

INTISARI

MODEL MONITORING SEBARAN PENYAKIT DEMAM BERDARAH DI INDONESIA BERDASARKAN ANALISIS PESAN TWITTER

Oleh

PUJI WINAR CAHYO

13/356442/PPA/04412

Kejadian Luar Biasa (KLB) demam berdarah dapat dihitung dari jumlah angka kejadian demam berdarah pada daerah suatu daerah. Informasi KLB yang diperoleh dapat digunakan sebagai tindakan *preventif* terhadap penyebaran penyakit demam berdarah. Menggunakan analisis data dari Twitter maka dapat dibentuk sebuah model monitoring sebaran penyakit demam berdarah dengan memanfaatkan NER (*Named Entity Recognition*) sebagai metode pengambilan informasi. NER tersebut terbagi kedalam dua tahapan, tahap pertama adalah pelatihan dengan menggunakan algoritma *Improve Iterative Scaling* dan tahap kedua adalah deteksi entitas menggunakan *maximum entropy*. Pengujian deteksi entitas menunjukkan nilai *Precision*, *Recall* dan *F-Measure* sangat rendah, untuk itu perlu dilakukan evaluasi pada proses deteksi entitas. Proses evaluasi menemukan beberapa masalah diantaranya adalah WSD (*Word Sense Disambiguation*) dan deteksi kata secara frase terlebih pada kata yang terdeteksi sebagai nama lokasi. Diterapkannya *string matching* sebagai solusi permasalahan tersebut terbukti dapat meningkatkan nilai *Precision*, *Recall* dan *F-Measure*, sehingga data yang dihasilkan dapat digunakan sebagai sumber data utama model *monitoring* demam berdarah. Pengujian data hasil model *monitoring* dengan dinas kesehatan setempat menunjukkan bahwa untuk mencapai kesesuaian hasil *monitoring* perlunya melihat konteks kalimat pada sumber informasi pesan Twitter yang diambil.

Kata-kata kunci : *social media*, *monitoring*, demam berdarah, *text mining*, *named entity recognition*

ABSTRACT

MODEL FOR MONITORING THE SPREAD OF DENGUE FEVER IN INDONESIA BASED ON ANALYSIS OF TWEETS

By

PUJI WINAR CAHYO

13/356442/PPA/04412

Kejadian Luar Biasa (KLB) dengue fever can be calculated from amount of dengue fever incidence in the region. KLB information can be used as a preventative measure against the spread of dengue fever. Using analysis of data from Twitter, it constructed a model of monitoring the spread of dengue fever by utilizing NER (Named Entity Recognition) as a method of information retrieval. NER is divided into two stages, the first stage is training using improve iterative scaling algorithm and the second stage is the detection of entities using the maximum entropy. The test of entity detection shows value of Precision, Recall and F-Measure is very low, it is necessary to evaluate the entity detection process. The evaluation process found some problems which are WSD (Word Sense disambiguation) and detection of words in the phrases structure, especially the word that is detected as the location name. Implementation of string matching as a solution to these problems are proven to increase the value of Precision, Recall and F-Measure, so that the resulting data can be used as source of data primary in monitoring model of dengue fever. Testing data of monitoring model compared with local health authorities showed that in order to achieve conformance of monitoring results need to see the context of sentence on resources taken from Twitter message.

Key words: social media, monitoring, dengue fever, text mining, named entity recognition