

Ekstraksi Informasi Bencana Alam di Indonesia dari Berita di Media Siber = Information Extraction on Natural Disaster Event in Indonesia from Cyber Media News

Gusman Dharma Putra, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20511254&lokasi=lokal>

Abstrak

Jenis bencana alam, lokasi, dan waktu kejadian adalah informasi minimal bisa mengindikasikan terjadinya sebuah bencana alam. Salah satu sumber informasi kejadian bencana alam adalah dari berita di media siber. Suatu sistem informasi tentang bencana alam bisa memanfaatkan berita di media siber sebagai sumber data, namun harus mengubah data teks berita menjadi bentuk data terstruktur. Teknik penambangan teks yang bisa digunakan untuk mendapatkan data terstruktur dari suatu kumpulan teks. Penelitian ini melakukan eksplorasi efektivitas teknik penambangan data untuk mengekstrak informasi jenis bencana alam, lokasi, dan waktu kejadian. Metode *web scraping* digunakan untuk mengumpulkan data teks berita dari media siber dan anotasi manual dilakukan untuk membuat data *gold standard*. Penelitian ini menggunakan klasifikasi teks dengan *machine learning* untuk mengetahui jenis bencana alam yang diberitakan. Klasifikasi biner diterapkan untuk mengetahui pemberitaan tentang bencana angin topan, banjir, erupsi, gempa, karhutla, kekeringan, longsor, dan tsunami. Algoritma yang diuji untuk klasifikasi teks adalah Multinomial Naive Bayes, Support Vector Machine, Random Forest, Linear Regression dan Adaboost. Penelitian ini memanfaatkan aplikasi Stanford NER untuk mengetahui entitas lokasi di suatu teks, kemudian gazetteer digunakan untuk pemetaan wilayah administrasi. Penelitian ini menggunakan pencocokan pola teks dengan regular expression untuk mengekstrak informasi tanggal kejadian bencana alam. Nilai F1 dari model klasifikasi penelitian ini untuk berita bencana angin topan, banjir, erupsi, gempa, karhutla, kekeringan, longsor, dan tsunami adalah 0,731, 0,767, 0,760, 0,761, 0,749, 0,680, 0,763, dan 0,600. Sedangkan Nilai F1 untuk hasil ekstraksi lokasi dan waktu adalah 0,795 dan 0,881.

.....The minimal information to notice the occurrence of a natural disaster is its type, location, and event time. News in the cyber media can be a source of information to discover disaster events. Furthermore, natural disaster information systems can utilize that news as the data source. The news needs to be converted into structured data to be processed by system information. Text mining is a method to extract structured information from a text collection. This research explored the effectiveness of data mining to extract natural disaster type, location, and event time reported by news in cyber media. The web scraping method was used to collect news in cyber media and manual annotation was performed to create gold-standard data. This study used text classification with a machine learning approach to identify the types of natural disasters reported. Binary classification was applied to label news for following disaster types: hurricanes, floods, eruptions, earthquakes, forest and land fires, droughts, landslides, and tsunami. This research evaluated Multinomial Naive Bayes, Support Vector Machines, Random Forests, Linear Regression, and AdaBoost algorithm for text classification tasks. This study utilized the Stanford NER application to recognize location entities in a text, then the gazetteer was used to get administrative area information. This study applied text patterns with regular expressions to extract date information of disaster events. The F1 value of 8 classification model in this research for following disaster news type: hurricanes, floods, eruptions, earthquakes, forest and land fires, droughts, landslides, and tsunami, are 0.731, 0.767, 0.760, 0.761, 0.749,

0.780, 0.680, 0.763, and 0.600. The F1 value of method to extract location and event time information are 0.795 and 0.881.