

## Efek penerapan pemotongan imbuhan kata, operator kedekatan kata, struktur dokumen web, serta umpan balik relevan semu terhadap kinerja perolehan informasi web

Okky Hendriansyah Tri Firgantoro, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=123572&lokasi=lokal>

---

### Abstrak

Perolehan informasi Web merupakan sebuah cabang perolehan informasi yang mengkhususkan pencarian informasi terhadap dokumen-dokumen Web. Dokumen Web atau HTML memiliki karakteristik yang berbeda dengan dokumen teks biasa, sehingga teknik-teknik yang sudah terbukti efektif dilakukan pada koleksi dokumen teks biasa belum tentu memiliki efek yang sama pada dokumen Web. Penelitian yang dilakukan adalah untuk mempelajari efek melakukan proses stemming dan menghilangkan stopwords pada masing-masing bahasa, penggunaan operator kedekatan kata pada kueri, pembobotan pada struktur dokumen HTML, serta efek dari melakukan umpan balik relevan semu pada koleksi dokumen Web yang multilingual. Penelitian dilakukan dengan menggunakan koleksi pengujian EuroGOV. Ternyata efek menggunakan stemmer cukup bervariasi pada masing-masing bahasa, namun secara keseluruhan penggunaan stemmer dapat meningkatkan hasil dokumen yang diperoleh hingga 1,61% dibandingkan dengan tanpa menggunakan stemmer. Sedangkan menghilangkan stopwords dapat meningkatkan kinerja perolehan Web hingga 4,06%. Penggunaan operator kedekatan kata pada kueri juga dapat mengefektifkan hasil dokumen yang diperoleh, dengan peningkatan terbesar sekitar 4,68% jika jarak antarkata yang diberikan tetap dan sekitar 5,71% pada jarak antarkata yang sesuai dengan panjang kueri. Eksperimen pembobotan pada struktur dokumen HTML membuahkan kesimpulan bahwa bagian judul, inlink, dan kepala dokumen HTML memiliki peran yang lebih penting dalam peningkatan kinerja jika dibandingkan dengan bagian struktur HTML lainnya. Melakukan umpan balik relevan semu pada koleksi dokumen Web multilingual justru menurunkan hasil dokumen relevan yang diperoleh hingga 15,73% dibandingkan dengan tanpa melakukan umpan balik relevan semu sama sekali.