

ABSTRAK

Beatrice Kerenhapukh Handoko

PEMBLOKIRAN HALAMAN WEBSITE PORNO BERDASARKAN TEXT CONTENT MELALUI BROWSER EXTENSION DAN WEB SERVICE

(xiv + 90 halaman; 24 gambar; 12 segmen program; 29 tabel)

Keberadaan internet membuka gerbang bagi semua jenis informasi untuk seluruh lapisan masyarakat. Namun, tidak semua informasi tersebut adalah informasi yang baik dan bermanfaat, tetapi bisa juga merupakan informasi yang merusak moral seperti pornografi. Oleh karenanya, bahaya pornografi dapat mempengaruhi anak-anak yang telah mengenal dunia internet sejak kecil. Dibutuhkan sebuah sistem yang dapat mengenali dan memilah halaman website mana yang layak dilihat oleh mereka. Halaman yang mengandung pornografi harus segera diblokir.

Sistem ini dibuat dengan menerapkan konsep *information retrieval* dan metode klasifikasi Naive Bayes serta memanfaatkan keberadaan browser extension dan web service. Suatu halaman website akan melalui tahap-tahap *cleaning*, tokenisasi (*tokenization*), *stop word*, *stemming*, serta pengelompokan pembobotan. Semua tahap tersebut akan menghasilkan kata kunci untuk selanjutnya diproses dengan perhitungan Naive Bayes. Pada akhirnya akan dihasilkan suatu kesimpulan apakah halaman tersebut mengandung pornografi atau tidak. Setelah itu, halaman yang termasuk kategori porno akan diblokir.

Berdasarkan hasil uji coba yang dilakukan, disimpulkan bahwa sistem ini dapat mengklasifikasikan dan memblokir halaman website yang mengandung pornografi. Penelitian ini menguji 120 halaman website, dengan 4 set pelatihan dan uji coba. Rata-rata tingkat akurasi semua set tersebut adalah sebesar 89,58%.

Kata kunci: pornografi, blokir, web service, browser extension, *information retrieval*, Naive Bayes.

ABSTRACT

Beatrice Kerenhapukh Handoko

BLOCKING PORN WEBSITE PAGES TEXT CONTENT BASED WITH BROWSER EXTENSION AND WEB SERVICE

(xiv + 90 pages; 24 pictures; 12 program segments; 29 tables)

The existence of internet opens the gates to all kinds of information for all society levels. Unfortunately, not all of those information are good and useful ones, but could also be morally damaging such as pornography. Therefore, the danger of pornography could affect children that had use internet since their childhood. This requires a system that could identify and pick which website page that are worth seeing by them. Web page that contains pornography should be blocked right away.

This system is created by implementing the concept of information retrieval and classification method Naive Bayes, and also utilizing the existence of browser extension and web service. A web page would go through several phase of cleaning, tokenization, stop word, stemming, and grouping-weighting. All these steps will generate key words to be further processed by Naive Bayes calculation. This will ultimately reach a conclusion whether the page contains pornography or not. After that, the page that being categorized as porn will be blocked.

Based on the results of experiments performed, it was concluded that this system can classify and block pages that contain pornography. This study examined 120 pages, with 4 sets of training and testing. The average level of accuracy for all training and testing set is 89.58%.

Keywords: pornography, block, web service, browser extension, information retrieval, Naive Bayes.