

NEWS WEB CRAWLER DENGAN METODE XPATH

Oleh : Fajri Abdillah

Pembimbing I : Ichsan Taufik, M.T.

Pembimbing II : Jumadi, S.T, M.Cs

ABSTRAK

Data yang tersebar di dunia Internet itu sangat beragam dan penuh dengan *noise*. Seperti konten berita yang diselingi dengan banyak iklan, sehingga mengganggu *user* dalam membaca konten berita. Data berita (*News*) juga sangat cepat dalam hal *publish*, dalam 20 menit saja bisa ter-*publish* lebih dari 50 berita baru. Data berita tersebut dapat digunakan dalam bermacam – macam hal, salah satunya adalah untuk *Social Media Analytic Service*. Kecepatan penangkapan data (*Crawl*), pemrosesan data (*Scraping*) menjadi salah satu hal yang sangat penting, agar tidak terjadi data yang kurang sempurna, dan data yang diterima adalah data yang paling baru. Untuk itu, maka dibuatlah aplikasi yang dapat menangkap data – data berita tersebut hingga semua data tidak ada yang terlewat. Aplikasi ini dibuat dengan bahasa pemrograman *Python* untuk menangkap data – data berita tersebut, menggunakan *Web Framework Django* untuk tampilan berbasis *Website* dan pemodelannya menggunakan UML dengan *tool* Star UML. Aplikasi ini akan digunakan oleh PT Mediawave Analytic sebagai *database news*. *Database* yang sudah ada yaitu *Facebook* dan *Twitter*. *Database* yang sudah ada dapat membantu dalam proses *analytic* kepada *client* yang berguna untuk *brand* yang dipergunakan oleh *client*. *Client* akan menerima *report* yang berguna dalam pengambilan keputusan, dengan menyertakan hasil *analytic* dari *Facebook*, *Twitter* dan *News*.

Kata kunci: *Crawl*, *Scraping*, *Python*, *Django*, *News*

The logo of Universitas Islam Negeri Sunan Gunung Djati Bandung, featuring the letters 'UIN' in a stylized, lowercase font.

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
BANDUNG



uin

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
BANDUNG