

Pembangunan Korpus Terhadap Kebakaran Hutan untuk Bahasa Indonesia dan Bahasa Melayu

Sadam Andhika Putra Darmawan¹, Warih Maharani²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹sadamandhikapd@students.telkomuniversity.ac.id, ²wmaharani@telkomuniversity.ac.id

Abstrak

Kebakaran hutan dan kabut asap merupakan hal yang sangat umum terjadi, bahkan merupakan kejadian tahunan di Indonesia. Kebakaran hutan ini biasa terjadi pada musim kemarau. Dampak dari kebakaran hutan ini sangat buruk dalam berbagai aspek, mengganggu kesehatan, ekonomi, transportasi, dan bahkan masuk ke ranah politik, terutama politik luar negeri, karena asap dari kebakaran hutan dapat menyebar hingga ke negara tetangga, mengganggu kehidupan masyarakat di negara tetangga tersebut dan menjadi bahan perbincangan di media sosial Twitter. Twitter merupakan salah satu media sosial terpopuler di dunia, dimana pengguna dapat mengirim dan membaca pesan berbasis teks yang disebut dengan tweet. Hal ini menjadikan Twitter sebagai tempat untuk mendiskusikan berbagai topik. Opini yang diutarakan oleh pengguna dapat digunakan untuk membangun korpus baik untuk Bahasa Indonesia maupun Bahasa Melayu. Namun, hanya sedikit penelitian yang membahas Bahasa Indonesia dan Bahasa Melayu, dan metodologi anotasi manual belum banyak digunakan untuk kedua korpus tersebut. Oleh karena itu, penelitian ini berfokus pada kata-kata dengan kosakata yang sama namun memiliki arti yang berbeda dalam kebakaran hutan yang terjadi saat ini. Data diperoleh dari Twitter, sehingga ditemukan kata-kata informal yang kemudian dipilih secara acak dengan kosakata yang sama tetapi memiliki arti yang berbeda.

Kata kunci : *Anotasi, Bahasa Indonesia, Bahasa Melayu, Kebakaran Hutan, Korpus*