

ABSTRAK

Seiring dengan semakin berkembangnya dunia internet, semakin banyak pula web berita online yang tersedia di dunia maya. Pada dasarnya web berita online banyak menyediakan informasi penting. Berbagai teknik dalam *text mining* dapat diterapkan dengan tujuan untuk memperoleh manfaat yang lebih banyak dari informasi yang disediakan, diantaranya yaitu dengan menggunakan *page-based clustering* maupun *keyword-based search*. Namun, *page-page* pada *newspaper* biasanya terdiri dari beragam *item* berita dengan topik yang saling tidak berhubungan satu sama lain, sehingga *page-based clustering* kurang memberikan hasil yang optimal.

Pada Tugas Akhir ini dilakukan pendekatan dengan melakukan ekstraksi terhadap item-item berita pada *web pages* secara individual dan melakukan *mining* secara terpisah dengan menggunakan *pattern based strategy*. Pendekatan ini menggunakan *pattern URL text* dan *anchor text* dalam mengekstrak link berita serta menggunakan crawler untuk penelusuran link berita dalam rangka mencari *full story* dari masing-masing item berita.

Tahap analisis dan pengujian memberikan hasil bahwa pendekatan *pattern based strategy* yang dibangun terbukti dapat mengekstrak *full story* pada halaman Web berita meskipun tidak semua *full story* dari setiap item berita dapat diekstrak. Hasil ekstraksi item berita akan mencapai nilai optimal jika link-link pada web input bersifat homogen.

Kata Kunci : *pattern based strategy* , *URL text* , *anchor text*, web berita, *full story*.