

ABSTRAK

Pemerintah Sumatera Barat mempunyai banyak website pemerintahan yang mengandung banyak sekali informasi yang sangat berguna bagi warganya, akan tetapi masing-masing website ini belum mempunyai fitur pencarian yang powerfull dan masih banyak yang belum terindeks dengan baik di mesin pencarian populer. Penelitian ini bertujuan untuk menampilkan semua hasil yang relevan antara kata kunci dan konten yang terkandung pada semua website pemerintah Sumatera Barat yang telah diindeks dengan teknik crawling. Konten website yang dikumpulkan bersumber dari semua website yang dikelola pada server Dinas Komunikasi dan Informatika Provinsi Sumatera Barat. Sebuah engine akan dijalankan untuk melakukan crawling pada semua website tersebut. Selanjutnya konten yang didapat dibersihkan terlebih dahulu dari karakter yang tidak diperlukan dengan bantuan regex dan kemudian akan disimpan pada database. Ketika pengguna melakukan pencarian dengan memasukkan kata kunci, sistem akan menggunakan regex untuk mengekstraks setiap kata dan akan mencocokkan dengan akurat pada semua row di database. Setelah ditemukan konten yang sesuai, sistem akan menghitung score berapa banyak kata kunci yang terkandung dalam row tersebut dan mengurutkannya berdasarkan score tertinggi. Hasil dari pengujian ini adalah sistem mampu menampilkan semua konten yang memiliki kata kunci yang dicari dengan akurat. Penelitian ini dapat membantu pemerintah dan warga Sumatera Barat dalam membagikan dan mendapatkan informasi yang diperlukan secara cepat dan presisi.

Kata kunci : *Search Engine, Crawling, Regex, Global Search, Content Extractin*