

ABSTRAK

Nama : Derian Adlan (140 2015 120)
Program Studi : S-1 Teknik Informatika
Judul Skripsi : Membangun *Web Crawler* Menggunakan Metode *Incremental* serta Tinjauannya Menurut Agama Islam
(Studi Kasus: Situs *Web* Kesehatan Indonesia)

Internet merupakan sumber informasi terbesar dan terkaya di dunia serta mengalami tren pertumbuhan informasi yang eksponensial setiap tahunnya. Namun di internet, ketersediaan korpus kesehatan berbahasa Indonesia masih sangat terbatas. Dengan internet sebagai sumber informasi terbesar di dunia, hal tersebut dapat dimanfaatkan untuk membuat dan menyediakan korpus kesehatan berbahasa Indonesia dengan mengambil dokumen *web* di internet. Pada penelitian ini akan membangun *web crawler* yang bernama MyCrawler. MyCrawler mempunyai tujuan untuk mengambil dokumen *web* kesehatan di internet dengan menggunakan metode *incremental*. MyCrawler mempunyai tiga tahapan utama yaitu tahap *scraping*, tahap *crawling* dan tahap *incremental*. Tahap *scraping* bertujuan untuk mendapatkan *seed url* berdasarkan sepuluh kata topik kesehatan terbaik, tahap *crawling* bertujuan untuk mengunduh dan mengambil dokumen *web* di internet, dan tahap *incremental* bertujuan untuk memperbarui dokumen *web* yang berada di basis data. MyCrawler lebih baik dibandingkan dengan aplikasi sejenis yaitu HTTrack karena dapat mengambil dokumen *web* lebih banyak dengan durasi 80% lebih cepat dan dapat menentukan dokumen *web* yang akan diperbarui dengan lebih efektif. Membangun *web crawler* menurut Agama Islam merupakan hal yang diperbolehkan.

Keyword: Korpus Kesehatan Indonesia, *Web Crawler*, *Incremental Crawler*