

Abstraksi

Information Retrieval System(IRS) merupakan bagian dari *computer science* yang berhubungan dengan pengambilan informasi dari dokumen-dokumen yang didasarkan pada isi dan konteks dari dokumen-dokumen itu sendiri. Proses dalam *Information Retrieval* dapat digambarkan sebagai sebuah proses untuk mendapatkan *relevant documents* dari *collection documents* melalui pencarian *query* yang diinputkan *user*[14].

Salah satu cara untuk meningkatkan performansi IRS adalah penggunaan pengindeksan kata(*word indexing*) dan pengindeksan frase(*phrase indexing*). Pada *word indexing*, pembobotan *term* dilakukan pada setiap kata yang ada pada dokumen, sedangkan *phrase indexing*, *term* yang dibobot adalah kumpulan kata atau frase. Adapun pemodelan yang dipakai adalah model probabilistik, IRS menghitung kemungkinan munculnya elemen indeks dalam suatu dokumen.

Dalam penerapan *word indexing* dan *phrase indexing* pada model probabilistik menghasilkan dokumen-dokumen hasil pencarian dengan nilai performansi yang berbeda. *Phrase indexing* tidak selamanya memberikan nilai performansi lebih yang baik dari *word indexing* karena penggunaan frase yang tidak tepat pada suatu dokumen mempengaruhi nilai relevansi dokumen tersebut terhadap *query*. Selain itu, jumlah *term* dari masing-masing metode pengindeksan juga mempengaruhi nilai korelevanan suatu dokumen. Pada tugas akhir ini, *word indexing* memberikan nilai performansi yang lebih baik daripada *phrase indexing*.

Kata Kunci: *information retrieval*, *word indexing*, *phrase indexing*, model probabilistik, performansi.