

Abstrak

Mesin pencari merupakan sebuah aplikasi yang mampu mengindeks berbagai URL berada di internet sehingga mempunyai *database* besar tentang web lengkap dengan deskripsinya. Sedangkan aplikasi “mesin pencari teks pada gambar” ini melakukan pencarian yang lebih spesifik, yaitu teks yang terdapat pada gambar. Proses kerja dari mesin ini dapat dibagi menjadi 3 bagian, yaitu *crawling*, *text recognition*, dan *searching*.

Proses *crawling* bertugas menelusuri URL, mencari, dan men-*download* gambar yang berawal dari *feed* URL. Di sisi lain, proses *text recognition* bekerja paralel dengan proses *crawling* ini. Gambar yang telah diperoleh, diproses secara sekuensial. Proses *text recognition* bertugas melakukan ekstraksi teks yang terdapat dalam gambar. Kemudian teks tersebut diubah menjadi *term-term*, disimpan, dan di-indeks dalam *database*. Aplikasi yang berhubungan langsung dengan *user* yaitu *subsystem search*. Di sini, pengguna memasukkan teks pada *interface* yang telah disediakan, lalu mesin akan mencari teks ke *database* sesuai dengan pola teks inputan. Kemudian file gambar dari data yang bersesuaian akan ditampilkan.

Keseluruhan sistem ini mampu menyediakan mesin pencari dengan fitur *similar match* yang mencari kata sama dengan inputan user, dan *exact match* yang mencari kata mirip dengan inputan user. Fitur ini disediakan karena hasil ekstraksi teks tidak selalu tepat dengan teks sesungguhnya pada gambar.

Kata kunci : *crawling*, *text recognition*, *searching*, *term*