

REKAYASA SISTEM TEMU KEMBALI INFORMASI DOKUMEN TEKS BERBAHASA JAWA METODE *COSINE SIMILARITY* DAN *RULE BASE STEMMING* BAHASA JAWA

Fatkul Amin¹, Eddy Nurraharjo²

^{1,2}Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Stikubank
e-mail: ¹fatkhulamin@edu.unisbank.ac.id, ²eddynurraharjo@edu.unisbank.ac.id

ABSTRAK

Saat ini penggunaan Bahasa Jawa di Indonesia yang mulai ditinggalkan. Perlunya pelestarian bahasa Jawa dalam bentuk online yang bisa diakses bagi penggunanya sehingga akan mempermudah dalam pencarian dokumen teks khususnya dokumen bahasa Jawa. Sistem Temu Kembali Informasi (STKI) yang ada saat ini memberikan hasil pencarian dokumen dengan hasil perolehan dokumen dalam jumlah banyak (*recall* tinggi) dan akurasi yang rendah (*precision* rendah). Tujuan pembuatan STKI menggunakan metode *Cosine Similarity* dan *Rule Base Stemming* Bahasa Jawa agar user mudah dalam melakukan pencarian dokumen teks berbahasa Jawa. Software STKI dirancang untuk memberikan hasil pencarian dokumen yang memiliki tingkat presisi tinggi, sehingga user akan mendapatkan hasil pencarian cepat dan akurat. Evaluasi hasil pencarian STKI dilakukan dengan uji *recall* dan *precision*. Studi kasus yang telah dilakukan menggunakan STKI ini didapatkan hasil sistem mampu melakukan proses preprosesing (*tokenisasi*, *filtering*, dan *stemming*) dan perhitungan *Cosine Similarity* dengan hasil mampu melakukan pencarian dokumen teks dan menampilkan hasil pencarian dokumen teks berbahasa Jawa dengan disertai bobot tiap dokumen beserta letak dokumen. Uji *recall* dan *precision* menggunakan analisa persepsi dan STKI menunjukkan hasil pencarian dokumen teks memiliki rata-rata *recall* = 0,02 dan rata-rata *precision* = 0,94. STKI yang dibangun memiliki keunggulan mampu melakukan pencarian dokumen teks bahasa Jawa dan hasil pencarian yang akurat (*precision* = 0,94), serta dilengkapi dengan bobot dan letak dokumen pada database

Kata Kunci: *Rule Base Stemmer, Bahasa Jawa, Cosine Similarity*

1. PENDAHULUAN

Pemanfaatan Mesin pencari atau Sistem Temu Kembali Informasi (STKI) saat ini sudah menjadi kebutuhan masyarakat pada umumnya. Beberapa mesin pencari menghasilkan hasil pencarian yang berbeda-beda karena menggunakan metode pencarian yang berbeda. Mesin pencari atau sistem temu kembali informasi dibuat untuk mendekatkan keinginan pengguna (*user*) berupa informasi dengan hasil pencarian yang ditampilkan bisa sesuai dengan keinginan pengguna. Sistem temu kembali informasi akan memberikan nilai tambah dalam pencarian informasi jika keinginan *user* bisa terpenuhi. Implementasi mesin pencari sudah banyak dilakukan dan mudah ditemukan dalam dunia daring, namun sedikit yang banyak menggunakan bahasa daerah khususnya bahasa Jawa. Penggunaan bahasa Jawa dalam kehidupan sehari-hari terus berkurang dan terus terpinggirkan. Mesin pencari Jawa yang sudah ada dan mudah ditemukan di dunia daring juga sedikit sekali digunakan.

Generasi muda mulai dari anak SD, SMP, dan SMA pelan tapi pasti mulai meninggalkan bahasa Jawa dan menggantinya dengan bahasa Indonesia. Bahasa Jawa sebagai bahasa yang paling banyak digunakan di wilayah Indonesia setelah bahasa Indonesia, dewasa ini mulai banyak ditinggalkan oleh kebanyakan orang. Media offline dan media online juga kurang mengangkat bahasa Jawa sehingga dikhawatirkan bahasa Jawa lama-kelamaan akan ditinggalkan oleh bangsa kita. Beberapa media online berbahasa Jawa ada, namun belum menggunakan atau belum menyediakan pencarian informasi menggunakan mesin pencari khusus berbahasa Jawa. Pencarian informasi saat ini dilakukan dengan menggunakan mesin pencari atau sistem temu kembali informasi, *user* menuliskan *query* dan mesin pencari akan menampilkan hasil pencarian. Mesin pencari yang sudah ada dan banyak digunakan saat ini memberikan hasil perolehan pencarian yang banyak (banyak dokumen yang terambil), sehingga diperlukan waktu untuk menentukan hasil pencarian yang relevan. Menentukan hasil yang relevan sesuai dengan keinginan *user* dengan jumlah hasil pencarian yang banyak akan menyulitkan *user*. Hal ini terjadi karena dokumen yang terambil oleh sistem jumlahnya banyak, maka sistem berkemungkinan menampilkan hasil pencarian yang tidak relevan. Banyaknya dokumen hasil pencarian ini membuat waktu yang dibutuhkan dalam pencarian menjadi lebih banyak dari yang diharapkan. Perkembangan penelusuran informasi saat ini menghasilkan *recall* yang tinggi dan *precision* yang rendah. *Recall* yang tinggi diartikan bahwa dokumen yang dihasilkan dalam penelusuran dokumen adalah banyak, sedangkan *precision* rendah dapat diartikan bahwa dokumen yang diharapkan dapat ditemukan sedikit.

Solusi untuk mengatasi masalah ini adalah dengan membuat *software Information Retrieval System (IRS)* menggunakan metode *Cosine Similarity*. Metode *Cosine Similarity* dipilih karena cara kerja model ini efisien, mudah dalam representasi dan dapat diimplementasikan pada *document-matching*. *Software IRS* diharapkan menghasilkan pencarian dokumen dengan hasil pencarian sedikit (*recall* rendah) dan data hasil pencarian informasi akurat (*precision* tinggi).

2. TINJAUAN PUSTAKA

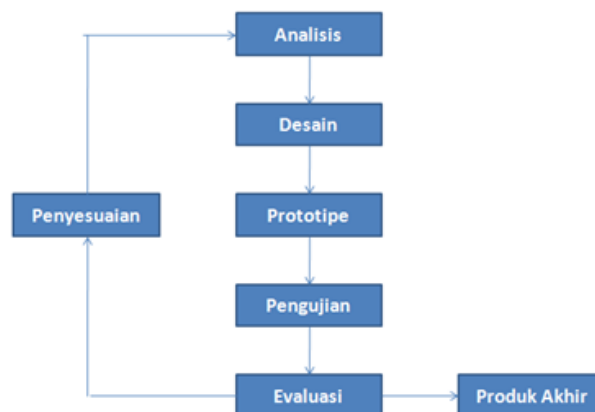
Penelitian terkait dengan menggunakan metode *Cosine Similarity* dilakukan antara lain pada bidang *LinguisticR*. Umamaheswari [10]. Clustering adalah teknik yang berguna yang menyelenggarakan jumlah besar dokumen teks menjadi sejumlah kecil kelompok yang bermakna dan koheren, sehingga memberikan dasar untuk intuitif dan informatif navigasi dan browsing mekanisme. Teks-clustering untuk membagi koleksi text dokumen ke dalam kategori yang berbeda sehingga dokumen dalam kategori yang sama menggambarkan samatopik seperti musik klasik. teks Clustering efisien elompok dokumen dengan isi yang serupa dalam cluster yang sama. Kesamaan antara objek adalah diukur dalam penggunaan fungsi kesamaan. Itu skema pengelompokan hirarki dapat secara efektif digunakan untuk memproses dataset besar.

Penelitian tentang Aplikasi Pendeteksi Kemiripan Pada Dokumen Teks Menggunakan Algoritma Nazief & Adriani Dan Metode *Cosine Similarity* dilakukan oleh Azhar Firdaus [2]. Aplikasi pendeteksi kemiripan pada dokumen teks menggunakan algoritma Nazief & Adriani dan metode *Cosine Similarity* dapat membandingkan berkas berbeda ekstensi dan membandingkan lebih dari dua dokumen secara bersamaan. Salah satu cara yang dapat digunakan untuk mengoptimasikan nilai kemiripan yang diperoleh dari perhitungan metode *Cosine Similarity* adalah dengan menerapkan algoritma Nazief & Adriani sebelum proses perhitungan. Berdasarkan hasil yang didapat dari uji kelayakan aplikasi dapat disimpulkan bahwa aplikasi pendeteksi kemiripan pada dokumen teks menggunakan algoritma Nazief & Adriani dan Metode *Cosine Similarity* tergolong ke dalam.

Penelitian tentang cosine juga dilakukan oleh Sugiyanto [9], dengan topik Analisa Performa metode *Cosine* dan *Jaccard* pada pengujian kesamaan dokumen. Teknologi informasi memudahkan distribusi data-data digital melalui berbagai media, salah satunya adalah dokumen. Namun hal ini menyebabkan adanya penyalahgunaan dalam bentuk duplikasi yang mengarah kepada kegiatan plagiarisme terutama untuk naskah-naskah akademik seperti skripsi atau tugas akhir. Berbagai metode dikembangkan untuk meminimumkan terjadinya duplikasi ilegal. Performa metode *Cosine* dan *Jaccard* menguji tingkat kemiripan dokumen dalam bentuk abstrak. Hasil pengujian menunjukkan bahwa pengujian kemiripan menggunakan *Cosine* memiliki tingkat akurasi lebih tinggi dibanding dengan *Jaccard*.

3. METODE PENELITIAN

Penelitian ini menggunakan model *prototype*. Di dalam model ini sistem dirancang dan dibangun secara bertahap dan untuk setiap tahap pengembangan dilakukan percobaan-percobaan untuk melihat apakah sistem sudah bekerja sesuai dengan yang diinginkan. Sistematika model *prototype* terdapat pada Gambar 1 memperlihatkan tahapan pada *prototype*.



Gambar 4.1. Tahapan *Prototype* (Pressman, 2001)

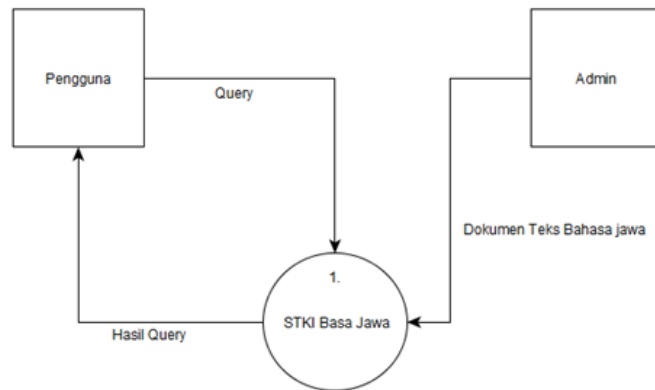
Pada tahap analisa dilakukan analisa tentang masalah penelitian dan menentukan pemecahan masalah yang tepat untuk menyelesaikannya. Menentukan tujuan pembuatan mesin pencari. Pada tahap desain dibangun rancangan Sistem Temu Kembali Informasi bahasa Jawa (DFD dan Flow Chart). Pada tahap *Prototype* dibangun Sistem Temu Kembali Informasi Bahasa Jawa. Tahap ini di mulai dari proses tokenisasi, Penyaringan (filtering), Pembuatan kata dasar bahasa Jawa (stemming), tfidf, dan perhitungan *Cosine Similarity* yang diaplikasikan dengan program PHP. Pada tahap Pengujian dilakukan pengujian *Recal* dan *Precision*. Pada tahap Evaluasi dilakukan evaluasi apakah performa aplikasi sudah sesuai dengan yang diharapkan, apabila belum maka dilakukan penyesuaian-penyesuaian secukupnya. Tahap Penyesuaian dilakukan apabila pada evaluasi performa aplikasi kurang memadai dan dibutuhkan perbaikan, tahap ini melakukan penyesuaian dan perbaikan pada aplikasi sesuai dengan kebutuhan

4. HASIL DAN PEMBAHASAN

4.1. Diagram Kontek

Pada STKI Jawa ini Diagram kontek Sistem Temu Kembali Informasi menggambarkan arus data yang terjadi yang dimulai dari *user* yang menginputkan *query* kemudian user akan mendapatkan informasi dari sitem

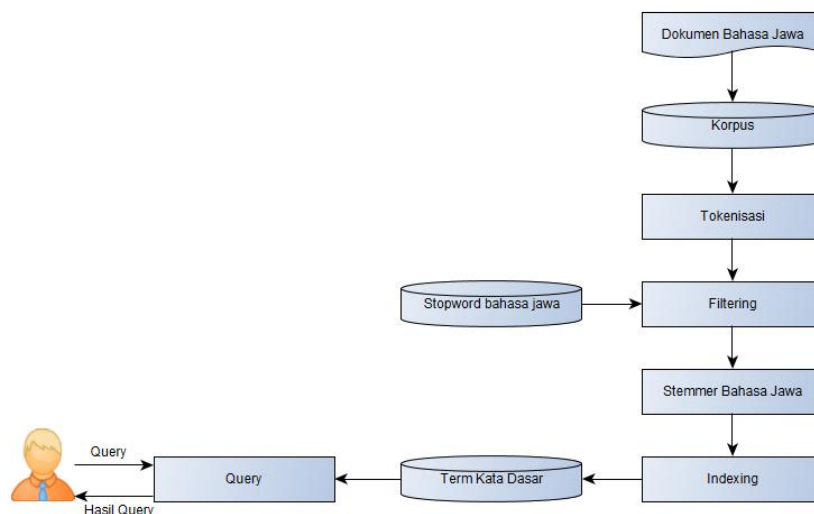
berupa hasil pencarian. Administrasi bisa memasukkan data (*input data*) dan edit dokumen teks bahasa Jawa kedalam STKI, dalam hal ini input dokumen teks ke dalam korpus. Selengkapnya diagram kontek STKI bisa dilihat pada gambar 2.



Gambar 2. Diagram Kontek Sistem Temu Kembali Informasi

4.2. Arsitektur Sistem Temu Kembali Informasi

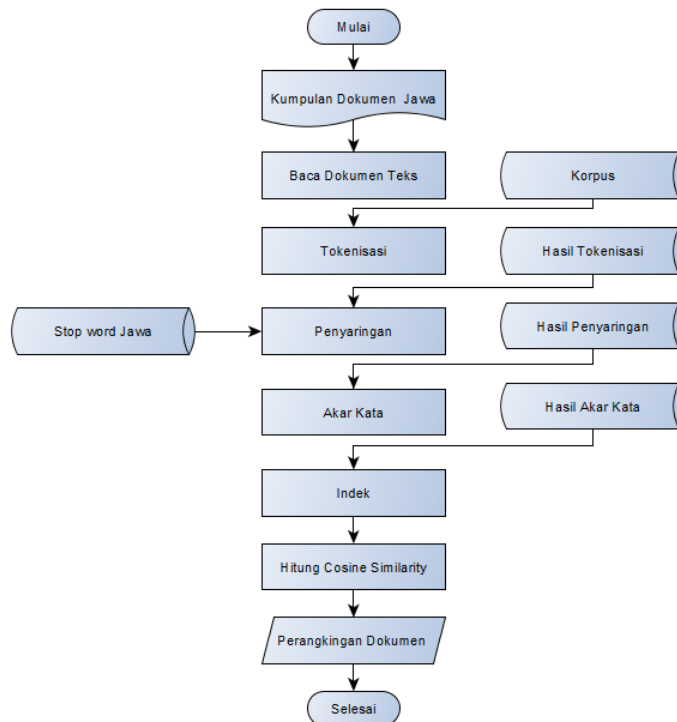
Pada tahap ini dibangun rancangan sistem dengan menggunakan *tools* pengembangan sistem informasi yaitu *flowchart* (*flowchart* sistem temu kembali informasi, terdiri dari *Flowchart* sistem untuk pembentukan dokumen, *Flowchart* proses tokenisasi, *Flowchart* proses pembuangan *stopword*, *Flowchart* proses *stemming*, *Flowchart* proses *indexing*, dan *Flowchart* proses hitung *Cosine*). Arsitektur STKI bisa dilihat pada gambar 3.



Gambar 3. Arsitektur Sistem Temu Kembali Informasi

4.3. Flowchart STKI Jawa

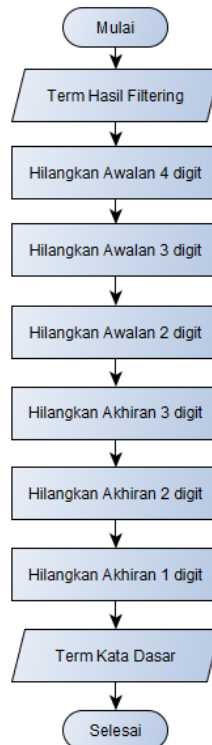
Flowchart STKI dibuat untuk memudahkan pengguna dan menampilkan model pencarian yang umum digunakan saat ini. STKI didesain untuk menemukan informasi secara akurat kepada pengguna (*user*), amin [1]. Proses STKI oleh sistem melalui proses-proses seperti gambar 4. *Flowchart* diawali dengan *input* dokumen-dokumen kedalam korpus. Selanjutnya dokumen melalui proses preprosesing, dihitung bobotnya dan dibuat rankingnya berdasarkan bobot dokumen yang tertinggi. Hasil STKI adalah dokumen yang relevan dengan permintaan *user*.



Gambar 4. flowchartSTKI Jawa

4.4. Flowchart Proses Stemmer Jawa

Stemming dibuat untuk term hasil filtering diubah menjadi term kata dasar. Proses stemmer bahasa jawa dimulai dengan menghilangkan awalan dan akhiran. Proses ini juga dirancang dapat melakukan pergantian (*replace*) ketika awalan dihilangkan dan menggantinya dengan huruf yang sesuai. Proses menghilangkan awalan, akhiran, dan *replace* sisipan dilakukan dalam satu tahap proses. Gambar 5 menunjukkan flowchart stemming.



Gambar 5. Flowchart Proses Stemming

4.5. Perhitungan STKI

Cosine Similarity atau $Sim(q,d_j)$ digunakan untuk mengevaluasi tingkat similaritas atau kemiripan antar dokumen (d_j) berkaitan dengan query (q) sebagai korelasi antar vektor d_j dan q . Korelasi ini bisa diukur dengan persamaan (1)

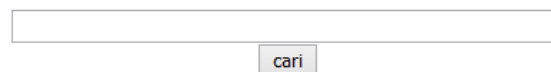
$$\begin{aligned} \text{Sim}(q, d_j) &= \frac{q \cdot d_j}{|q| * |d_j|} \\ &= \frac{\sum_{i=1}^t W_{iq} \cdot W_{ij}}{\sqrt{\sum_{j=1}^t (W_{iq})^2} * \sqrt{\sum_{i=1}^t (W_{ij})^2}} \end{aligned} \quad (1)$$

Dengan $\text{Sim}(q,d_j)$ adalah Similaritas antara *query* dan dokumen, q adalah Bobot *query*, d_j adalah Bobot dokumen, $|q|$ adalah Jarak *query*, dan $|d_j|$ adalah Jarak dokumen.

4.6. Aplikasi Tampilan STKI Jawa

STKI Jawa dibuat dengan model mesin pencari pada umumnya. Pada Aplikasi *interface* ini ditampilkan kolom *query* yang bisa digunakan untuk memasukkan *query* oleh pengguna. Kotak *button* dengan label *go* digunakan untuk memproses setelah *query* di *input*. Tombol *button* cari jika sudah diklik akan menampilkan abstraksi hasil pencarian. Gambar rancangan *Interface* dapat dilihat pada gambar 6.

STKI Basa Jawa



Gambar 6. Tampilan (*Interface*)STKI Bahasa Jawa

4.7. Aplikasi Hasil Pencarian STKI Bahasa Jawa

Aplikasi menu Hasil Pencarian ini akan tampil setelah STKI Jawa melakukan pencarian. Sistem akan menampilkan sejumlah halaman yang relevan dengan *query* setelah melalui proses perankingan. Tata letak hasil pencarian menggunakan bobot yang dimiliki oleh tiap dokumen berdasarkan proses perhitungan *Cosine Similarity*. Hasil pencarian dengan bobot terbesar akan diletakkan pada posisi pertama diikuti oleh bobot yang lebih kecil (*descending*). Rancangan *Interface* hasil pencarian akan menampilkan informasi-informasi tentang dokumen-dokumen yang relevan dengan *query* yang diinput oleh user. *Interface* hasil pencarian akan menampilkan informasi berupa dokumen-dokumen yang relevan dengan hasil pencarian, dilengkapi dengan nilai similaritas dokumen dan posisi dokumen tersebut. Mesin pencari *Cosine Similarity method* (STKI) bekerja melalui beberapa tahapan proses sebagai berikut; Proses Tokenisasi dilakukan dengan mekanisme jika dokumen pada korpus ditemukan spasi, maka *term* yang ada diantara spasi akan di *retrieved* (akan diambil oleh sistem) kemudian *term* ditempatkan dalam tabel tabelawal. Hasil proses berupa *term* asli (*term* yang masih memiliki imbuhan, tanda baca yang melekat, dan angka).

Proses penyaringan (*Filtering*) dilakukan dengan mekanisme jika *term* pada tabel tabelawal ditemukan tanda baca, huruf kapital, atau angka. Maka program akan menghilangkan (tanda baca dan angka) dan mengganti (huruf kapital menjadi huruf kecil), kemudian memeriksa *term* dengan *stopwords* (*stopword* bahasa Jawa). Hasil proses berupa *term* pilihan (tanpa tanda baca, tanpa huruf kapital, dan bukan termasuk *stopwords*). Keunggulan proses ini sistem mampu mereduksi tanda baca, angka, merubah *term* menjadi huruf kecil, dan memeriksa *term* *stopwords* dengan waktu komputasi yang cepat. Proses *Stemming* dilakukan program dengan cara menghilangkan imbuhan yang terdapat pada *term* hasil *filtering*. Proses menghilangkan dilakukan dengan menghilangkan awalan, sisipan, dan akhiran. Hasil proses ini dimasukkan dalam tabel tabelfreq. Hasil proses berupa *term* kata dasar.

Proses Pembobotan dokumen dengan metode *Cosine Similarity method* dilakukan dalam proses pencarian dokumen. Program akan bekerja ketika user melakukan *query*, selanjutnya program akan memproses *query* tersebut dengan perhitungan-perhitungan *tf*, *idf*, *tfidf*, jarak *query* dan dokumen, similaritas dan *cosine similarity*. Hasil proses pembobotan dengan metode *Cosine Similarity method* berupa dokumen hasil pencarian disertai dengan bobot dokumen, letak dokumen dan disusun *descending*. Selengkapnya proses atau cara kerja STKI bahasa Jawa mulai dari persiapan dokumen hingga pencarian dokumen beserta hasilnya akan dibahas pada proses STKI.

4.8. Studi Kasus menggunakan Keyword Bahasa Jawa

Studi kasus pada aplikasi STKI Bahasa Jawa ini menggunakan dokumen teks Bahasa Jawa pada Majalah Penjebar Semangad, Majalah Jayabaya dan majalah Jaka Lodang. *Query* yang dimasukkan pada STKI bahasa Jawa adalah *keyword* dengan 1 *term* yaitu "Raja" 2 *term* yaitu "Garwa Raja", "Anak Raja" 3 *term* "Sunan Muria Kudus". "Miturut sejarah budha" 4 *term* "Guru mangan tangi turu", "Ora wani karo anak". 5 *term* "Sesuk isuk rung mesti ono", "bali mrene mangkat mengko sore", dan 7 *term* "kabeh wong sesuk isuk rung mesthi ono".

STKI Basa Jawa

Anak Raja

Dicari : anak raja
Jml Kata : 2

Judul : HARJUNA KAWTWAHA

Similaritas : 0.835
Dokumen : JL022009PD

sang harjuna nuli nithi kreta per-rang, angiuwa langkap si gand-hiwa sarta anyebul sangkakala dewardatta, sesumbar nantang yuda njalukdilawan prabu niwatakawaca. sakala geger para wadya ian prawira negara hima-himantaka, dhasarwis padha siaga ing pupuh, mula age pad-ha mbudidaya ngroyok satriya sing dianggep neneka ian murangtata. katgadeng-tyas sri narendra yaksa niwatakawaca mireng sangkakala panantang kang swarane kaya pan-gempretng gajah sakethi angebaki akasa. mula srinata age nilar bathari supriba, miyos menyang alun-alun methukake sapa sing kumawani nantang, kacarita nalika semana kreta-perange sang harjuna mobat-mabit nRajang bebarisan para buta sarwi ngubal jemparing kang maewu-ewu mbrubul kaya sulung ngebaki langit. giris gusis tapis para wadya raseksa. nadyan para manggala ian prawira sisane kang padha mati, age padha ngungsi marang samburine prabu niwatakawaca. "hhrrrr e e e e! hong tete hyang kadalodra maspatik Raja dewa-ku! satriyaaaa, sapa aranmu dene ku-mawani soroh amuk ana ing susuhe singabarong, hei? hhr!!!" "mara waspadakna kang cetha ian rungokna kanthi wening, heh yaksa kang nedya lumawan kodrat! ya insun iki kangsesilih raden har-juna kangsinaraya ing dewa kanggo nyimakake titah angkara kang dadi memalaningjagad! mara age timban-gana kridhaku yen pancen kowe wis jeleh urip!!!" "kelakon ana satriya mulih kari aran, keparat!!!" paraga lelorone nuli padha nem-puh perang sarwi madeg suraning ndriya. sakah gelar ian keprikelan angunggar yuda ginelar dening sang yaksendra, nanging kabeh bisa katandhingan dening sang harjuna. esthining penggalih sang harjuna nalika semana, ora arep pamer kasudiran, nanging mung mestuti dhawuhe sanghyang bathara indra, yen paprangan ing hima-himantaka mono dadia pendadaran, amarga sa-liyane sekti mandraguna apengawak sak-prabata-suta, sang niwatakawaca pinunjul ing bab gelaring perang dharat ian sandhuwuring kreta ing sa-jagad datanpa tandhing. yen harjuna bisa nandhingi niwatakawaca, bakal gedhe undhuh-undhuhane ing tembe, lamun kaleksanan dumadi perang agung bharatayuda jayabinangun. kedadak kretane sang niwatakawaca ambruk ambyar remuk kena

Gambar 7. Hasil Pencarian keyword "Anak Raja"

Hasil pencarian dokumen dengan keyword "Anak Raja"(gambar 7), menunjukkan dokumen dengan bobot tertinggi adalah dokumen letak dokumen JL022009PD (bobot 0,835). Dokumen JL022009PD (dokumen Teks Majalah Jaka Lodang bulan Februari tahun 2009 kategori Padalangan) memiliki bobot tertinggi atau memiliki tingkat kemiripan tertinggi dibandingkan dengan dokumen lain yang ada pada korpus.

4.9. Deklarasi Pengujian keyword dengan Analisa Persepsi

Pendeklarasian penting dilakukan untuk proses perhitungan uji recall dan precision. Semua keyword yang akan di uji harus dideklarasikan dulu makna atau arti dari keywordnya. Hal ini penting agar proses pengkategorian keyword relevan atau tidak relevan menjadi benar dan ada dasarnya. Deklarasi persepsi bisa dilihat pada tabel 1.

Tabel 1. Tabel Persepsi

Jawa	Indonesia	Arti menurut KKBI
Anak	Anak	anak n 1 keturunan yang kedua: ini bukan -- nya, melainkan cucunya; 2 manusia yang masih kecil:
Raja	Raja	raja/ra-ja/ n 1 penguasa tertinggi pada suatu kerajaan (biasanya diperoleh sebagai warisan); orang yang mengepalai dan memerintah suatu bangsa atau negara: negara kerajaan diperintah oleh seorang --; 2 kepala daerah istimewa; kepala suku; sultan; 3 sebutan untuk penguasa tertinggi dari suatu kerajaan; 4 orang yang besar kekuasaannya (pengaruhnya) dalam suatu lingkungan (perusahaan): -- minyak; 5 orang yang mempunyai keistimewaan khusus (seperti sifat, kepandaian, kelicikan)
Garwa	Istri	1 wanita (perempuan) yang telah menikah atau yang bersuami; 2 wanita yang dinikahi: almarhum meninggalkan seorang -- dan dua orang anak; -- gelap perempuan simpanan; perempuan piaraan (yang tidak dinikahi);
Sunan	Sunan	1 sebutan raja untuk keraton Surakarta (di Jawa); 2 penyebutan nama untuk para wali: -- Kalijaga; kesunanan/ke-su-nan-an/ n daerah sunan
Muria	Muria	Nama wisata di Muria
Kudus	Kudus	suci; murni; nama kota
Guru	Guru	orang yang pekerjaannya (mata pencahariannya, profesinya) mengajar;-- kencing berdiri, murid kencing berlari, pb kelakuan murid (orang bawahan) selalu mencontoh guru (orang atasannya);
Mangan	Makan	memasukkan makanan pokok ke dalam mulut serta mengunyah dan menelannya: mereka -- tiga kali sehari; 2 v memasukkan sesuatu ke dalam mulut, kemudian mengunyah dan menelannya: ia sedang -- pisang; 3 v memasukkan sesuatu ke dalam mulut dan mengunyah-ngunyahnya: Nenek sedang -- siri; 4 v memasukkan sesuatu ke dalam mulut dan menelannya: pasien harus -- pil;
Tangi	Bangun	1 bangkit; berdiri (dari duduk, tidur, dan sebagainya): anak itu berkali-kali terjatuh, namun ia selalu dapat -- kembali; 2 jaga (dari tidur): setiap pagi ia -- pukul 04.00; 3 belum (tidak) tidur; jaga: sewaktu suaminya pulang larut malam, ia masih --; 4 siuman dari pingsan; mendusin: ia -- setelah kepalanya diguyur air; kalau tidak mendengar ledakan itu, saya tidak --; 5 mulai sadar (insaf) akan nasibnya: bangsa terjajah mulai -- menuntut kemerdekaannya; kaum buruh mulai -- menuntut perbaikan nasib; 6 mulai memuai (tentang adonan): dengan ditutup rapat-rapat adonan itu akan cepat --; 7 mulai menjadi cair (tentang minyak kelapa yang beku):

4.10. Perhitungan Recall dan Precision

Pengujian recall(P) dan precision (R) dilakukan dengan cara inputquery ke dalam STKI input 1 term, 2 term dan 3 term, 4 term, dan 5 term dan 7 term. Perhitungan recall dan precision menggunakan persamaan (8) dan persamaan (9). Hasil pengujian recall dan precision dengan menguji 1 term, 2 term dan 3 term sampai dengan 7 term menunjukkan bahwa jika recall rendah maka precision akan tinggi, selengkapnya terlihat pada tabel 2.

Hasil perhitungan Recall untuk keyword "Anak Raja" adalah sebagai berikut;

$$\text{Recall (R)} = \frac{\text{Number of relevant items retrieved}}{\text{Total number of relevant items in collection}} \tag{2}$$

$$\text{Recall} = \frac{15}{340} = 0.04$$

Hasil perhitungan Precision untuk keyword “seneng ngalah” adalah sebagai berikut;

$$\text{Precision (P)} = \frac{\text{Number of relevant items retrieved}}{\text{Total number of items retrieved}} \tag{3}$$

$$\text{Precision} = \frac{15}{19} = 0.79$$

Hasil perhitungan rata-rata untuk Recall dan precision adalah sebagai berikut;

$$\text{Rata - rata Recall} = \frac{0.19}{10} = 0.02$$

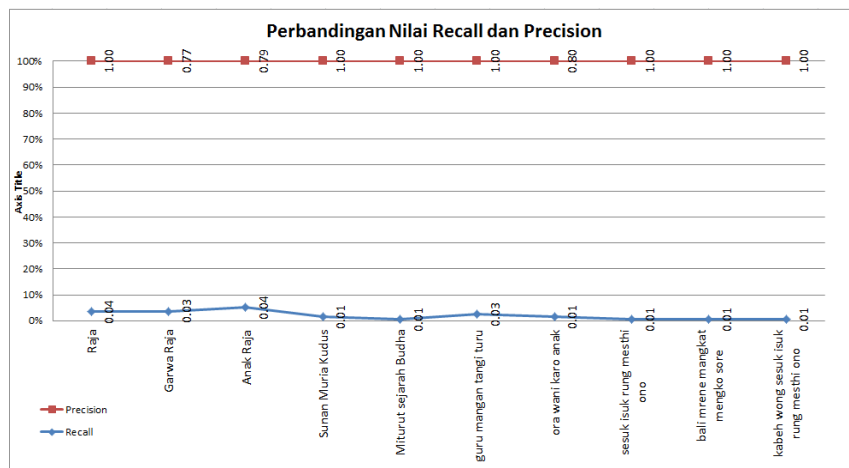
$$\text{Rata - rata Precision} = \frac{9.36}{10} = 0.94$$

Tabel 2. Hasil Pengujian Recall dan Precision

No	Query	Recall	Precision
1	Raja	0.04	1.00
2	Garwa Raja	0.03	0.77
3	Anak Raja	0.04	0.79
4	Sunan Muria Kudus	0.01	1.00
5	Miturut sejarah Budha	0.01	1.00
6	guru mangan tangi turu	0.03	1.00
7	Ora wani karo anak	0.01	0.08
8	Sesuk isuk rung mesthi ono	0.01	1.00
9	Bali mrene mangkat mengko sore	0.01	1.00
10	Kabeh wong sesuk isuk rung mesthi ono	0.01	1.00

4.11. Diagram Hasil pengujian Recall dan Precision

Hasil pengujian recall dan precision pada keyword 1 suku kata, 2 suku kata, 3 suku kata, 4 suku kata, 5 suku kata dan 7 suku kata bisa dilihat pada gambar 8.



Gambar 8. Diagram Hasil perhitungan Recall dan Precision

5. KESIMPULAN

- Sistem Temu Kembali Informasi bahasa Jawa yang dibuat memiliki tingkat keakuratan tinggi sehingga efektif dalam pencarian informasi. Berdasarkan uji recall dan precision menggunakan analisa persepsi, Hasil Uji recall dan precision STKI Jawa Metode Cosine Similarity dan Rule Base Stemmer menunjukkan hasil pencarian dokumen teks Jawa memiliki rata-rata recall = 0,02 dan rata-rata precision = 0,94.
- STKI Jawa yang dibangun memiliki keunggulan mampu melakukan pencarian dokumen teks bahasa Jawa dan hasil pencarian yang akurat (precision = 0,94), serta dilengkapi dengan bobot dan letak dokumen pada database.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Direktorat Riset Pengabdian Kepada Masyarakat Direktorat Jenderal Penguatan Riset dan Pengembangan Kementerian Riset, Teknologi, dan Pendidikan Tinggi yang telah memberi “dukungan financial” terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] Amin, Fatkhul. 2015. Rancang Bangun Information Retrieval System (IRS) Bahasa Jawa ngoko pada Palintangan Penjebar Semangad dengan Metode Vector Space Model (VSM) Dinamik Jurnal Teknologi Informasi, UNISBANK Semarang, Volume 20, Nomor 1, , ISSN : 0854 - 9524, Halaman: 25 – 35
- [2] Firdaus, Azhar Dkk. 2014. Aplikasi Pendeteksi Kemiripan Pada Dokumen Teks Menggunakan Algoritma Nazief & Adriani Dan Metode Cosine Similarity. Jurnal Teknologi Informasi, Volume 10 Nomor 1, ISSN 1414-9999
- [3] Haryono, M.EA., Wahyudi., 2005. *Customer Information Gathering* menggunakan Metode Temu Kembali Informasi dengan Model Ruang Vektor. Seminar Nasional Aplikasi Teknologi Informasi, SNATI 2005.
- [4] Kadir, A., 2001. Dasar Pemrograman Web Dinamis menggunakan PHP. Penerbit Andi. Yogyakarta.
- [5] Manning, C., Raghavan, P., 2007. *An Introduction to Information Retrieval*, Stanford. USA.
- [6] Meadow, C.T., 1997. *Text Information Retrieval Systems*. Academic Press. New York.
- [7] Tala, F.Z., 2003, *A Study of Stemming Effects on Information Retrieval in bahasa Indonesia*. Institut for logic, Language and Computation Universiteit van Amsterdam The Netherlands.
- [8] Salton, G., 1989, *Automatic Text Processing, The Transformation, Analysis, and Retrieval of information by computer*. Addison – Wesley Publishing Company, Inc. USA.
- [9] Sugiyamto Dkk., 2014. Analisa Performa Metode Cosine dan Jacard pada Pengujian Kesamaan Dokumen. Jurnal Masyarakat Informatika, Volume 5, Nomor 10, ISSN 2086–4930. Universitas Diponegoro
- [10] Umamaheswari, K Dkk. 2014. Text Clustering Using Cosine Similarity and Matrix Factorization. International Journal of Research in Computer and Communication Technology, Vol 3, Issue 10. ISSN 2320- 5156
- [11] Yates, R.B, 1999. *Modern Information Retrieval*, Addison Wesley-Pearson international edition, Boston. USA.