

Optimalisasi *Feedback Query* Istilah Kultur dan Subkultur terhadap Konten Bahasa Resmi Menggunakan Metode Rocchio Relevace Feedback (Feedback Query Optimization of Culture Terms and Subculture Content of The Official Language Using Rocchio Relevace Feedback)

Arifin Paulus Tindi, Alfrianus Papuas, Adi Pandu Nababan

Kampus Polnustar Tahuna

arifintindi@live.com, papuasalfri@yahoo.com, nababanschoolboy@gmail.com

Abstrak: Indonesia merupakan negara dengan kultur dan sub kultur adat budaya yang sangat variatif dari berbagai macam daerah dan hal tersebut mempengaruhi kehidupan bermasyarakat dalam penggunaan istilah yang cukup beragam dan telah dipakai secara turun temurun. Bahasa Indonesia dipakai sebagai bahasa resmi dalam kehidupan berbangsa dan bermasyarakat. Semua konten resmi menggunakan bahasa Indonesia, yang bagi sebagian masyarakat tersebut belum memahami dengan baik. Menjembatani antara istilah kultur atau pun sub kultur masyarakat diperlukan interpreter yang dapat mengerti dan memahami istilah tersebut kemudian mencocokkan dengan penggunaan bahasa Indonesia yang benar dan tepat. Metode Rocchio Relevance Feedback digunakan sebagai interpreter untuk menjembatani hal tersebut.

Kata Kunci: kultur dan sub kultur, bahasa Indonesia, Metode Rocchio Relevance Feedback

Abstract: Indonesia is a country with a culture and sub-culture is very varied cultural customs of the various regions and it affects the life of society in the use of the term is quite diverse and has been used for generations. Indonesian is used as an official language in the life of the nation and society. All content is authorized to use the Indonesian language, which for some people is not yet well understood. Bridging between the terms of culture or sub-culture of society is required interpreters who can see and understand these terms then match the use of the Indonesian language is correct and appropriate. Rocchio Relevance Feedback Method is used as interpreters to bridge it.

Keywords: culture and sub-cultures, Indonesian, Rocchio Relevance Feedback Method

Menurut sensus penduduk tahun 2010, perkembangan penduduk Indonesia saat ini mencapai jumlah 237.556.363 jiwa, yang menempatkan Indonesia pada urutan keempat dunia setelah Cina, India dan Amerika Serikat. Penduduk Indonesia tersebar dari ujung Barat hingga Timur, mulai dari Sumatra sampai Papua dengan kondisi geografis yang berbeda-beda seperti wilayah pesisir, tepian hutan, pedesaan, perkotaan, dataran rendah dan pegunungan/dataran tinggi. Beragam suku bangsa hidup berdampingan dengan latar belakang kehidupan yang berbeda, Kondisi geografis tempat tinggal yang berbeda tersebut menjadikan masyarakat di Indonesia memiliki

kehidupan beraneka ragam yang dipengaruhi oleh budaya masing-masing sebagai warisan dari tiap generasi sebelumnya (Widiastuti, 2013).

Negara Indonesia yang saat ini memiliki lebih dari 240 juta jiwa penduduk, mempunyai ratusan bahasa daerah yang tersebar dari ujung pulau Sumatera hingga Papua. Dalam *Ethnologue: Languages of the World*, tercatat Indonesia memiliki 726 bahasa. Dari jumlah itu 719 bahasa masih digunakan oleh penuturnya, dua bahasa menjadi bahasa kedua tanpa penutur bahasa ibu (*mother tongue*) dan lima bahasa sisanya diklaim punah karena tidak ada lagi penuturnya (Rusdi, 2012).

Salah satu aspek yang bisa dikaitkan dengan relevansi tersebut adalah bagaimana mengatur sebuah produk hukum daerah dimana keterkaitan erat antara panduan hukum dan kehidupan masyarakat membuat intensitas kebutuhan akan kedua hal tersebut sangat berkaitan erat tak terkecuali dalam organisasi pemerintahan.

Dengan adanya kewenangan yang dimiliki oleh Pemda tersebut, maka Pemda diberikan kekuasaan untuk menuangkan pengakuan dan perlindungan masyarakat hukum adat dalam produk hukum daerah yang dibentuknya. Dengan mengelaborasi urusan pemerintahan yang menjadi kewenangan Pemda, maka materi muatan produk hukum daerah bisa sangat luas. Sehingga tidak seharusnya Pemda mempersempit kewenangannya sendiri dalam menuangkan materi muatan dalam produk hukum daerah. Hal ini dimaksudkan agar produk hukum daerah tersebut menjadi solusi dari berbagai permasalahan ketidakadilan dan pelanggaran hak asasi manusia atas masyarakat hukum adat (Anggraeni, 2015).

Semua organisasi memiliki kekuatan dan kelemahan dalam area fungsional bisnis. Tidak ada perusahaan yang sama kuatnya atau lemahnya dalam semua area bisnis. Kekuatan internal, digabungkan dengan peluang atau ancaman dari eksternal dan pernyataan misi yang jelas, menjadi dasar untuk penetapan tujuan dan strategi. Tujuan dan strategi ditetapkan dengan maksud memanfaatkan kekuatan internal dan mengatasi kelemahan (David, Fred, 2006).

Dalam penerapan konten tentang kehidupan berbangsa dan bernegara, adakalanya masyarakat sebagai pengguna suatu sistem masih cenderung menguasai bahasa kultur atau subkultur (bahasa ibu) dan belum memahami secara jelas tentang penggunaan bahasa atau istilah resmi yang termuat sehingga apa yang dicari masyarakat dalam suatu muatan konten cenderung tidak bisa ditemukan jika menggunakan keywords yang tidak sama persis dalam querying terhadap suatu sistem.

Untuk menjembatani kerenggangan tersebut perlu dibangun sebuah sistem katalog yang dapat menjembatani antara query yang diberikan oleh pengguna dan konten yang termuat dalam suatu sistem katalog. Penerapan Metode Rocchio Relevance Feedback dapat memaksimalkan suatu feedback yang lebih relevan terhadap query umum yang diberikan kepada masyarakat dan memberikan hasil penelusuran yang lebih optimal dan relevan.

Metode

(Uden, 2011) Metode *Rocchio Relevance Feedback* adalah strategi *reformulasi query* paling populer karena sering digunakan untuk membantu user pemula satu *information retrieval system*. Dalam siklus *relevance feedback*, kepada user disajikan hasil pencarian dokumen, setelah itu user dapat memeriksa dan menandai dokumen yang benar-benar relevan.

Rocchio classifiers merupakan salah satu metode pembelajaran *supervised document classification*. Metode klasifikasi *rocchio* membandingkan kesamaan isi antara data training dan data tes dengan merepresentasikan semua data ke dalam sebuah vector. Kedekatan kesamaan isi dihitung dari kedekatan sudut yang terbentuk antara bobot data training dan bobot data test menggunakan aturan sodinus. Untuk menghitung bobot setiap kata dalam dokumen digunakan skema pembobotan *tfidf* (*Term Frequency/Invers Document Frequency*) karena komponen heuristic utama adalah klasifikasi *rocchio* yaitu skema pembobotan *tfidf*, metode pembelajaran *rocchio* disebut juga dengan *tfidf Classifier*

Di dalam membandingkan kesamaan isi antara data training dan data test, *tfidf classifiers* menggunakan *prototype vector* untuk merepresentasikan kategori yang terbentuk dari data *training*, dengan kata lain *prototype vector* merupakan *vector* yang mewakili seluruh *vector* data training dalam setiap kategori. Tiga hal utama yang dapat dipakai pada klasifikasi *tfidf* adalah menggunakan skema pembobotan *tfidf* yang berguna untuk merepresentasikan dokumen ke dalam sebuah vector, merepresentasikan *prototype* setiap kategori dengan menjumlahkan vector-vector dalam satu kategori dari data training, membandingkan kedekatan sudut antara vector data test dengan semua *prototype vector*

Teknik ini menggunakan *vector space model* untuk merepresentasikan setiap dokumen dalam korpus. Algoritma *rocchio* diasumsikan bahwa sebagian pengguna memiliki konsep umum pada dokumen yang relevan dan non-relevan sebagai sarana meningkatkan pencarian yang singkat dan presisi.

Langkah-langkah pengolahan query:

- a. Text Mining dan Klasifikasi Teks: mencari dan mengelompokkan dokumen ke dalam kategori tertentu
- b. Parsing: memilah isi dokumen menjadi unit-unit kecil (token), yang berupa kata, frase, atau kalimat

- c. Stemming: proses penghilangan prefiks (awalan) dan surfixs (akhiran)
- d. Inverted Index: struktur yang dioptimasi untuk menemukan kembali dokumen

Dalam menggunakan *vector space* model diperlukan batas-batas antar kelas untuk mengetahui klasifikasi yang sesuai. Teknik Rocchio menggunakan centroid untuk batas-batas tersebut

$$\vec{\mu}(c) = \frac{1}{|D_c|} \sum_{d \in D_c} \vec{v}(d)$$

Di mana D_c adalah himpunan dokumen di dalam korpus pada kelas c , sedangkan merupakan vektor dokumen yang telah dinormalisasi. Untuk menentukan kemiripan dua vektor space model ada dua cara yaitu dengan mengukur jarak atau dengan mengukur kemiripan. Dalam menentukan jarak (distance) antara dua vektor space model digunakan jarak euclidean.

$$d(d_1, d_2) = \sqrt{\sum_{i=1}^n (w_{i,d1} - w_{i,d2})^2}$$

Dan dengan menghitung kemiripan (similarity) antara dua vektor dokumen adalah sebagai berikut:

$$sim(d_1, d_2) = \frac{v(d_1) \cdot v(d_2)}{|v(d_1)| |v(d_2)|}$$

Jika terdapat suatu kueri diproses menjadi sebuah vektor space, maka dapat dibandingkan dengan masing-masing centroid kelas yang ada pada korpus. Dengan dua pendekatan mencari kemiripan dua vektor space, vektor kueri dianggap mirip dengan sebuah centroid kelas dapat dilakukan dengan menggunakan jarak, yang dicari adalah kelas yang memiliki jarak yang terkecil dengan kueri. Dan jika menggunakan kemiripan yang dicari adalah kelas yang memiliki kemiripan yang paling besar dengan kueri, seperti yang ada di bawah:

1. Menggunakan jarak

$$\min |\vec{\mu}(c) - \vec{v}(d)|$$

2. Menggunakan kemiripan

$$\max \cos(\vec{\mu}(c), \vec{v}(d))$$

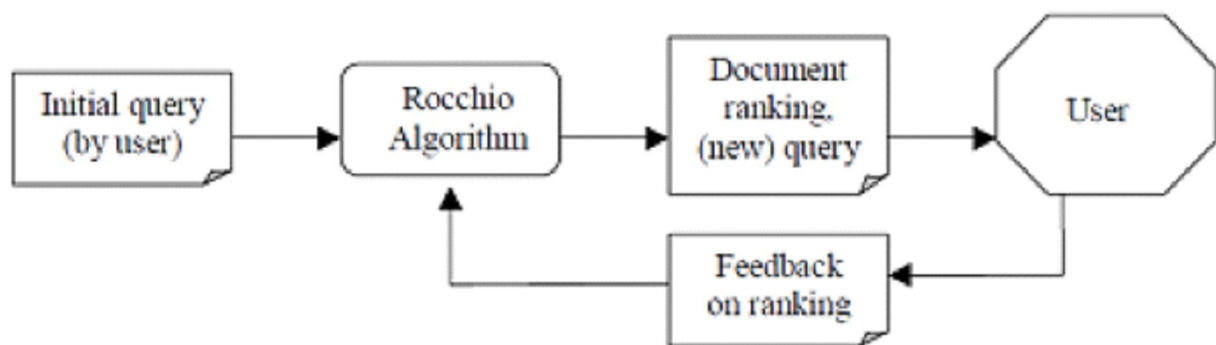
Hasil Pembahasan

Hasil yang dicapai adalah rancang sistem menggunakan Metode *Relevance Feedback* untuk Optimalisasi Pencarian Data Katalog yang dapat membantu dalam proses pencarian data terutama dalam istilah-istilah non formil (atau bahasa daerah), menyajikan informasi-informasi yang berkaitan antar konten dan konteks dalam menentukan relevansi feedback.

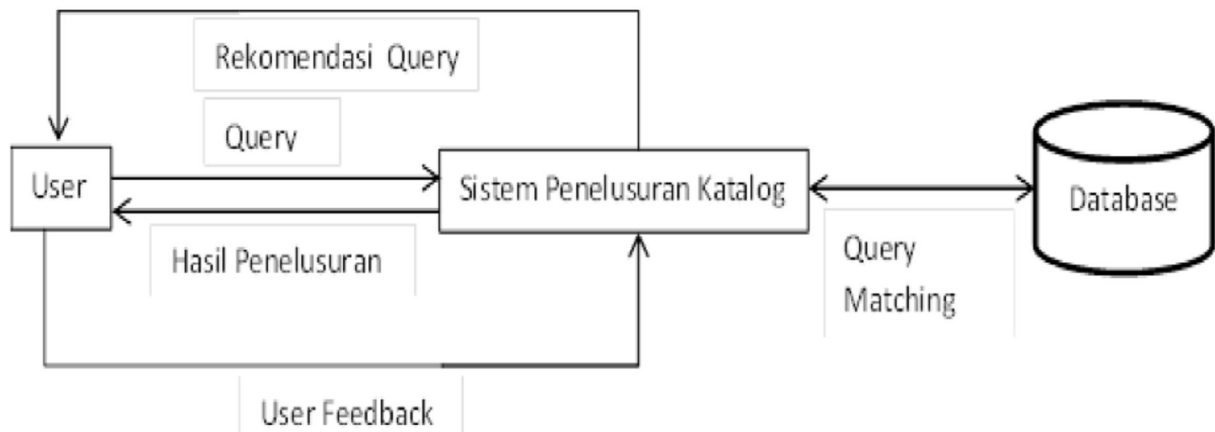
Gambar 1 menunjukkan bahwa metode *Rocchio relevance feedback* memberikan rekomendasi hasil penelusuran katalog kepada *user* dalam hal ini adalah masyarakat pengguna bahasa non formil (bahasa daerah) dengan melakukan modifikasi *query* awal sehingga menghasilkan *query* yang lebih relevan sesuai dengan informasi yang ingin ditelusuri oleh pengguna berdasarkan bahasa resmi (bahasa Indonesia) yang termuat di dalam katalog konten terkait.

Kinerja sistem penelusuran katalog yang akan dikembangkan adalah sebagai berikut User memasukkan *query* dalam bentuk istilah atau bahasa daerah (bahasa/istilah subkultur), *query* yang dicari oleh user akan diolah oleh sistem untuk menghasilkan rekomendasi *query* kepada *user*, sehingga user dapat memilih *query* mana yang tepat seperti yang diinginkan *user* dengan muatan konten yang telah menggunakan bahasa resmi atau bahasa Indonesia. Selanjutnya *user* memberikan *feedback* kepada sistem berdasarkan rekomendasi yang telah dipilih. Rekomendasi *query* yang dipilih oleh user akan ditelusuri oleh sistem.

Dari Arsitektur Sistem seperti yang terlihat, maka dikembangkan beberapa modul fungsi yang digunakan dalam sistem temu kembali informasi meliputi:



Gambar 1. Konsep Rocchio Relevance Feedback



Gambar 2. Konsep Sistem dengan Metode Rocchio Relevance Feedback

- Tokenisasi dan *indexing*, Merupakan fungsi untuk melakukan pembentukan vektor *space* dokumen dari file-file dokumen
- Pembentukan *Centroid*, Merupakan fungsi membentuk vektor *centroid* masing-masing kelas yang ada dokumen korpus
- Rekomendasi, Merupakan fungsi untuk merekomendasikan klasifikasi dokumen berdasarkan kelas keilmuan yang ada
- Pemeriksaan *query*. Merupakan fungsi untuk membandingkan vektor *query* dengan *centroid* kelas
- Interface Pencarian. Merupakan fungsi untuk memasukkan *query* dan menampilkan hasil kueri pencarian

diantara dokumen lain sesuai dengan *input term* dari *user*.

Saran

- a. Dalam perkembangannya Konsep Rocchio Relevance Feedback Dapat Dikembangkan sebagai metode untuk memberikan solusi terhadap multi istilah terutama di Indonesia yang memiliki berbagai macam kultur dan sub kultur masyarakat.
- b. Metode Rocchio dapat dikembangkan lebih dalam lagi dan memberikan implikasi positif pada daerah yang memiliki kecenderungan tinggi terhadap pemakaian istilah nonformil atau istilah resmi sebagai contoh dalam sistem katalog produk hukum daerah.

KESIMPULAN DAN SARAN

Kesimpulan

Dari hasil penelitian yang telah dilakukan maka diperoleh kesimpulan sebagai berikut:

- a. Proses *query* di dalam sistem penelusuran katalog produk hukum menggunakan metode *rocchio relevance feedback* terdiri dari empat tahap yaitu tokenizing, filtering, stemming dan term weighting.
- b. Penggunaan metode *rocchio relevance feedback* dapat melakukan penelusuran dalam suatu konten sesuai *input term* dari *user*.
- c. Sistem penelusuran katalog produk hukum dengan mengimplementasikan metode *rocchio relevance feedback* menampilkan hasil penelusuran yang memiliki nilai kemiripan tertinggi

DAFTAR RUJUKAN

- Uden, M.V. 2011. *Rocchio Relevance Feedback in Learning Classification Algorithms*, Thesis. Departmentn of Computing Science University of Nijmegen.
- Risca, A. 2015. *Produk Hukum Daerah*, Info Hukum Vol. 2, Epistema Institute.
- Widiastuti. 2013. *Analisis SWOT Keberagaman Budaya Indonesia*, Jurnal Ilmiah WIDYA Vol. 1, Universitas Darma Persada.
- Farid Rusdi. 2012. *Bahasa Daerah dan Industri Radio*, Menggagas Pencitraan Berbasis Kearifan Lokal, Jakarta.
- David, Fred, R. 2006. *Manajemen Strategi*. Buku 1, Edisi Kespuluh. Jakarta: Penerbit Salemba Empat.