

Abstrak

Text mining merupakan bagian dari data mining yang memfokuskan bidangnya pada pengambilan informasi berharga dari *database* yang berupa teks. Sebagian besar proses *data mining* menggunakan dua metode yaitu kategorisasi yang dikenal juga dengan klasifikasi, dan klasterisasi. Kategorisasi adalah teknik yang bertujuan untuk menentukan topik dari suatu artikel atau *text document* berdasarkan atribut kelas yang sudah ditentukan sebelumnya (*supervised*). Sedangkan klasterisasi adalah pengelompokan topik dari suatu artikel berdasarkan kemiripannya dengan artikel yang lain tanpa melihat atribut kelasnya (*unsupervised*).

Jika diperhatikan lebih lanjut, sebagian besar teknik pengelompokan pada *text mining* bekerja berdasarkan analisis statistik dari frekuensi kemunculan *term* pada suatu dokumen tanpa menghiraukan semantik atau arti kata dari *term* tersebut. Sehingga bisa saja terjadi dua *term* mempunyai frekuensi yang sama pada suatu dokumen padahal salah satu *term* diantara kedua *term* tersebut memberikan kontribusi yang lebih besar dalam proses penentuan topik sebuah dokumen. Oleh karena itu perlu diterapkan pendekatan baru yang melibatkan *term* yang mempunyai peranan dalam semantik kalimat untuk mencapai hasil yang lebih akurat.

Graphical concept-based mining model memberikan pendekatan baru dalam pemisahan *term* yang berarti dan yang tidak berarti bagi penentuan topik sebuah dokumen berdasarkan peranan *term* dalam semantik. Hal tersebut yang menyebabkan *graphical concept-based mining model* memberikan hasil yang lebih akurat.

Kata kunci: TF IDF, graphical concept-based mining model, conceptual ontological graph, concept, semantik.