

SISTEM TEMU KEMBALI CERITA RAKYAT NUSANTARA MENGUNAKAN METODE *VECTOR SPACE MODEL*

Erik Ratu Prawira Negara¹, Fatkhul Amin²

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Stikubank Semarang
e-mail : ¹rick_licious@yahoo.com ²fatkhulamin@edu.unisbank.ac.id

ABSTRAK

Perkembangan teknologi informasi memberikan dampak banyaknya data dokumen yang tersedia sehingga dibutuhkan banyak waktu untuk menemukan sebuah dokumen. Banyaknya data yang terambil pada saat pencarian dianggap tidak efektif, yang berdampak pada rendahnya akurasi hasil pencarian. Cerita rakyat nusantara merupakan cerita yang berasal dari masyarakat indonesia yang disampaikan secara turun-temurun secara lisan dengan harapan adanya nilai-nilai kebaikan dan pelajaran moral yang didapat dari cerita tersebut. Banyaknya jumlah cerita rakyat nusantara dianggap menjadi kendala karena terbatasnya sarana yang tersedia untuk menemukan cerita yang relevan dengan kebutuhan. Sehingga dibuatlah sebuah sistem temu kembali informasi (STKI) dengan menggunakan vector space model (VSM). Hasil dari pembuatan mesin pencari yang dibuat memiliki tingkat presisi yang tinggi.

Kata Kunci : *Vector Space Model, STKI, cerita rakyat.*

1. PENDAHULUAN

Berkembangnya teknologi saat ini berdampak pada banyaknya informasi yang beredar di masyarakat, informasi ini beragam macam jenis dan jumlahnya. Diperlukan adanya sebuah alat yang dapat memudahkan untuk melakukan pencarian data informasi. Salah satu cara yang dapat dilakukan untuk menemukan data yang relevan adalah dengan menggunakan sebuah mesin pencari otomatis. Mesin pencarian otomatis yang di kenal juga sebagai Sistem Temu Kembali Informasi (STKI) merupakan sebuah alat yang dapat membantu untuk menemukan sebuah data dengan cara memproses sebuah kata kunci yang nantinya akan membantu menemukan sebuah data informasi yang dibutuhkan oleh pengguna (Boubekeur, 2013).

Pelajaran moral dan budi pekerti mulai dilupakan dan ditinggalkan oleh sebagian masyarakat indonesia. Salah satu cara orang tua untuk menanamkan nilai moral dan budi pekerti adalah melalui cerita. Cerita rakyat di indonesia ada beragam, dan cerita itu yang telah lama berkembang di masyarakat dan terus diceritakan secara turun-temurun. Ada beberapa cara untuk dapat memperoleh informasi tentang cerita rakyat nusantara, diantaranya melalui berbagai macam buku cerita rakyat, bertanya kepada orang dan melalui cerita yang banyak beredar di internet. mencari cerita di dalam buku dan internet dianggap kurang efisien karena memerlukan banyak waktu dan biaya untuk dapat menemukan cerita yang sesuai dan bila bertanya kepada orang yang sering terkendala oleh lemahnya ingatan orang tersebut yang dapat merubah jalur cerita sebagian maupun secara keseluruhan. Diperlukan adanya sebuah alat yang dapat digunakan untuk menampung cerita rakyat sehingga dapat mempermudah untuk menemukan sebuah cerita rakyat nusantara yang relevan dengan kebutuhan dengan waktu yang singkat.

Berdasarkan hal tersebut maka diperlukan sebuah alat pencarian informasi yang dapat menampilkan hasil pencarian cerita rakyat nusantara dengan menggunakan Sistem Temu Kembali Informasi (STKI) dengan menggunakan *Vector Space Model* (VSM). VSM dipilih karena dapat menampilkan hasil temuan pencarian yang tidak terlalu banyak, sehingga dapat meningkatkan ketepatan hasil pencarian sehingga dapat mempersingkat waktu pencarian yang dibutuhkan untuk melakukan pencarian cerita yang relevan.

Berdasar latar belakang di atas guna untuk membantu proses pencarian Cerita Rakyat Nusantara maka penulis mengambil penelitian dengan judul “Rancang Bangun Sistem Temu Kembali Dengan VSM Studi Kasus Cerita Rakyat Nusantara”.

2. TINJAUAN PUSTAKA

2.1. Sistem Temu Kembali Informasi

Sistem Temu Kembali Informasi (STKI) atau yang dikenal sebagai *Information Retrieval System* (IRS) adalah sebuah sistem yang berfungsi untuk dapat menemukan (*Retrieve*) informasi yang dibutuhkan oleh User secara otomatis (Amin, 2012). Prinsip kerja sistem temu kembali informasi adalah dengan mencari sebuah dokumen yang relevan dari sekumpulan dokumen yang ada pada sistem dan membuang dokumen yang tidak relevan (Salton, 1989).

2.2 Tokenisasi

Tokenisasi merupakan proses pemisahan serangkaian karakter yang dipisahkan oleh karakter spasi, dan pada waktu yang bersamaan juga dilakukan penghapusan karakter tertentu seperti tanda baca. Token pada umumnya disebut juga sebagai istilah (*term*) atau kata, sebagai contoh sebuah token adalah serangkaian karakter yang berurutan dari suatu dokumen yang dikelompokkan sebagai unit semantik yang berguna untuk diproses (Salton, 1989).

2.3 Filtering

Proses *Filtering* atau penyaringan kata dengan menggunakan daftar *stopword* yang digunakan oleh Tala (2003). *Stopword* yang digunakan merupakan *stopwords* dalam bahasa Indonesia yang berisi kata-kata seperti; ke, ada, kepada, yang dan lain sebagainya. Penelitian ini menggunakan list *stopwords* Tala sebagai daftar *stopwords* yang digunakan dalam penelitian ini.

Stopwords adalah kata yang umum dan biasanya sering muncul dalam suatu dokumen dalam jumlah yang besar serta dianggap tidak memiliki suatu makna yang spesifik. Pada umumnya *stopwords* dimanfaatkan dalam bagian dari sebuah task information retrieval. Mesin pencari berbasis web seperti *google* juga memanfaatkan *stopword* sebagai bagian dari *task information retrieval* miliknya. Contoh dari *stopword* dalam bahasa Inggris diantaranya adalah “of”, “the”. Sedangkan *Stopword* yang umum digunakan dalam bahasa Indonesia diantaranya adalah “ke”, “di”, “yang” (Tala, 2003).

2.4 Stemming

Stemming adalah proses yang terdapat dalam sistem temu kembali informasi yang berfungsi untuk mengubah kata-kata yang terdapat dalam suatu dokumen menjadi bentuk kata-kata dasarnya (*root word*) dengan menggunakan peraturan tertentu (Tala, 2003). Proses *stemming* dalam bahasa Indonesia berbeda dengan *stemming* dalam bahasa Inggris yang hanya menghilangkan sufiks dari sebuah kata, sedangkan dalam teks berbahasa Indonesia proses *stemming* juga akan menghilangkan sufiks, prefiks, dan konfiks.

Proses *stemming* digunakan untuk mengubah *term* yang melekat pada *term* itu sendiri seperti awalan, sisipan dan akhiran. Proses *stemming* dilakukan dengan cara menghilangkan semua imbuhan (*affixes*) yang ada pada sebuah *term*. Imbuhan tersebut dapat terdiri dari awalan (*prefixes*), sisipan (*infixes*) dan akhiran serta dapat juga kombinasi antara awalan dan akhiran (*confixes*). *Stemming* digunakan untuk mengubah atau mengganti bentuk dari suatu kata menjadi kata dasar yang sesuai dengan struktur morfologi bahasa Indonesia yang benar (Nazief, Bobby dan Mirna Adriani, 1996).

2.5 Vector Space Model

VSM adalah salah satu metode pencarian informasi yang dilakukan dengan cara merepresentasikan sebuah dokumen sebagai sebuah vektor (Manning, 2007). Dokumen dipandang sebagai sebuah vektor yang memiliki jarak (*magnitude*) dan arah (*direction*). VSM digunakan dalam STKI karena metode ini merupakan metode yang efisien karena mudah diimplementasikan serta dapat melihat tingkat kedekatan dan kesamaan (*Similarity*) sebuah kata kunci dengan cara menghitung bobot dari sebuah kata (*term*) (Amin, 2012). Kesesuaian antara dokumen dengan kata kunci berdasar dari kesamaan antara sebuah vektor dokumen dengan vektor dari kata kunci (Baeza, 1999).

VSM memungkinkan untuk dapat memberikan kecocokan parsial dari sebuah kerangka. Hal ini dapat dicapai dengan menetapkan bobot non-biner untuk istilah indeks dalam *query* dan dokumen. Bobot merupakan istilah yang digunakan untuk menghitung kesamaan (*similarity*) antara tiap-tiap dokumen yang terdapat dalam sistem dengan permintaan pengguna. Dokumen yang terambil dipilih dan diurutkan berdasarkan tingkat kemiripan dengan kata kunci yang digunakan, model vektor memperhitungkan dan

mempertimbangkan dokumen yang memiliki kesamaan dan relevan dengan kebutuhan pengguna. Hasil akhirnya adalah sekumpulan dokumen yang terambil lebih akurat dan sesuai dengan kebutuhan pengguna. Persamaan berikut digunakan untuk menghitung jarak *query*.

$$|q| = \sqrt{\sum_{j=1}^t (w_{i,q})^2} \dots\dots\dots(1)$$

Nilai jarak *query* |q| dihitung untuk mendapatkan jarak query dari bobot *query* dokumen W_{iq} yang terambil oleh sistem. Jarak *query* dihitung dengan persamaan akar jumlah kuadrat dari *query*.

Persamaan berikut digunakan untuk menghitung jarak dokumen.

$$|d_j| = \sqrt{\sum_{i=1}^t (w_{i,j})^2} \dots\dots\dots(2)$$

Nilai Jarak dokumen |dj| dihitung untuk mendapatkan jarak dokumen dari bobot dokumen dokumen W_{ij} yang terambil oleh sistem. Jarak dokumen dihitung dengan persamaan akar jumlah kuadrat dari dokumen.

Perhitungan pengukuran similaritas query document (inner product). Menggunakan persamaan berikut :

$$Sim(q,d_j) = \sum_{i=1}^t W_{iq} \cdot W_{ij} \dots\dots\dots(3)$$

Similaritas antar *query* dan dokumen atau *inner product/Sim* (q,dj) digunakan untuk mendapatkan bobot dengan didasarkan pada bobot *term* dalam dokumen (W_{ij}) dan bobot *query* (W_{iq}) atau dengan cara menjumlah bobot q dikalikan dengan bobot dokumen.

Pengukuran *Cosine similaritas* (menghitung nilai kosinus sudut antara dua vector) menggunakan persamaan berikut :

$$CosSim(q,d_j) = \frac{q \cdot d_j}{|q| \cdot |d_j|} = \frac{\sum_{i=1}^t W_{iq} \cdot W_{ij}}{\sqrt{\sum_{i=1}^t (w_{iq})^2} \cdot \sqrt{\sum_{i=1}^t (w_{ij})^2}} \dots\dots(4)$$

Similaritas antara *query* dan dokumen atau $Sim(q,d_j)$ berbanding lurus terhadap jumlah bobot query (q) dikali bobot dokumen (d_j) dan berbanding terbalik terhadap akar jumlah kuadrat q (|q|) dikali akar jumlah kuadrat dokumen (| d_j |). Perhitungan similaritas menghasilkan bobot dokumen yang mendekati nilai 1 atau menghasilkan bobot dokumen yang lebih besar dibandingkan dengan nilai yang dihasilkan dari perhitungan *inner product*.

2.6 Pembobotan Kata

Sekumpulan dokumen yang terambil dengan VSM direpresentasikan sebagai sebuah matriks *term* dokumen (atau matriks *term frequency*). Setiap bagian sel yang terdapat dalam matrik bersesuaian dengan bobot yang diberikan dari suatu term dalam suatu dokumen. Nilai nol menunjukkan *term* tersebut tidak ada (tidak ditemukan) dalam dokumen.

Dalam pembobotan dokumen hasil perhitungan yang telah didapat dengan VSM menggunakan persamaan (5) dan (6) selanjutnya dihitung bobotnya berdasarkan banyaknya dokumen dan *term* yang muncul dalam dokumen tersebut.

$$tf = tf_{ij} \dots\dots\dots(5)$$

tf adalah *term frequency* (*tf*) yang nilainya di dapat melalui penghitungan banyaknya kemunculan *term* t_i dalam dokumen d_j .

Perhitungan nilai *Inverse Document Frequency* (*idf*) diperoleh melalui persamaan berikut :

$$idf_i = \log \frac{N}{df_i} \dots\dots\dots(6)$$

Sehingga perhitungan nilai *idfi* digunakan untuk mengetahui banyaknya *term* yang dicari (*dfi*) yang muncul dalam dokumen yang terdapat di dalam *database*.

Perhitungan *term frequency Inverse Document Frequency (tfidf)* dilakukan dengan persamaan berikut :

$$W_{ij} = tf_{ij} \cdot \log \left(\frac{N}{df_i} \right) \dots \dots \dots (7)$$

Dengan W_{ij} adalah bobot dokumen, N adalah banyaknya jumlah dokumen yang terambil oleh sistem, tf_{ij} adalah banyaknya kemunculan *term* t_i dalam dokumen d_j . Nilai *term frequency (tf)* didapat dari banyaknya kemunculan *term* t_i dalam dokumen d_j . serta df_i adalah banyaknya dokumen didalam Koleksi dokumen dimana *term* t_i yang muncul di dalamnya. Nilai dari W_{ij} yang merupakan bobot dari sebuah dokumen di hitung dari perkalian atau kombinasi antara *term frequency (tf_{ij})* dan *Inverse Document Frequency (df_i)*.

2.7 Recall dan Precision

Nilai dari hasil pencarian dari sistem temu kembali informasi dapat dilihat dari tingkat *recall* dan *precision* nya. *Precision* adalah ukuran untuk ketepatan atau ketelitian dari sebuah dokumen dan *recall* adalah kesempurnaan dari sebuah dokumen. Nilai *precision* adalah proporsi dokumen yang terambil adalah relevan. Nilai *recall* adalah proporsi dokumen yang terambil oleh sistem (Salton, 1989).

Recall adalah dokumen yang terpanggil dari sebuah sistem temu kembali informasi yang relevan dengan kebutuhan yang diminta oleh pengguna sesuai dengan pola sistem temu kembali informasi. Nilai *recall* yang tinggi belum dapat menggambarkan bahwa suatu sistem temu kembali informasi itu dikatakan baik atau tidak.

Persamaan berikut adalah persamaan yang digunakan untuk mendapat nilai dari *recall*

$$R = \frac{\text{Number of relevant items retrieved}}{\text{Total number of relevant items in collection}} \dots (8)$$

R adalah nilai dari *Recall*, nilai dari R didapatkan dengan membandingkan antara nilai *Number of relevant item retrieved* dengan *Total number of relevant item in collection*.

Precision adalah dokumen yang terpanggil dari sekumpulan dokumen dalam sebuah basis data (*database*) yang sesuai dengan kebutuhan informasi pengguna setelah dinilai sendiri oleh pengguna tersebut. Semakin besar nilai *precision* dari sebuah sistem temu kembali informasi, maka sistem temu kembali informasi tersebut dapat di katakan baik. Persamaan berikut adalah persamaan yang digunakan untuk mendapat nilai dari *Precision*

$$P = \frac{\text{Number of relevant items retrieved}}{\text{Total number of items retrieved}} \dots \dots (9)$$

Dengan P adalah *precision*. Sehingga nilai dari P di dapat dengan membandingkan nilai antara *Number of relevant items* dengan *Total number of item retrieved*.

2.8 Cerita Rakyat

Menurut Kamus Besar Bahasa Indonesia, cerita rakyat adalah cerita yang terjadi di jaman dahulu yang terus hidup di kalangan masyarakat dan diwariskan secara lisan. Cerita rakyat adalah cerita yang berasal dari masyarakat dan berkembang dalam masyarakat. Umumnya cerita rakyat nusantara berupa cerita dongeng. Menurut Retno Purwandari dan Qani'ah cerita dongeng adalah suatu cerita yang bersifat khayal.

Dongeng memiliki berbagai macam jenis dan ragamnya, yaitu sebagai berikut :

- a. Fabel
Fabel merupakan cerita lama yang menokohkan binatang sebagai lambang pelajaran moral (bisa pula di sebut sebagai cerita binatang).
- b. Mitos (*mite*)

Mitos adalah cerita-cerita yang berhubungan dengan kepercayaan terhadap suatu benda atau hal yang dipercayai mempunyai kekuatan gaib.

c. Legenda

Legenda adalah cerita lama yang menceritakan tentang riwayat terjadinya suatu tempat atau wilayah.

d. Sage

Sage merupakan cerita lama yang berhubungan dengan sejarah, yang menceritakan tentang keberanian, kepahlawanan, kesaktian, dan keajaiban seseorang.

e. Parabel

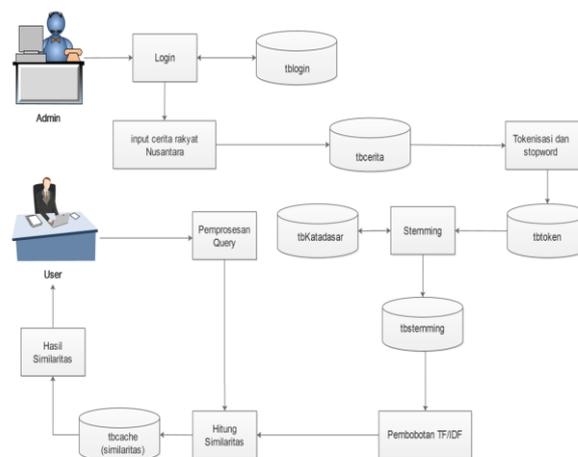
Parabel adalah cerita rekaan yang menggambarkan sikap moral atau keagamaan dengan menggunakan ibarat atau perbandingan.

f. Dongeng Jenaka

Dongeng Jenaka adalah cerita tentang tingkah laku orang bodoh, malas, atau cerdik dan masing masing dilukiskan secara humor.

1.1. Arsitektur Sistem

Arsitektur sistem digunakan untuk memperlihatkan langkah langkah yang digunakan dalam membangun sebuah aplikasi sistem temu kembali informasi serta menjelaskan alur kerja sistem secara sederhana dengan menggunakan gambar yang sederhana yang berupa tahapan tahapan preprocessing pada sistem temu kembali informasi. Serta hubungan antar tiap-tiap prosesnya.



Gambar 1. Arsitektur Sistem

2. METODOLOGI PENELITIAN

Metode yang digunakan pada rancang bangun sistem temu kembali informasi ini adalah sebagai berikut :

a. Objek Penelitian

Objek yang dijadikan bahan dalam penelitian ini adalah sekumpulan dari cerita atau sinopsis Cerita rakyat nusantara yang berasal dari pulau sumatera.

b. Metode Pengumpulan Data

Pengumpulan data yang digunakan untuk penelitian ini diperoleh dari berbagai sumber, sumber perolehan data penelitian ini antara lain :

i. Sumber Data Primer

Data yang di jadikan sumber data primer adalah data yang diperoleh secara langsung kepada nasrasumber dengan melakukan wawancara dan dengan pencarian cerita rakyat nusantara melalui buku cerita rakyat.

ii. Sumber Data Sekunder

Data sekunder merupakan data pendukung yang diperoleh melalui media internet. Cerita yang beredar di internet ini akan disimpan dan dibuat sinopsis ceritanya.

c. Metode Pengumpulan Data

Metode yang digunakan untuk mengumpulkan data adalah sebagai berikut :

i. Wawancara

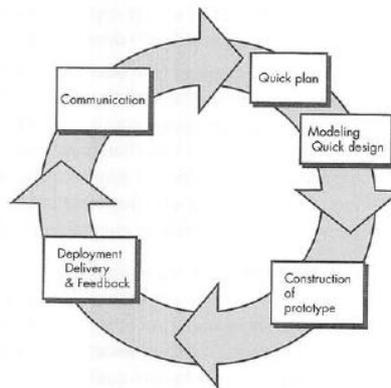
Pengumpulan data dengan wawancara dilakukan dengan melakukan percakapan tanya jawab dengan narasumber dan pihak-pihak yang terkait, baik pihak yang menguasai bidang terkait dengan penelitian maupun narasumber cerita rakyat nusantara.

ii. Studi Pustaka/Literatur

Pengumpulan data dengan metode kepustakaan ini dilakukan dengan mencari data pustaka yang berhubungan dan dapat menunjang penelitian ini. Pustaka-pustaka yang ditemukan akan dipelajari dan digunakan dalam penelitian ini. Pustaka tersebut dapat berupa buku maupun artikel yang terkait dengan penelitian yang berada dan berdasar di internet.

d. Metode Pengembangan Sistem

Metode yang Penulis gunakan untuk dijadikan metodologi pengembangan sistem adalah metode prototyping (prototipe). Metode prototyping dipilih karena adanya komunikasi timbal balik dari pengguna sistem dan pembuat sistem yang dapat membantu perkembangan sistem menjadi lebih baik sehingga dapat mempersingkat waktu pengembangan sistem untuk dapat menghasilkan sistem yang sesuai dengan keinginan pengguna. Hal ini dimungkinkan karena terdapat evaluasi yang dilakukan pada sistem secara berkala, sehingga dengan adanya evaluasi ini bisa didapatkan sebuah sistem yang baik (Amin, 2012).



Gambar 2. *Evolutionary Prototyping* (Pressman, 2005)

Adapun tahapan pembuatan perangkat lunak dengan prototype adalah sebagai berikut :

a. Komunikasi (Communication)

Pada tahap ini pihak pembuat sistem (Developer) dengan pihak pengguna (User) menentukan gambaran umum sistem yang akan dikembangkan. Baik kebutuhan tujuan dan manfaat sistem agar sesuai dengan kebutuhan pengguna.

b. Perencanaan Secara Cepat (Quick Plan)

Tahap ini hasil dari tahap komunikasi yang telah dilakukan oleh Developer dan User akan menghasilkan gambaran yang sama terhadap kebutuhan perangkat lunak yang akan dibuat.

c. Pemodelan Perancangan Secara Cepat (Modeling Quick Design)

Tahap ini memulai hasil dari gambaran yang didapat dari tahap sebelumnya. Aspek-aspek yang menjadi fokus dari kebutuhan pengguna masih berupa gambaran prototipe seperti desain tampilan antarmuka yang menghubungkan pengguna dan sistem.

d. Pembentukan Prototipe (Construction of Prototype)

Tahap ini adalah tahapan pembuatan kerangka dari perangkat lunak yang telah direncanakan. Kerangka ini akan dibuat menjadi perangkat lunak yang terdiri dari kode-kode yang diperlukan untuk membuat fitur pada perangkat lunak.

e. Penyerahan Kepada User dan Umpan Balik (Development Delivery & Feedback)

Tahapan ini adalah tahapan penilaian dari pembuatan perangkat lunak. Pihak Developer akan menyerahkan perangkat lunak kepada User untuk dievaluasi. Evaluasi ini akan menghasilkan sebuah umpan balik baik berupa keluhan, komentar maupun kekurangan yang terdapat didalam perangkat lunak tersebut. Dari umpan balik ini maka proses pengembangan perangkat lunak ini akan kembali ke tahap komunikasi. Hal ini akan terus-menerus berulang sampai semua kebutuhan terpenuhi.

3. IMPLEMENTASI SISTEM

3.1. Implementasi Database

Implementasi dari program sistem temu kembali informasi cerita rakyat nusantara yang telah dirancang memerlukan sebuah *database* sebagai sarana pengelolaan data. Pembuatan *database* pada STKI cerita rakyat nusantara ini menggunakan phpMyAdmin dengan database MySQL. Database yang digunakan bernama “cerita”, didalam database cerita tersebut terdapat beberapa tabel yang bernama *tbcache*, *tbcerita*, *tbkatadasar*, *tbkeyword*, *tblogin*, *tbstemming*, *tbtoken*, dan *tbvektor*.

3.2. Implementasi Program

Berdasarkan pada perancangan tampilan antar muka yang telah dibahas pada bab sebelumnya, dapat dilakukan pembuatan program untuk menjalankan keseluruhan sistem yang ada. Program sistem temu kembali cerita rakyat nusantara yang telah diuji dengan menggunakan Browser Google Chrome versi 47.0.2526.80 m yang sebelumnya telah di instal *appserv*. URL yang digunakan untuk dapat masuk ke halaman sistem temu kembali informasi cerita rakyat nusantara adalah <http://localhost/cerita/index.php>. Melalui alamat URL untuk tersebut dapat digunakan untuk menampilkan halaman utama sistem temu kembali informasi untuk pencarian cerita rakyat nusantara. Setelah memasukkan kata kunci, sistem akan menuju ke halaman hasil pencarian, pada tampilan tersebut juga terdapat halaman login yang digunakan untuk masuk ke halaman admin. Halaman *admin* digunakan untuk menambah, menghapus, melihat cache, melihat bobot dan vektor dari dokumen yang telah dimasukkan ke dalam *database*.

3.3. Halaman Pencarian

Setelah mengakses *link* <http://localhost/cerita/index.php>, sistem akan menuju ke halaman pencarian. Halaman ini terdapat *form*, *button* cari dan *button* login. Untuk melakukan pencarian pengguna harus memasukkan kata kunci pada *form* pencarian, setelah itu dengan menekan *button* cari maka sistem akan menampilkan hasil pencarian pada halaman hasil pencarian. Sedangkan jika *button* login diklik maka pengguna akan di alihkan ke halaman *login* yang dapat digunakan untuk mengakses halaman *admin*.

3.4. Halaman Hasil Pencarian

Setelah menginputkan kata kunci (*keyword*) dari dokumen yang akan di cari, setelah itu user dapat mengklik tombol “Cari” yang terdapat pada halamn pencarian. Halaman ini terdapat *form* untuk memasukkan kata kunci, dan sebuah tombol “Cari” untuk melihat hasil pencarian serta pada halaman pencarian ini terdapat tombol “About” yang berguna untuk menampilkan *biodata* dari pembuat program serta tata cara penggunaan program sistem temu kembali cerita rakyat. Selain itu pada halaman hasil pencarian ini juga akan menampilkan hasil pencarian cerita rakyat nusantara yang telah diinputkan sebelumnya pada *form* pencarian. Hasil pencarian ini menampilkan judul cerita rakyat beserta dengan isi cerita rakyatnya dan nilai kemiripan antara suatu *query* yang telah diinputkan oleh *user* dengan dokumen-dokumen yang terdapat pada sistem temu kembali informasi.

Pengguna dapat langsung mengunjungi situs web dari cerita rakyat nusantara yang ditampilkan pada halaman pencarian dengan mengklik pada judul cerita rakyat nusantara yang telah ditampilkan oleh sistem. Hasil pencarian cerita rakyat nusantara yang memiliki nilai kemiripan tertinggi dengan *query* yang telah di inputkan maka akan menempati urutan teratas pada hasil pencarian, sebaliknya apabila semakin kecil nilai kemiripannya dengan *query* yang telah di inputkan, maka cerita rakyat tersebut akan menempati urutan berikutnya.

3.5. Halaman Login

Apabila admin ingin masuk ke halaman *admin*, Admin dapat mengklik *Button* Login yang terdapat pada halaman pencarian. *Button* login ini menggunakan *Icon* layang-layang yang terdapat dikanan atas halaman *admin*. Setelah mengklik *button* untuk *login*, selanjutnya sistem akan masuk ke halaman yang

berisi tampilan *login*. Halaman *login* ini terdapat *form username*, *form password* serta *button login*. *Form username* di isi dengan nama yang telah di buat oleh *admin* dan *form password* diisi dengan *password* yang telah dibuat *admin* sebelumnya. Setelah mengisi *username* dan *password* dengan benar, maka selanjutnya sistem akan mengarahkan pengguna ke halaman *admin*.

3.6. Halaman Admin

Setelah berhasil *login* dengan memasukkan *username*, *Password* dan mengklik *button login* maka sistem akan menampilkan halaman admin. Halaman *admin* ini merupakan halaman yang ditampilkan setelah berhasil masuk ke dalam hak akses *administrator* dari sistem temu kembali informasi. Halaman *Admin* ini menampilkan beberapa menu dalam sistem temu kembali informasi cerita rakyat nusantara, diantaranya adalah menu lihat cerita, untuk dapat melihat data cerita rakyat nusantara yang telah dimasukkan ke dalam sistem sebelumnya, selain itu pada menu lihat cerita rakyat nusantara ini terdapat fitur hapus yang berguna untuk menghapus dokumen cerita yang telah dimasukkan sebelumnya. Setelah itu terdapat juga menu masukkan cerita nusantara yang berfungsi untuk menginputkan dokumen cerita rakyat nusantara baru kedalam sistem temu kembali informasi. Kemudian ada menu preprosesing, menu ini berupa menu *drop-down* yang berisi tiga menu yaitu menu hasil *token*, menu hasil *stemming*, dan menu *keyword*. Menu hasil *token* berguna untuk menampilkan seluruh kata yang telah dilakukan proses *token* dan dibuang *stopword*nya. Lalu ada menu hasil *stemming* yang digunakan untuk menampilkan hasil proses *stemming*, menu ini menampilkan kata-kata yang telah dilakukan proses *stemming* sehingga menjadi bentuk kata dasarnya. Setelah itu terdapat menu tampil *keyword* untuk melihat berapa kata kunci yang sesuai dengan jumlah presentase kata yang ingin ditampilkan oleh *admin*, selain itu juga menu ini digunakan untuk melihat bobot dari tiap-tiap kata pada *keyword*. Kemudian ada menu panjang vektor yang berfungsi untuk menampilkan panjang vektor dari tiap dokumen. Setelah itu terdapat menu *cache* yang berfungsi untuk menampilkan nilai kemiripan antara suatu *query* dengan dokumen-dokumen yang terdapat pada sistem temu kembali informasi cerita rakyat nusantara. Lalu yang terakhir ada menu *logout* yang berfungsi untuk keluar dari halaman *admin* dan akan menampilkan halaman pencarian.

4. HASIL PENELITIAN DAN PEMBAHASAN

4.1 Pengujian Sistem Dengan Satu Term

Hasil penelitian yang didapat dengan memasukkan sebuah kata kunci atau *keyword* dengan satu *Term* pada *form* yang tersedia pada sistem temu kembali informasi cerita rakyat nusantara menunjukkan bahwa sistem dapat berjalan dengan baik. Pengujian sistem dengan satu *term* untuk hasil pencarian “sangkanar” didapatkan dokumen yang terambil sebanyak 2 dokumen.

Tabel 1. Hasil uji satu term

Nama Dokumen	Nomor Dokumen	Bobot
Asal Mula Pulau Sangkar	118	0.0629182
Dongeng Nusantara Asal Muasal Pulau Sangkar	117	0.0602468

4.2 Pengujian Sistem Dengan Dua Term

Hasil penelitian yang didapat dengan memasukkan kata kunci atau *keyword* dengan dua *Term* pada *form* yang tersedia pada sistem temu kembali informasi cerita rakyat nusantara menunjukkan bahwa sistem dapat berjalan dengan baik. Pengujian sistem dengan dua *term* untuk hasil pencarian “antu banyu” didapatkan dokumen yang terambil sebanyak 3 dokumen.

Tabel 2. Hasil uji dua term

Nama Dokumen	Nomor Dokumen	Bobot
Mitos Antu banyu	103	0.704718
Legenda Antu Banyu : Hantu Di Sungai Musi Palembang	106	0.559579
Mitos Hantu Banyu di Sungai Musi Palembang	104	0.360116

4.3 Pengujian Sistem Dengan Tiga Term

Hasil penelitian yang didapat dengan memasukkan kata kunci atau *keyword* dengan tiga *term* pada *form* yang tersedia pada sistem temu kembali informasi cerita rakyat nusantara menunjukkan bahwa

sistem dapat berjalan dengan baik. Pengujian sistem dengan tiga *term* untuk hasil pencarian “istana sultan aceh” didapatkan dokumen yang terambil sebanyak 11 dokumen.

Tabel 3. Hasil uji tiga term

Nama Dokumen	Nomor dokumen	Bobot
Asal Usul Tari Guel	1	0.233981
Sejarah asal usul tari guel	5	0.205848
Cerita Rakyat Asal Mula Tari Guel Aceh	3	0.20512
Cerita Rakyat Tari Guel	4	0.20512
Cerita Asal Usul Tari Guel	2	0.171465

4.4 Pengujian Sistem Dengan Lebih Dari Tiga Term

Hasil penelitian yang didapat dengan memasukkan kata kunci atau *keyword* dengan lebih dari tiga *term* pada *form* yang tersedia pada sistem temu kembali informasi cerita rakyat nusantara menunjukkan bahwa sistem dapat berjalan dengan baik. Pengujian sistem dengan lebih dari tiga *term* untuk hasil pencarian “malin kundang anak durhaka” didapatkan dokumen yang terambil sebanyak 95 dokumen.

Tabel 4. Hasil uji lebih dari tiga term

Nama Dokumen	Nomor Dokumen	Bobot
Kisah Legenda Malin Kundang	77	0.771766
Cerita Rakyat Malin Kundang	75	0.763959
Malin Kundang	70	0.763685
Kisah Legenda Indonesia : Malin Kundang	83	0.763685
Cerita Rakyat Nusantara Malin Kundang	80	0.755188

4.5 Pengujian Sistem Dengan Keyword Tidak Relevan

Hasil penelitian yang didapat dengan memasukkan kata kunci atau *keyword* dengan *term* yang tidak relevan atau tidak tersedia pada tabel *tbkeyword* pada *form* pencarian yang tersedia pada sistem temu kembali informasi cerita rakyat nusantara menunjukkan bahwa sistem dapat berjalan dengan baik.

4.6 Pengujian Recall Dan Precision

Pengujian *recall* dan *precision* pada Rancang Bangun Sistem Temu Kembali Informasi cerita rakyat nusantara dilakukan dengan memasukan 1 *term* (tari, batu, sultan) untuk 2 *term* (Malin Kundang, tari guel, anak durhaka, pulau kemarau) selanjutnya 3 *tem* (istana sultan aceh, tari guel aceh, legenda putri hijau) dan lebih dari 3 *term* (malin kundang anak durhaka).

Tabel 5. Recall dan precision

No	Keyword	Terambil	Relevan	Tidak Relevan	Total Doc	Relevan			R	P
						A	B	C		
1	tari	12	11	1	128	11	128	12	0.08	0.91
2	batu	35	24	11	128	24	128	35	0.18	0.68
3	sultan	22	9	13	128	9	128	22	0.07	0.40
4	malin kundang	16	16	0	128	16	128	16	0.12	1.00
5	tari guel	11	11	0	128	11	128	11	0.08	1.00
6	anak durhaka	95	20	75	128	20	128	95	0.15	0.21
7	pulau kemarau	41	7	34	128	7	128	41	0.05	0.17
8	istana sultan aceh	11	10	1	128	10	128	11	0.07	0.90
9	tari guel aceh	33	10	23	128	10	128	33	0.07	0.30
10	legenda putri hijau	67	11	56	128	11	128	67	0.08	0.16
11	malin kundang anak durhaka	97	16	81	128	16	128	97	0.12	0.16

Nilai rata-rata dari recall adalah 0.097 dan nilai rata-rata dari precicion adalah 0.535.

5. KESIMPULAN

5.1 Kesimpulan

Adapun kesimpulan pada penelitian ini adalah sebagai berikut :

- a. Hasil uji pencarian pada sistem temu kembali informasi dengan satu term memiliki bobot tertinggi 0.0629182. Pengujian dengan dua term memiliki bobot tertinggi 0.704718. Pengujian dengan tiga term memiliki bobot tertinggi 0.233981. Hasil pencarian lebih dari tiga term memiliki bobot tertinggi 0.771766.
- b. Hasil dari rata-rata pengujian nilai recall didapat 0.097 dan nilai rata-rata dari precision 0.535 dari perbandingan antara nilai recall dan precision menunjukkan angka recall yang lebih rendah dari precision. Nilai Precision menunjukkan ketepatan hasil pencarian dengan term semakin besar nilai precision dari sebuah sistem temu kembali informasi, maka sistem temu kembali informasi tersebut dapat dikatakan baik. Sedangkan Nilai recall yang tinggi belum dapat menggambarkan bahwa suatu sistem temu kembali informasi itu dikatakan baik atau tidak.
- c. Sistem temu kembali cerita rakyat nusantara dapat menampilkan hasil pencarian dengan memberikan bobot pada tiap dokumen yang terpanggil dengan nilai presisi yang tinggi.

5.2 Saran

Untuk pengembangan lebih lanjut mengenai aplikasi sistem temu kembali informasi ini, diberikan saran sebagai berikut :

- a. Aplikasi ini masih terbatas dengan data cerita rakyat nusantara yang berasal dari pulau sumatera, maka diharapkan untuk pengembangannya dapat dilanjutkan dengan menambah cerita dari daerah-daerah lainnya.
- b. Proses stemming yang digunakan masih belum bisa sepenuhnya mengubah sebuah term menjadi bentuk kata dasarnya dengan benar. Proses ini akan mempengaruhi hasil dari indexing dokumen yang akan mempengaruhi hasil akhir perhitungan.
- c. Apabila pada proses pencarian cerita rakyat nusantara tidak terdapat dokumen yang sesuai atau relevan dengan kata kunci yang diinputkan oleh pengguna maka dibutuhkan sebuah metode yang dapat mengembangkan term pencarian berdasarkan persamaan dari sebuah kata (sinonim).

DAFTAR PUSTAKA

- [1] Agusta, L. (2009). *Perbandingan Algoritma Stemming Porter dengan Algoritma Nazief & Adriani untuk Stemming Dokumen Teks Bahasa Indonesia*. Konferensi Nasional Sistem dan Informatika, 196-201.
- [2] Amin, Fatkhul., (2012). *Sistem Temu Kembali Informasi dengan Metode Vector Space Model*.
- [3] Arief M.Rudyanto, (2012). *Pemrograman Web Dinamis Menggunakan PHP dan MySQL*. Yogyakarta.
- [4] Baeza-Yates R., Neto R., (1999). *Modern Information Retrieval*, Addison Wesley-Pearson international edition, Boston. USA.
- [5] Boubekour, Fatiha., Azzoug, Wassila. (2013). *Concept-Based Indexing In Text Information Retrieval*.
- [6] Febrian, Jack., (2007). *Kamus Komputer & Teknologi Informasi*. Cetakan ke dua, Informatika Bandung, Bandung.
- [7] Hariyanto, Bambang., (2004). *Rekayasa Sistem Berorientasi Objek*, Cetakan Pertama Informatika Bandung, Bandung. Halaman 59, 259, 267.
- [8] Iswanti, K., (2013). *Sistem Informasi Hadits Menggunakan Teknik Temu Kembali Informasi Dengan Model Ruang Vektor*. Pekanbaru.
- [9] Kosasih, E., Hendriyani, L. (2014). *Cerdas Berbahasa dan Bersastra Indonesia untuk SMA/MA kelas XI*. Erlangga.
- [10] MADCOM. (2011). *Aplikasi Web Database dengan Dreamweaver dan PHP-MySQL*, Edisi pertama, Andi, Yogyakarta
- [11] Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze. (2007). *Introduction to information retrieval*. Vol. 1. Cambridge: Cambridge university press.
- [12] Mashadi, I. (2015). *Sistem Temu Kembali Informasi Dengan Metode Vector Space Model (Studi Kasus Biografi Pahlawan Indonesia)* (Doctoral dissertation, UNIVERSITAS STIKUBANK).

- [13] Nazief, Bobby dan Mirna Adriani, 1996, *Confix-Stripping : Approach to Stemming Algorithm for Bahasa Indonesia*, Faculty of Computer Science University of Indonesia.
- [14] Nugroho, Bunafit., (2004). *PHP dan MySQL dengan Editor Dreamweaver MX*, Edisi pertama, Andi, Yogyakarta.
- [15] Pressman, Roger S., (2005). *Software Engineering A Practicioners Approach*.
- [16] Putra, T. K. (2012). *Sistem Temu Kembali Artikel IPTEK* (Doctoral dissertation, UNIVERSITAS STIKUBANK (UNISBANK) SEMARANG).
- [17] Raharja, D. T. Y. (2015). *Sistem Temu Kembali Informasi Dengan Metode Vector Space Model (Studi Kasus Abstraksi Skripsi Fti Unisbank)* (Doctoral dissertation, UNIVERSITAS STIKUBANK).
- [18] Salton, G., 1989, *Automatic Text Processing, The Transformation Analysis, and Retrieval of information by computer*. Addison – Wesley Publishing Company, Inc. USA.
- [19] Sutanta, E., (2005). *Pengantar Teknologi Informasi*. Cetakan Pertama, Graha Ilmu, Yogyakarta
- [20] Sutarman., (2007). *Membangun Aplikasi Web Dengan PHP & MySQL*. Edisi ke 2, Cetakan Pertama, Graha Ilmu, Yogyakarta.
- [21] Tala, F.Z., 2003, *A Study of Stemming Effects on Information Retrieval in bahasa Indonesia*. Master Thesis, Institut for logic, Language and Computation Universiteit van Amsterdam The Netherlands.
- [22] Tim Edukatif. (2007). *Kompeten Berbahasa Indonesia Jilid 1 untuk SMA/MA Kelas X*. Erlangga.
- [23] ZA, CHOIRUN NAFIES., 2013. *Sistem Temu Kembali Informasi Teks Berbahasa Indonesia Berobjek Resep Masakan*. Information Technology and Telematics, 1(1).s