1. Pendahuluan

Media sosial adalah aplikasi yang berbasis internet dimana para penggunanya bisa menciptakan atau saling menukar konten dengan pengguna lainnya[1]. Media sosial saat ini telah mencakup berbagai hal dan menjadi sebuah sarana yang sangat direkomendasi dalam berbagai bidang.

Dengan jumlah akun yang begitu banyak maka dalam penelitian ini akan dicari akun media sosial apa saja yang dimiliki oleh seorang pengguna menggunakan *search engine*. *Search engine* atau mesin pencari merupakan mesin telusur otomatis yang mengandalkan pencocokan kata kunci yang kemudian diolah dalam kueri yang kemudian hasilnya kemiripannya akan ditampilkan hasilnya di *browser*[2]. Pada umumnya hasil pencarian pada media sosial hanya menampilkan pencarian dari media sosial itu sendiri, namun hasil yang ditampilkan oleh aplikasi akan lebih efektif karena aplikasi menampilkan hasil pencarian dari empat media sosial sekaligus.

Sistem dirancang untuk melakukan proses *crawling* menggunakan struktur DOM dan berikutnya dilakukan *Brute Force* yang dipadukan dengan SQL untuk mencocokkan kata kunci yang sudah dimasukkan ke mesin pencari dengan data hasil *crawling* yang sudah dimasukkan ke dalam *database*. Berdasarkan riset[3,4] mengenai mesin pencari yang menggunakan metode *Brute Force*, riset tersebut melakukan pencarian dengan memeriksa kecocokan perhuruf dari kata yang dimasukkan ke mesin pencari ke seluruh data. Namun pada penelitian ini penulis mencoba untuk menyederhanakan skema pencocokan *Brute Force* yang dipakai dalam aplikasi untuk mempersingkat waktu yang diperlukan dalam proses pencocokan sehingga saat proses pencocokan sistem hanya perlu mencocokkan data dengan atribut yang ditentukan dan tidak memeriksa atribut lainnya yang tidak berhubungan dengan data.

Batasan masalah dalam topik ini adalah :

1. *Input* data yang diperlukan adalah nama, kota asal/tinggal, tempat kerja, dan status verifikasi dari akun tersebut. *Input* data ini dipilih berdasarkan informasi yang bisa diambil dari detail akun Facebook dan detail dari data *input* tersebut juga akan digunakan untuk menyaring data *crawling* yang akan dimasukkan ke *database*.
2. Status akun saat *crawling* dibagi menjadi dua yaitu akun terverifikasi dan belum terverifikasi.
3. Media sosial yang akan diambil datanya adalah Facebook, Twitter, Instagram dan Google+ untuk status akun yang belum terverifikasi. Untuk akun yang terverifikasi media sosial yang akan diambil datanya untuk ditaksir adalah Facebook, Twitter dan Instagram. Penentuan media sosial tersebut dilakukan karena beberapa media sosial lainnya ditemui kendala saat melakukan proses *crawling* seperti masalah session yang diperlukan pada halaman, atau halaman mengidentifikasi alamat IP sebagai penyerang karena melakukan *crawling* menggunakan sintaks pemrograman, atau halaman membatasi akses halaman. Jumlah media sosial yang akan dicrawl juga nantinya akan menentukan waktu yang diperlukan aplikasi untuk melakukan *crawling*.
4. Nama yang dimasukkan dalam *input* data bukanlah nama yang terlalu umum. Batasan ini digunakan agar hasil pencarian di media sosial tidak terlalu banyak. Jika nama yg digunakan terlalu umum, maka akan menyebabkan ketepatan data yang dicari akan berkurang.
5. Saat melakukan *crawling* Facebook, akun yang dituju harus membuka akses pencarian dari perangkat luar dan memiliki setidaknya 1 detail sehingga dapat mempermudah proses pencarian.
6. Jumlah akun yang dimunculkan aplikasi untuk halaman Facebook maksimal 20 akun dan untuk Twitter maksimal 18 akun. Jumlah akun ini berdasarkan jumlah default yang ditampilkan halaman saat pertama kali halaman dimuat dan mempermudah proses *crawling*.
7. *Crawling* data terhadap media sosial dilakukan sebelum 25 Mei 2018.

Adapun tujuan dari Tugas Akhir ini dibuat berdasarkan topik dan pembahasan yang ditulis. Tujuan dari Tugas Akhir ini adalah :

1. Memahami cara mengimplementasikan algoritma *Brute Force* untuk melakukan pencarian pada data yang diambil dari media sosial.
2. Memahami cara melakukan *crawling* pada media sosial menggunakan struktur DOM.
3. Mengetahui atribut apa saja yang digunakan untuk menaksir akun media sosial milik seseorang.