

ABSTRAK

Nama : Clara Vania

Program Studi : Ilmu Komputer

Judul : Perolehan Opini pada Dokumen Blog Berbahasa Indonesia

Perolehan opini merupakan sebuah proses perolehan dokumen, dimana dokumen yang diperoleh dan diurutkan berdasarkan opini mengenai sebuah topik kueri yang terdapat pada dokumen. Sebuah dokumen yang relevan harus relevan terhadap topik kueri dan mengandung opini mengenai kueri. Salah satu penelitian perolehan opini yang sedang diminati adalah perolehan opini pada dokumen blog. Perolehan opini pada dokumen blog dapat membantu dalam menemukan pendapat publik mengenai suatu target, yang kemudian dapat memantau perkembangan merek suatu produk, analisis media, dan menelusuri kebutuhan pasar.

Pada tugas akhir ini, perolehan opini dilakukan pada dokumen blog berbahasa Indonesia dengan menggunakan pendekatan statistik. Korpus blog yang digunakan berupa *feeds* dan *permalinks* dengan jumlah total sebanyak 21.411 dokumen HTML. Penelitian dilakukan secara bertahap, yaitu dengan mencoba berbagai teknik pemrosesan kueri, teknik perbaikan kueri, dan teknik pengurutan dokumen berdasarkan opini. Hasil evaluasi terbaik dari setiap penelitian akan digunakan untuk penelitian selanjutnya. Evaluasi dilakukan dengan membandingkan nilai MAP dari masing-masing percobaan.

Hasil penelitian menunjukkan adanya peningkatan sebesar 49% pada penggunaan pemotong kata berimbuhan bahasa Indonesia dan sebesar 20% pada pencarian berdasarkan frase dengan menggunakan judul dan deskripsi pada topik kueri. Selain itu, penambahan bobot apabila kueri terdapat pada *field body* dan penggunaan Indri *pseudo-relevance feedback* juga dapat meningkatkan akurasi walau hanya sedikit, yaitu masing-masing sebesar 4,09% dan 0.62%. Eksperimen pengurutan dokumen memberikan hasil terbaik untuk metode pengurutan dengan menambahkan nilai *similarity* dokumen berdasarkan kata-kata opini, namun tidak terlalu berpengaruh terhadap akurasi perolehan blog, yaitu hanya sebesar 0.11%.

Kata kunci: perolehan opini, blog, *blog retrieval*, Indri *pseudo-relevance feedback*, MAP

ABSTRACT

Name : Clara Vania

Study Program : Computer Science

Title : Opinion Retrieval from Indonesian Blogs

Opinion retrieval is a document retrieval process, which requires documents to be retrieved and ranked based on their opinions about a query topic. A relevant document must be relevant to the query topic and also contains opinions about the query. An example of research in opinion retrieval is opinion retrieval from blogs. Opinion retrieval from blog can help to uncover public sentiment about a target, which can be used to do brand monitoring, media analysis, and tracking consumer's need.

In this research, we used Indonesian blogs as our corpus for opinion retrieval using statistical technique. We used both blog feeds and permalinks which are made up of 21.411 HTML documents. We do research step by step by doing some experiments in query processing techniques, pseudo-relevance feedback techniques, and document ranking based on opinion techniques. The best result from each experiment will be used in the next experiment. To evaluate the accuracy of retrieval, we used the MAP score.

Our result showed that the use of Indonesian stemmer can improve 49% in performance, while query searching with phrase using title and description in query topic can improve 20% in performance. Term weighting in field body can improve performance for about 4.09%, while the use of Indri's pseudo-relevance feedback gives 0.62% performance improvement. Document ranking technique by adding similarity score based on opinion words give the best result, but do not provide any significant performance improvement, that is only about 0.11%

Key words: opinion retrieval, blog, blog retrieval, Indri pseudo-relevance feedback, MAP