Evaluasi

Reranking menggunakan LETOR dilakukan pada hasil query dengan BM25 dengan k1=0 dan b=0. Dataset yang digunakan untuk testing menggunakan model test grels.txt.

Berikut adalah hasil evaluasi.

	Tanpa LETOR	Dengan LETOR
RBP	0.87	0.87
DCG	7.12	7.14
AP	0.86	0.86

Hasil evaluasi mengalami peningkatan, namun tidak tampak signifikan.

Sementara itu berikut adalah hasil evaluasi untuk BM25 dengan k1=2 dan b=0.75.

	Tanpa LETOR	Dengan LETOR
RBP	0.93	0.92
DCG	7.49	7.44
AP	0.92	0.91

Hasil percobaan tersebut mengeluarkan hasil di mana LETOR memiliki score yang lebih buruk.

Dari hasil percobaan tersebut, ditarik kesimpulan bahwa LETOR dapat berfungsi, akan tetapi bila hasil SERP sudah cukup akurat, LETOR dapat merusak hasil yang sudah bagus. Semuanya bergantung pada model yang digunakan. Dalam contoh ini, LGBMRanker bisa saja memiliki data training yang kurang bagus. Hal lain yang bisa saja mempengaruhi hal training adalah tipe karakter yang ada di dalam docs dan query. Ketika saya menjalankan experimen pertama kali, terdapat error yang mengatakan character dalam file tidak dapat terbaca, sehingga perlu di-*encode* oleh utf8.