *library* CURL pada PHP untuk mengunjungi halaman sebuah website dan *Reguler Expression* untuk *parsing* data atau pengambilan data yang dibutuhkan. Setelah mendapatkan data, selanjutnya melakukan proses kategorisasi menggunakan *library natural processing tools*. Data yang akan dikategorikan yaitu berdasarkan pria, wanita, dan general.

Hasil penelitian yang diperoleh yaitu aplikasi web service ini mampu menampilkan data-data sepatu dari tiga e-commerce sekaligus tanpa harus mengunjungi website e-commerce satu persatu menggunakan teknik *crawling* dan algoritma *depth first crawling*. Rata-rata waktu yang diperlukan untuk melakukan penelusuran dari algoritma *depth first crawling* yaitu 15,70 detik. Proses kategorisasi menggunakan *library natural language processing tools* mendapatkan akurasi 96.35.

Kata kunci : Sepatu, *E-commerce*, *Web Service*, *Crawling*, Algoritma *depth first crawling*, *CURL*, *Reguler Expression*, *Natural Language Processing Tools*.