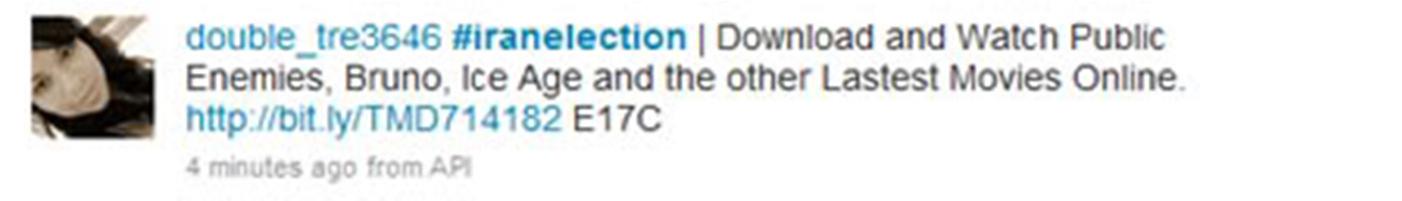
# BAB I DESKRIPSI MASALAH

*Electronic-spam* merupakan pesan elektronik yang tidakdiinginkan penerimanya, bisa dalam bentuk surat elektronik, SMS, *posting* atau komentar di media sosial yang muncul di *timeline* kita, ataupun pesan pada *chat-messenger*. *Spammer* melakukan *spamming* untuk tujuantertentu, paling banyak untuk menyebarkan iklan. Penentuan *spam* sangat bersifat subjektif, artinya *spam* untuk kita, belum tentu spam untuk pengguna lain. Gambar 1.1 dan 1.2 merupakan contoh *spam* pada *twitter* dan *facebook*.



(a)



(b)

https://econsultancy.com/blog/4218-the-seven-twitter-sins-a-comprehensive-list-of-twitter-spam-techniques

**Gambar 1.1** Contoh spam pada media sosial*twitter*:

(a) hashtag/trending topic spam; (b) autoresponder spam



https://m.facebook.com/security/photos/a.36604031885.58084.31987371885/36604061885/?type=3&source=43

**Gambar 1.2** Contoh*spam*pada media sosial*facebook*

Algoritma pencocokan string (*pattern*) Knuth-Morris-Pratt (KMP) dan Algoritma Boyer-Moore merupakan algoritma yang lebih baik daripada *brute force*. Pada Tugas Besar III kali ini Anda diminta membuat aplikasi sederhana deteksi *spam* pada media sosial dengan kedua algoritma tersebut, plus menggunakan *regular expression* (*regex*). Teks yang akan Anda proses adalah *posting* berbahasa Indonesia. Pengguna aplikasi ini akan memberikan masukan berupa *keyword* spam, dan menandai daftar *posting* yang dikategorikan *spam* terurut berdasarkan tanggal.

Pencocokan string yang anda buat adalah *exact matching* (untuk KMP dan BM) jadi *posting* yang diproses mengandung string yang tepat sama dengan keyword spam daripengguna. Sedangkan bila menggunakan *regex* maka tidak selalu *exact matching*. Pencarian juga tidak bersifat *case sensitive*, jadi huruf besar dan huruf kecil dianggap sama (hal ini dapat dilakukan dengan mengganggap seluruh karakter di dalam pattern dan teks sebagai huruf kecil semua atau huruf kapital semua).

# BAB II DASAR TEORI

Di zaman sekarang ini, penggunaan media sosial telah menjadi bagian dari hidup manusia. Setiap hari manusia menggunakan media sosial sebagai salah satu sarana yang paling sering digunakan untuk berkomunikasi dengan orang lain. Selain untuk berkomunikasi dengan orang lain, banyak juga media sosial yang menyediakan layanan-layanan lain seperti melakukan *posting*, baik berupa gambar, teks, bahkan video.

Layanan ­*posting* di media sosial tak jarang digunakan oleh orang-orang. Biasanya terdapat 2 jenis *posting* yang dapat kita lakukan di media sosial, yaitu *public* dan *private*. *Posting*-an yang bersifat *public* dapat dibaca oleh semua orang, termasuk orang-orang yang belum menjadi teman kita di media sosial yang bersangkutan.Layanan ini sering digunakan oleh orang-orang untuk mempublikasikan sesuatu yang layak untuk dibaca oleh orang banyak. Contohnya perusahaan berita melakukan *public posting* untuk mempublikasikan berita-berita baru, instansi pemerintahan untuk mengumumkan adanya infrastruktur baru atau perbaikan infrastruktur yang telah ada.

Dalam dunia media sosial, dikenal suatu istilah *spam*. *Spam* dapat berarti *message* yang tidak relevan atau tidak pantas, *message* sama yang dikirim berulang-ulang, atau iklan yang tidak diinginkan dan mengganggu. Dalam *public posting* itu sendiri, *spam* dapat berarti *posting* yang memiliki isi yang sama yang dilakukan berulang-ulang, atau *posting* berisi iklan yang tidak diinginkan. *Spamming* tidak jarang dilakukan orang-orang dalam *public posting*. Bahkan suatu akun seringkali melakukan *posting* berulang yang memiliki isi yang sama.

*String matching* (pencocokan *string*) merupakan suatu kakas yang dapat digunakan untuk mendeteksi *spam* pada *posting-posting* di media sosial. *String matching* itu sendiri merupakan kakas yang dapat digunakan untuk mendeteksi apakah suatu kata atau kumpulan kata terdapat di dalam kumpulan kata lainnya.

Pada umumnya, *posting-posting* yang memiliki isi yang sama memiliki suatu kata kunci yang sama untuk setiap *post*. Contohnya pada media sosial *Twitter*, orang sering menggunakan tagar yang sama untuk *posting-posting* dengan isi yang sama. Kakas *string matching* dapat digunakan untuk mendeteksi *posting-posting* yang berupa *spam* dengan mencocokkan suatu kata kunci di dalam *posting-posting* tersebut. Apabila kata kunci yang bersangkutan terdapat dalam suatu *post*, maka *post* tersebut dapat dikatakan merupakan *spam*.