

Pengklasifikasian artikel berita berbahasa Indonesia secara otomatis menggunakan ontologi

Prajna Wira Basnur, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=122485&lokasi=lokal>

Abstrak

Klasifikasi dokumen adalah sebuah metode untuk menentukan suatu dokumen termasuk ke suatu kategori secara otomatis berdasarkan isi dokumen. Metode Naïve Bayes dan ontologi merupakan metode klasifikasi dokumen teks yang digunakan dalam penelitian ini. Data yang digunakan dalam penelitian ini berupa artikel berita berbahasa Indonesia dari situs <http://www.kompas.com>. Dalam penelitian ini menggunakan lima kategori dalam domain olahraga untuk melakukan klasifikasi dokumen, yaitu kategori bulutangkis, basket, otomotif, sepakbola, dan tenis. Klasifikasi dokumen dengan menggunakan ontologi dilakukan dengan membandingkan nilai kemiripan diantara dokumen dan sebuah node yang ada di ontologi. Sebuah dokumen diklasifikasikan ke sebuah kategori atau node, jika memiliki nilai kemiripan paling tinggi diantara semua node yang ada di ontologi. Hasil penelitian menunjukkan bahwa ontologi dapat digunakan untuk melakukan klasifikasi dokumen. Nilai recall, precision, dan f-measure untuk klasifikasi dokumen menggunakan ontologi berturut-turut adalah 97.03%, 91.63%, dan 94.02%.

<hr>

Document classification is a method for determine document category automatically based on contents of document. In this research, we use Naïve Bayes and Ontology method for document classification. Mass media in Bahasa Indonesia is used as data in this research. Data is taken from <http://www.kompas.com>. We uses five category in sports domain for document classification that comprise with bulutangkis, basketball, automotive, soccer, and tennis category. Document classification uses ontology can be done with compare similarity value between document and a node in ontology. A document can classified to a category or node, if a document has highest similarity value between all node in ontology. In this research indicate that ontology can used for document classification. Recall, precision, and f-measure value for document classification using ontology in a row are 97.03%, 91.63%, and 94.02%.