

Abstrak

Seiring dengan perkembangan dunia internet, semakin banyak pula halaman-halaman web yang tersedia di dunia maya. Banyak orang yang mencari informasi mengenai kesehatan di internet karena kesehatan merupakan hal penting dalam kehidupan manusia. *User* akan mengalami kesulitan untuk mendapatkan halaman web dengan cepat, padat dan relevan sesuai dengan topik kesehatan. *Focused crawler* dapat memberikan halaman web sesuai dengan topik yang dimaksud. *Focused crawler* secara selektif menelusuri dan mengambil halaman *Web* yang relevan. Tentu saja selain crawler komponennya, terdapat juga classifier untuk membangun sebuah *focused crawler*.

Pada tugas akhir ini dilakukan implementasi dan analisis terhadap *focused crawler* menggunakan Naive Bayes Classifier (NBC). Urutan antrian penelusuran halaman *web* menggunakan *breadth-first search (BFS)*. Diharapkan dengan menggunakan *naive bayes classifier* didapatkan akurasi hasil klasifikasi yang tinggi. Sebelum melakukan pengujian terhadap *focused crawler*, dilakukan pengujian jumlah halaman *web* yang dilibatkan dalam proses *training* dan digunakan *stopword* atau tidak pada data training. Tujuannya ialah untuk menghasilkan akurasi hasil klasifikasi yang maksimal.

Ditunjukkan pada tugas akhir ini bahwa *focused crawler* menggunakan *naive bayes classifier* menghasilkan hasil klasifikasi halaman *web* yang relatif kurang baik karena terjadi penurunan akurasi di setiap levelnya.

Kata Kunci : *focused crawler* , *naive bayes classifier*, klasifikasi, halaman web, kesehatan, *training*.