SISTEM PENJAWAB PERTANYAAN PUBLIK DI KECAMATAN KALIWATES JEMBER MENGGUNAKAN METODE COSINE SIMILARTIY

[1]Rama Nur Jazuli Setiawan, [2]Deni Arifianto, S.Kom, [3]Hardian Oktavianto, S.Si

Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember

Email : rmjz@outlook.com

ABSTRAK

Kecamatan memiliki fungsi memberikan berbagai pelayanan publik yang diperlukan oleh masyarakat untuk memenuhi kebutuhannya dalam berbagai bidang. Pada era digitalisasi saat ini kebutuhan masyarakat akan informasi menjadi semakin kompleks. Akan tetapi perkembangan informasi yang cukup pesat ini dirasa masih cukup sulit untuk diterapkan dalam bidang pelayanan-pelayanan di kecamatan. Permasalahan yang dihadapi adalah bagaimana mengelola banyaknya dokumen agar dapat menjadi informasi yang dibutuhkan masyarakat. Sistem tanya jawab adalah salah satu cara yang dapat dikembangkan untuk mengelola banyaknya dokumen . Metode yang cukup populer untuk membantu pengembangan sistem tanya jawab yaitu metode *cosine similarity.* Dengan memanfaatkan fitur *SMS Gateway*, nantinya masyarakat dapat menanyakan informasi yang dibutuhkan hanya melalui sms. Dalam pembuatan aplikasi SMS Gateway dibutuhkan proses text mining untuk mengolah dokumen yang akan dijadikan acuan dasar berupa Penyusunan Standar Pelayanan Publik. Dokumen tersebut kemudian diklasifikasikan dengan cara menghitung *Term Frequency/Inverse Document Frequency (TF/IDF).* Dengan melakukan *text processing* : *tokenizing, stoplist* dan *stemming* pada dokumen kemudian melakukan pencarian bobot dengan mencari *term*, tf dan idf. Selanjutunya dilakukan pencarian relevansi antara SMS yang masuk dengan dokumen yang dijadikan acuan dasar menggunakan *cossine similarity* sehingga SMS yang masuk dapat dikategorikan sesuai dengan dokumen yang dijadikan acuan dasar. Dari hasil uji coba yang telah dilakukan metode *cosine similarity* dalam Sistem Penjawab Pertanyaan Publik ini memiliki tingkat akurasi sebesar 85% sehingga cocok diterapakan dalam Sistem ini.

Kata Kunci : *Cosine Similarity*, Sistem Penjawab Pertanyaan Publik, *SMS Gateway*

I. Pendahuluan

Kecamatan memiliki fungsi memberikan berbagai pelayanan publik yang diperlukan oleh masyarakat, mulai dari pelayanan dalam bentuk pengaturan atau pun pelayanan-pelayanan lain dalam rangka memenuhi kebutuhan masyarakat dalam berbagai bidang. Pada era digitalisasi saat ini kebutuhan masyarakat akan informasi menjadi semakin kompleks.Teknologi informasi pun berkembang dengan pesat, akan tetapi perkembangan informasi yang cukup pesat ini dirasa masih cukup sulit untuk diterapkan didalam bidang pelayanan pemerintahan khususnya di kecamatan. Masalah yang dihadapi adalah bagaimana mengelola banyaknya dokumen agar dapat menjadi informasi yang dibutuhkan masyarakat. Sistem tanya jawab adalah salah satu cara yang dapat dikembangkan untuk mengelola informasi yang dibutuhkan masyarakat. Metode yang digunakan untuk membantu pengembangan sistem tanya jawab yaitu metode *cosine similarity.* Dengan memanfaatkan fitur SMS Gateway, nantinya masyarakat dapat menanyakan informasi yang dibutuhkan hanya melalui sms sehingga diharapkan nantinya masyarakat dapat memperoleh informasi dengan mudah dan tepat mengenai

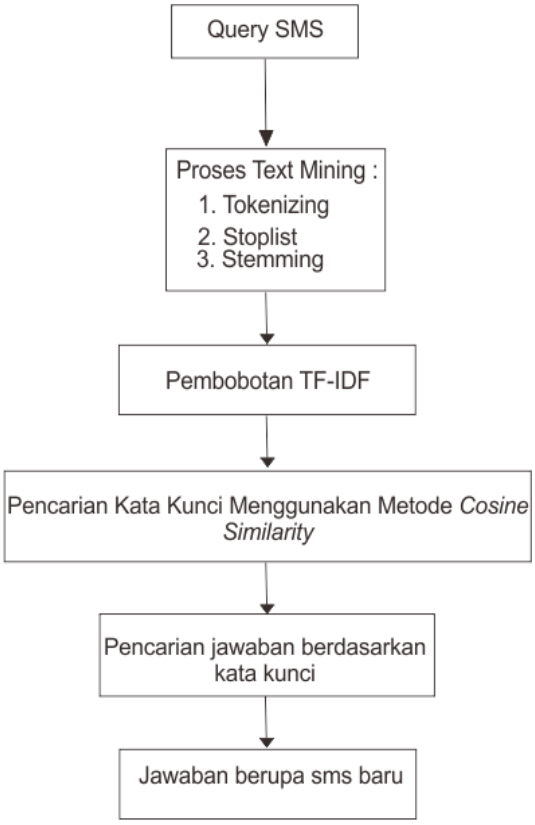
pelayanan-pelayanan yang ada di lingkungan kecamatan.

II. Metodologi Penelitian

Penelitian ini dilakukan dengan tahapan sebagai berikut :

1. Studi Literatur dalam metode ini, pengumpulan data dilakukan dengan cara meminta sampel SPP di Kecamatan Kaliwates Kabupaten Jember.
2. Data dokumen berupa kumpulan sms tentang pertanyaan-pertanyaan yang berhubungan dengan pelayanan publik di Kecamatan.
3. Implementasi merupakan tahapan untuk menerapkan metode *cosine similarity.*
4. Pengujian dilakukan dengan menggunakan aplikasi dekstop dan bahasa pemrogaman Visual Basic 6.0.
5. Pertanyaan SMS yang diterima server dari client akan dianalisa menggunakan metode *cosine similarity.*

Rancangan alur sistem dari SMS Gateway :



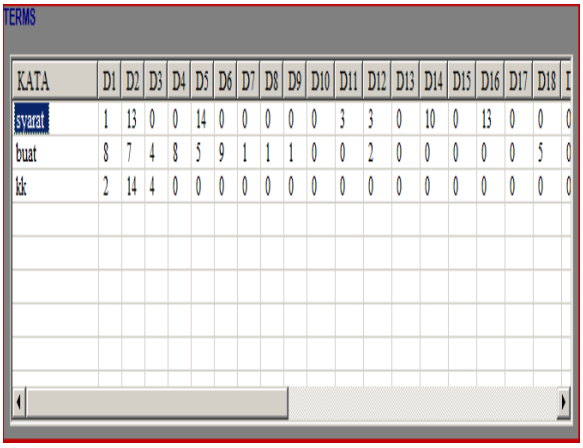
III. Hasil dan Pembahasan

1. Proses *Stemming*

Pada tahap ini sistem akan melakukan 3 proses pemecahan kalimat yaitu :

- *Tokenizing*
  Proses pemecahan kalimat menjadi kata.
- *Stemming*
  Proses pencarian kata dasar dari kata yang telah di *tokenizing*
- *Stopword*
  Proses pembuangan kata yang tidak diperlukan.

2. Menghitung Term Frequency dalam dokumen



TERMS

KATA	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18	D
syarat	1	13	0	0	14	0	0	0	0	0	3	3	0	10	0	13	0	0	0
buat	8	7	4	8	5	9	1	1	1	0	0	2	0	0	0	0	0	5	0
kk	2	14	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Proses ini mencari nilai *Term Frequency*(*tf*) yaitu jumlah kemunculan atau frekuensi(*f*) suatu *term*(kata) *i* didalam sebuah dokumen *j*.

$$tf_{ij} = f_{ij}$$

dan setelah itu mencari nilai *Inverse Dokumen Frequency* (*idf*) yaitu logaritma dari jumlah dokumen keseluruhan (n) dibagi dengan jumlah *term*.

$$idf_i = log\left(\frac{n}{df_i}\right)$$

3. Pembobotan TF-IDF

TFxIDF

D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D1
-0.2361	-3.0633	0	0	-3.3054	0	0	0	0	0	-0.7083	-0.7083	0
-1.6568	-1.4497	-0.8284	-1.6568	-1.0355	-1.8639	-0.2071	-0.2071	-0.2071	0	0	-0.4142	0
0.5106	3.5742	1.0212	0	0	0	0	0	0	0	0	0	0
-1.8929	-0.9448	0.1928	-1.6568	-4.3409	-1.8639	-0.2071	-0.2071	-0.2071	0	-0.7083	-1.1225	0

Pembobotan TF-IDF dilakukan dengan mengalikan hasil pembobotan TF dan IDF.

$$W_{ij} = tf_{ij}.idf_{ij}$$

4. Perhitungan Wdq x Wdi

WDQxWDi

D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11
0.7247	0.7247	0	0	0.7804	0	0	0	0	0	0.1672
0.3002	0.3002	0.1716	0.3431	0.2145	0.386	0.0429	0.0429	0.0429	0	0
0.9125	0.9125	0.2607	0	0	0	0	0	0	0	0
1.0249	1.9374	0.4323	0.3431	0.9949	0.386	0.0429	0.0429	0.0429	0	0.1672

Dalam tahap ini bobot dari *query* dikalikan dengan bobot dari tiap dokumen. Kemudian nilai dari tiap term hasil perkalian di jumlahkan.

5. Normalisasi Term

NORMALISASI

D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11
0.0557	9.4206	0	0	10.9257	0	0	0	0	0	0.5017
2.745	2.1016	0.6862	2.745	1.0723	3.4741	0.0429	0.0429	0.0429	0	0
0.2607	12.7749	1.0428	0	0	0	0	0	0	0	0
2.8007	4.929...	1.314...	1.656...	3.463...	1.863...	0.207...	0.207...	0.207...	0	0.708...

Proses normalisasi dilakukan dengan cara mengkuadratkan bobot setiap *term* dalam setiap dokumen termasuk *Q*, kemudian nilai kuadrat dari setiap *term* di jumlahkan dan terakhir di akarkan.

6. *Cosine Similarity*

SIMILARITY

Nilai	Dokumen
D1 =	0.9042357...
D2 =	0.9712004...
D3 =	0.8123712...
D4 =	0.5117010...
D5 =	0.7097277...
D6 =	0.5117206...
D7 =	0.5117942...
D8 =	0.5117942...
D9 =	0.5117942...
D10 =	0

Proses terakhir yaitu proses pencarian *relevansi* antar dokumen menggunakan metode *cosine similarity*. Perhitungannya dilakukan dengan cara hasil penjumlahan dari proses perkalian Wdq x Wdi dibagi dengan hasil akar dari proses normalisasi. Sehingga memunculkan nilai terbesar dari *query* yang dianggap paling relevan dengan dokumen.

7. Perhitungan tingkat akurasi sistem

Rumus perhitungan

$$Tingkat\ kesalahan = \frac{Banyak\ data\ benar}{Jumlah\ Data}\ x\ 100\ \%$$

$$Tingkat\ kesalahan = \frac{17}{20}\ x\ 100\ \% = 85\ \%$$

Berdasarkan uji coba di atas, sistem ini memiliki tingkat akurasi sebesar 85%.

IV. Kesimpulan

Kesimpulan dari Sistem Penjawab Pertanyaan Publik dengan menggunakan metode *Cosine Similarity* adalah

1. Metode *Cosine Similarity* dalam Sistem Penjawab Pertanyaan Publik ini memiliki tingkat akurasi sebesar 85%.
2. Metode Cosine Similarity sesuai diterapkan dalam Sistem Penjawab Pertanyaan Publik karena tingkat kesalahannya relatif rendah.

V. Saran

Untuk pengembangan lebih lanjut dari sistem ini diberikan saran – saran yang berguna untuk pemikiran maupun implementasinya yaitu :

1. Semakin banyak data input yang digunakan dalam membangun Sistem Penjawab Pertanyaan Publik di Kecamatan Kaliwates ini akan sangat membantu sistem untuk menjawab pertanyaan Masyarakat dengan tingkat kesalahan yang rendah.
2. Adanya fitur ketika tidak ada jawaban dalam aplikasi maka akan dijawab manual oleh petugas atau admin.

Daftar Pustaka

[1] Fahmi, K., (2012). *Rancang Bangun Search Engine Ayat Al-Qur'an yang Mampu Memproses Teks Bahasa Indonesia Menggunakan Metode Cosine Similarity*. Universitas Islam Negeri Maulana Malik Ibrahim Malang: Malang

[2] Gunawan., Lovina, G. Jan 2005. *Question answering system dan penerapannya pada alkitab*. Jurnal Informatika, 7(1), 1-9.

[3] Kecamatan Lubai., (2013). "Pengertian Kecamatan". Diakses pada 29 Oktober 2014 http://keclubai.blogspot.com/2013/02/pengertian-kecamatan.html#.VFmP4xbVSKs.

[4] Muchlisin, R., (2012). "Short Message Service (SMS)". Diakses pada 27 Oktober 2014, dari http://www.kajianpustaka.com/2012/12/teori-sms-short-message-service.html.

[5] Ramadhika, A., (2012). "SMS Gateway". Diakses pada 27 Oktober 2014, dari http://www.ubaya.ac.id/2013/content/articles_detail/33/SMS-Gateway-menggunakan-Gammu-dan-MySQL.html.

[6] Riloff, E., Thelen, M. 2003. *A rule-based question answering system for reading comprehension tests*. Retrieved December 1, 2008, from University of Institut Pertanian Bogor.

[7] Singgasana Kecamatan Kaliwates., (2011). "Profil Kecamatan Kaliwates". Diakses pada 29 Oktober 2014, dari http://kecamatankaliwates.blogspot.com/2011/12/profil-kecamatan-kaliwates.html.

[8] Sulistyo., (2008). "*Trem Frekuensi-Invers Document*", Universitas Diponegoro, Semarang.

[9] Tekli., (2009). "*Skema Pembobotan TF-IDF*", Universitas Diponegoro, Semarang.

[10] Widodo, A., (2012). *The Text Mining Handbook*, Jurnal Ilmiah Ilmu Komputer, Surabaya.