

ABSTRAK

Nama : Danang Setiawan
Program Studi : Teknik Informatika
Judul : Sistem Pemeringkat Konten Suara Berdasarkan Bahasa Kasar Menggunakan Pendekatan NLP (*Natural Language Processing*)
Dosen Pembimbing : Dino Hariatma Putra M. Kom

Bahasa Kasar merupakan ekspresi yang berisi kata kata kasar atau ungkapan yang tidak senonoh. Penyebaran Bahasa kasar dapat dilakukan melalui berbagai media, salah satu contohnya yaitu melalui tontonan atau siaran. Perkembangan teknologi yang sangat pesat membuat tontonan dapat dengan mudah ditemui diberbagai media sosial. Hampir semua orang mulai dari anak kecil hingga orang tua sudah menggunakan media sosial dan memiliki akun pribadi. Hal tersebut membuat penerimaan konten negatif, konten yang berisi bahasa kasar dan konten yang tidak sesuai dengan umur menjadi lebih mudah, konten yang diterima dapat berbentuk konten video maupun konten suara. Masalah ini menjadi dasar dibutuhkannya sistem pemeringkat konten agar dapat terhindar dari konten yang mungkin berisi bahasa kasar maupun tidak sesuai umur. Salah satu yang dapat dilakukan adalah membuat sistem pemeringkat konten suara yang dapat memberikan rating umur pada konten berdasarkan bahasa kasar yang terdapat didalamnya. Proses dilakukan dengan beberapa tahap yaitu transkripsi teks untuk mendapatkan teks hasil transkripsi, *Text Preprocessing* untuk merapihkan teks agar mudah diolah, dan deteksi kata kasar pada teks untuk mendapatkan hasil klasifikasi pemeringkat konten suara. Pada proses transkripsi teks menggunakan *Speech Recognition Google Speech API* untuk pendeteksi suara dan proses *Text Preprocessing* menggunakan berbagai teknik mulai dari Case Folding hingga Stemming dengan library sastrawi yang menerapkan algoritma Nazief & Adriani. Hasil implementasi didapatkan bahwa sistem pendeteksi suara dapat diterapkan dengan baik pada pemeringkat konten dan membantu penonton terhindar dari konten negatif, konten berisi kata kasar, dan konten yang tidak sesuai umur.

Kata kunci : *Bahasa Kasar, Pemeringkat Konten, Speech Recognition,*

ABSTRACT

Crude Language is an expression that contains abusive words or expressions that are indecent. The spread of abusive language can be done through various media, one example is through spectacle or broadcast. The rapid development of technology makes the spectacle can be easily found on various social media. Almost everyone from small children to the elderly are already using social media and have personal accounts. This makes it easier to accept negative content, content that contains abusive language and content that is not age-appropriate, the content received can be in the form of video content or sound content. This problem is the basis for the need for a content rating system to avoid content that may contain abusive or age-appropriate language. One thing that can be done is to create a sound content rating system that can provide an age rating on content based on the abusive language contained in it. The process is carried out in several stages, namely text transcription to get the transcription text, Text Preprocessing to tidy up the text to make it easy to process, and detect harsh words in the text to get the content rating classification results. In the text transcription process using the Speech Recognition Google Speech API for voice detection and the Text Preprocessing process using various techniques ranging from Case Folding to Stemming with a literary library that applies the Nazief & Adriani algorithm.. The implementation results show that the sound detection system can be applied well to content rating and helps viewers avoid negative content, content containing abusive words, and content that is not age appropriate..

Keywords : *Crude Language, Content Rating, Speech Recognition*