

Penerapan *Cosine Similarity* dalam Aplikasi Kitab Undang-Undang Hukum Dagang (*Wetboek Van Koophandle Voor Indonesia*)

Hirroe Wijaya Ani Kesuma¹ dan Feddy Setio Pribadi²

Jurusan Teknik Elektro, Fakultas Teknik, Universitas Negeri Semarang, Indonesia
hirroewijaya.ptik@gmail.com¹, feddy.setio@mail.unnes.ac.id²

Abstrak— Kitab Undang-Undang Hukum Dagang (KUHD) merupakan salah satu kitab hukum yang dipakai di Indonesia. KUHD terbagi atas dua kitab dan 23 bab. Permasalahan yang muncul adalah mencari bagian pasal maupun ayat yang dilakukan secara manual. Oleh karena itu dibutuhkan suatu aplikasi yang menerapkan konsep sistem temu kembali (*Information Retrieval*) yang dapat melakukan pencarian kembali suatu indeks berdasarkan *query* yang diinput *user*. Salah satu algoritma yang dapat digunakan untuk mencocokkan dan menghitung kedekatan antara *query* dengan indeks suatu pasal tertentu dalam kitab adalah *Vector Space Model* yang menggunakan nilai cosinus untuk melihat kedekatan antara keduanya. Metode ini dikenal dengan *Cosine Similarity*. Kemudian untuk melihat kinerja sistem dalam menerapkan metode ini, dapat dilihat melalui perhitungan *performance measure* dengan nilai *recall* sebesar 55,04%, nilai *precision* sebesar 63,33%, dan nilai *f-measure* sebesar 56,93%.

Kata kunci— Kitab Undang-Undang Hukum Dagang, *information retrieval*, *Cosine Similarity*, *performance measure*

I. PENDAHULUAN

Hukum Dagang adalah hukum yang mengatur tingkah laku manusia yang turut melakukan perdagangan dalam usahanya memperoleh keuntungan. Dapat juga dikatakan, Hukum Dagang ialah hukum yang mengatur hubungan hukum antara manusia-manusia dan badan-badan hukum satu sama lainnya, dalam lapangan perdagangan [1].

Kitab Undang-Undang Hukum Dagang (KUHD) Indonesia kira-kira satu abad yang lalu telah dibawa orang Belanda ke tanah air kita, mula-mula ia hanya berlaku bagi orang-orang Eropa di Indonesia (berdasarkan asas konkordansi). Kemudian juga dinyatakan berlaku bagi orang-orang Timur Asing, akan tetapi tidak berlaku seluruhnya untuk orang-orang Indonesia (hanya bagian-bagian tertentu saja). KUHD yang mulai berlaku di Indonesia pada 1 Mei 1848 terbagi atas dua kitab dan 23 bab. Kitab I terdiri dari 10 bab dan Kitab II terdiri dari 13 bab [1].

Permasalahan yang muncul adalah mencari bagian pasal maupun ayat yang dilakukan secara manual. Dengan kemajuan teknologi yang sangat pesat pada saat ini dibutuhkan sebuah aplikasi yang dapat mengakomodir kebutuhan setiap pengguna akan sebuah aplikasi pengganti buku yang mudah dibawa serta dapat digunakan kapanpun dan dimanapun dengan efektif [2].

Melihat kebutuhan untuk tersedianya suatu alat bantu dalam mencari indeks dalam Kitab Undang-Undang Hukum Pidana, maka dibuat suatu aplikasi yang dapat menggantikan kitab. Aplikasi yang dibutuhkan tentunya perlu menerapkan suatu konsep sistem temu balik informasi (*Information*

Retrieval). Konsep ini bertujuan menghasilkan dokumen yang paling relevan berdasarkan *keyword* pada *query* yang diberikan pengguna. Dokumen dianggap relevan jika suatu dokumen cocok dengan pertanyaan pengguna [3].

Penentuan relevansi ini tentu membutuhkan suatu metode yang berfungsi untuk melakukan pencocokan. Di dalam *text mining*, dikenal suatu algoritma yang dapat digunakan untuk melakukan pencocokan yang dikenal dengan algoritma *Vector Space Model*.

II. METODE

Pada algoritma *vector space model* digunakan rumus untuk mencari nilai *cosinus* sudut antara dua *vector* dari setiap bobot dokumen (WD) dan bobot dari kata kunci (WK). Rumus yang digunakan dalam *vektor space model* sebagai berikut [4]

$$\text{Cosinus} \rightarrow \text{sim}(d_j, q) = \frac{d_j \cdot q}{|d_j| \cdot |q|} = \frac{\sum_i W_{ij} \cdot W_q}{\sqrt{\sum_i W_{ij}^2} \cdot \sqrt{\sum_i W_q^2}} \quad (1)$$

Keterangan :

W_{ij} = bobot kata i pada dokumen j

W_q = bobot *query*

Perhitungan nilai *cosinus* sudut antara dua vektor ini dikenal dengan metode *Cosine Similarity*. Nilai sudut *cosinus* antara dua vektor menentukan kesamaan dua buah objek yang dibandingkan dimana nilai terkecil adalah 0 dan nilai terbesar adalah 1.

Metode evaluasi yang biasanya digunakan pada *information retrieval* adalah metode intrinsik dengan *Precision / Recall* dan *F-Measures* yang dirumuskan dengan rumusan [5] :

$$Recall = \frac{\text{pasal relevan output sistem}}{\text{pasal relevan}} \quad (2)$$

$$Precision = \frac{\text{pasal relevan output sistem}}{\text{pasal output sistem}} \quad (3)$$

$$F - \text{measure} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

III. HASIL DAN PEMBAHASAN

Hasil pengujian dari penerapan *Cosine Similarity* ditunjukkan pada Tabel I. Pengamatan dilanjutkan pada pencocokan *query* dengan *document collection* yang berjumlah 984 pasal dilakukan dengan 4 variasi. Variasi ini didasarkan pada jumlah kata (*term*) yang digunakan pada *query* yaitu 1 hingga 4 buah kombinasi kata yang digunakan bersamaan. Pada masing-masing variasi diuji coba sebanyak 3 kali, sehingga terdapat 12 *query* yang diinputkan.

TABEL I. HASIL UJI UNTUK *PERFORMANCE MEASURE*

<i>Query</i>	Jumlah Output Sistem	Jumlah Pasal Relevan Output Sistem	Jumlah Pasal Relevan
kasir	3	2	2
makelar	10	6	12
pembukuan	3	2	5
pembukuan kasir	6	4	7
perusahaan firma	10	8	20
perseroan terbatas	10	4	21
penyimpangan perbuatan dagang	10	6	10
hak pimpinan firma	10	7	10
bongkar muat kapal	10	8	14
hak waris perusahaan dagang	10	3	5
alokasi dana pembukuan perusahaan	10	8	14
jual beli sewa kapal	10	6	12

Ada beberapa *query* yang menghasilkan output kurang dari 10 pasal. Hal ini berarti bahwa *term* tersebut disebutkan pada kurang dari 10 buah pasal jika dibandingkan dengan *term* yang lain. Salah satu *term* yang menghasilkan output pasal kurang dari 10 buah adalah *term* “kasir”.

Sangat jarang ditemukan suatu kombinasi *query* mewakili 1 pasal tertentu. Dampaknya untuk beberapa *query* yang memiliki salah satu *term* yang diindekskan secara khusus, justru tidak memberikan hasil yang tepat. Hal ini terjadi karena bobot dari salah satu *term* yang bersifat umum lebih besar daripada bobot *term* yang telah terindeks itu sendiri. Tentu besarnya bobot ini akan dapat menggeser pasal yang diwakili oleh *term* yang bersifat khusus saat dilakukan perhitungan dengan *Cosine Similarity*.

Pada masing-masing pasal, besarnya nilai *similarity* tidak terpengaruh dengan panjang pendeknya kalimat yang dimuat dalam setiap pasal. Meski suatu pasal memiliki jumlah kalimat / kata yang relative banyak, namun selama nilai *dot product* -nya (perkalian *W query* dan *W dokumen ke-i*) adalah 0 (nol), maka nilai *similarity*-nya akan 0 (nol).

Analisa tentang besarnya nilai *performance measure* dilakukan dengan melakukan perhitungan *recall*, *precision* dan *f-measure*. Nilai *performance measure* dapat dilihat pada Tabel II. Berdasarkan *query* yang diinput sejumlah 12 kata, diperoleh nilai rata-rata *recall* adalah 55,04%, nilai rata-rata untuk *precision* adalah 63,33% dan nilai rata-rata *f-measure* adalah 56,93%. Angka ini cukup dapat menggambarkan tingkat kemampuan sistem dalam melakukan fungsi temu kembali di dalam aplikasi.

TABEL II. NILAI *RECALL*, *PRECISION* DAN *F-MEASURE*

Jumlah Kata Query	<i>Recall</i>	<i>Precision</i>	<i>F-Measure</i>
1	100,00%	66,67%	80,00%
	50,00%	60,00%	54,55%
	40,00%	66,67%	50,00%
2	57,14%	66,67%	61,54%
	40,00%	80,00%	53,33%
	19,05%	40,00%	25,81%
3	60,00%	60,00%	60,00%
	70,00%	70,00%	70,00%
	57,14%	80,00%	66,67%
4	60,00%	30,00%	40,00%
	57,14%	80,00%	66,67%
	50,00%	60,00%	54,55%
Rata-Rata	55,04%	63,33%	56,93%

IV. SIMPULAN DAN SARAN

Berdasarkan hasil pengamatan dan uji coba, metode *Cosine Similarity* berhasil untuk melakukan fungsi pencarian pasal berdasarkan *query* pengguna terhadap indeks di dalam kitab. Tingkat keberhasilan dari metode ini dalam menjalankan fungsinya dapat dilihat melalui besarnya nilai uji kinerja sistem (*performance measure*) yang menghasilkan nilai 55,04% untuk *recall*, 63,33% untuk *precision* dan 56,93% untuk *f-measure*.

REFERENSI

- [1] Kansil. 2008. Pokok-Pokok Pengetahuan Hukum Dagang Indonesia. Sinar Grafika. Jakarta
- [2] Harnanto, AD. 2013. Aplikasi Pendukung Pembelajaran Bahasa Isyarat pada Anak Berkebutuhan Khusus. http://eprints.ums.ac.id/26127/15/Naskah_Publikasi.pdf. 2 Februari 2016 (12.21).
- [3] Fitri, Meisya. Perancangan Sistem Temu Balik Informasi dengan Metode Pembobotan Kombinasi TF-IDF untuk Pencarian Dokumen Berbahasa Indonesia. <http://jurnal.untan.ac.id/index.php/justin/article/view/1319>. 15 Agustus 2015 (13.42).
- [4] Andika, Ari. 2015. Perancangan Aplikasi Pengukuran Similaritas pada Dokumen dengan Metode Semantic. Majalah Ilmiah Informasi dan Teknologi Ilmiah (INTI) V(3): 13-19.
- [5] Zaman, B. dan Winarko, E. 2011. Analisis Fitur Kalimat untuk Peringkat Teks Otomatis pada Bahasa Indonesia. Indonesia Journal Computing and Cybernetics System (IJCCS) 5(2): 60-68