

PERANCANGAN DAN PEMBUATAN DIGITAL LIBRARY DENGAN MENGGUNAKAN METODE K-NEAREST NEIGHBOR (KNN)

Eka Ridho Septiyan
(10.04.1.1.1.00041)

Pembimbing I : Firdaus Solihin, S.Kom., M.Kom
Pembimbing II : Rika Yunitarini, ST., M.T.

ABSTRACT

Information technology development and data stream speed lead up the increase of information need rate. This condition imply heighten amount of digital document growth and push the user into digital document searching problem. Index list built to handle that problem in a system. This technique enhance information forwarding and help user to get an appropriate document.

Information retrieval is a way to get information from electronic book with tracking an electronic book in the database. Information gathering employ a web crawler will trace a web page and collect documents to be processed and downloaded using web crawler.

Those documents will be classified use K-Nearest Neighbour method (KNN). This method classify data based on highest vote rate between training data and testing data. KNN obtain K value accurateness (neighbourhood) k.1-k.7 80% – 70%, k.8-k.20 between 60%, 50%, 30%. From this experiment, we get a conclusion that bigger K value, lower accurateness we got during classification process.

Keyword: *Digital library, Information retrieval, Web crawler, K-Nearest Neighbour*

ABSTRAK

Seiring perkembangan teknologi dan cepatnya aliran data maka kebutuhan informasi akan meningkat dan menyebabkan peningkatan jumlah dokumen *digital* yang sangat pesat, dengan fakta tersebut maka seorang pengguna kesulitan dalam pencarian dan penemuan kembali dokumen *digital*. untuk menangani hal tersebut salah satunya dengan membangun daftar indeks dalam suatu sistem, Sehingga dengan dibangunnya *digital library* diharapkan dapat membantu dan meningkatkan penyampaian informasi tentang *electronic book*, serta memudahkan *user* dalam pencarian dokumen yang diinginkan. *Information retrieval* merupakan salah satu cara untuk memperoleh informasi tentang *electronic book* dengan melakukan penelusuran kembali atas *electronic book* yang tersimpan dalam *database*. Proses mengumpulkan semua informasi yang ada di dalam halaman web menggunakan *web crawler*, *web crawler* akan berjalan menelusuri halaman web dan mengumpulkan dokumen-dokumen kemudian melakukan proses *download* data yang telah dikumpulkan oleh *web crawler*. Dokumen-dokumen tersebut akan diklasifikasikan menggunakan metode *K-Nearest Neighbor* (KNN). *K-Nearest Neighbor* merupakan metode klasifikasi data berdasarkan voting terbanyak antara *data training* dan *data testing* dari klasifikasi tersebut didapatkan tingkat akurasi nilai k (ketetanggaan) k.1-k.7 adalah 80%-70% sedangkan akurasi k.8-k.20 antara 60%, 50%, 30% dengan demikian semakin banyak nilai k maka semakin berkurang tingkat akurasi pada proses klasifikasi.

Kata Kunci: *Digital Library, Information retrieval, Web Crawler, K-Nearest Neighbor.*