

ABSTRAK

Pencarian informasi dapat diperoleh dari internet dengan mudah dan cepat melalui halaman web. Web memuat banyak informasi yang dihasilkan dari waktu ke waktu secara kontinu dari berbagai sumber. Jumlah informasi yang terus bertambah dapat menyulitkan para pencari informasi dalam menemukan informasi yang relevan. Salah satu cara yang dirasa efektif untuk menyelesaikan permasalahan ini adalah dengan melakukan klasifikasi halaman web menurut topiknya.

Salah satu metode untuk mengklasifikasikan halaman web berdasarkan topiknya yaitu *categorization by context*. Metode *categorization by context* mengekstrak informasi yang berguna dari sebuah dokumen halaman web untuk mengklasifikasi dokumen dimana URL muncul sebagai rujukannya. Metode ini dirasa efektif karena tidak mengkategorisasikan dokumen pada halaman web menurut isinya namun berdasarkan *URL terkait* dan *tag<title>* yang ada pada suatu dokumen.

Sebuah sistem berbentuk sebuah katalog diimplementasikan yang isinya terdiri dari kategori-kategori dari halaman web dimana kategori ditentukan oleh penulis sendiri. Pada implementasi ini dokumen-dokumen diklasifikasikan ke dalam kategori tertentu berdasarkan tag <title> dan URL terkaitnya yang kemudian dicocokkan dengan parameter setiap kategori yang didapatkan. Hasil yang diperoleh menunjukkan sebagian besar dokumen terklasifikasi sesuai dengan kategori prediksinya.

Kata kunci : *categorization by context*, *URL terkait*, *tag<title>*