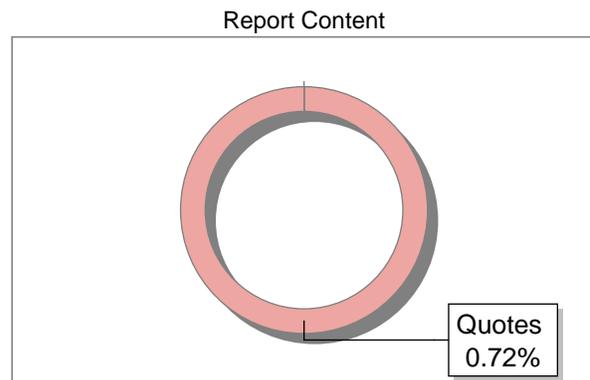
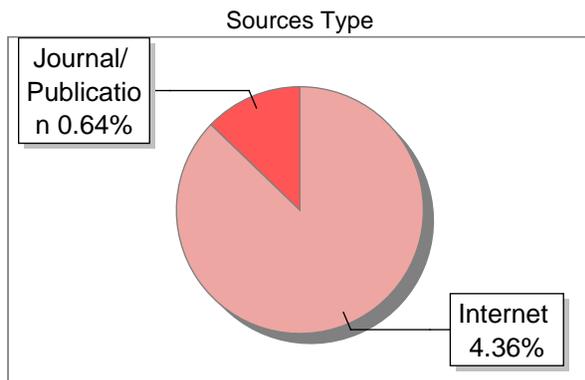


Submission Information

Author Name	Jeffri Alfa Razaq
Title	Pemeringkatan Hasil Pencarian Dokumen Teks Pada Sistem Temu Kembali Informasi Berbahasa Jawa Menggunakan Metode Dice Similarity
Paper/Submission ID	2247613
Submitted by	atikrahma@edu.unisbank.ac.id
Submission Date	2024-08-20 23:08:28
Total Pages, Total Words	10, 2647
Document type	Article

Result Information

Similarity **5 %**



Exclude Information

Quotes	Excluded
References/Bibliography	Excluded
Source: Excluded < 14 Words	Not Excluded
Excluded Source	92 %
Excluded Phrases	Not Excluded

Database Selection

Language	Non-English
Student Papers	No
Journals & publishers	No
Internet or Web	Yes
Institution Repository	No

A Unique QR Code use to View/Download/Share Pdf File





DrillBit Similarity Report

5

SIMILARITY %

8

MATCHED SOURCES

A

GRADE

A-Satisfactory (0-10%)

B-Upgrade (11-40%)

C-Poor (41-60%)

D-Unacceptable (61-100%)

LOCATION	MATCHED DOMAIN	%	SOURCE TYPE
4	docplayer.info	1	Internet Data
5	www.freepatentsonline.com	1	Internet Data
6	jurnal.ugm.ac.id	1	Publication
7	docplayer.info	1	Internet Data
8	ub.ac.id	1	Internet Data
10	adoc.pub	<1	Internet Data
11	docplayer.info	<1	Internet Data
12	repository.unair.ac.id	<1	Internet Data

EXCLUDED SOURCES

1	amikjtc.com	92	Publication
---	-------------	----	-------------

PEMERINGKATAN HASIL Pencarian Dokumen Teks Pada Sistem Temu Kembali Informasi Berbahasa Jawa Menggunakan Metode DICE Similarity

Fatkhul Amin, Jeffri Alfa Razaq

Universitas Stikubank Semarang

fatkhulamin@edu.unisbank.ac.id, mrjf@edu.unisbank.ac.id

ABSTRAK

Sistem temu kembali informasi dokumen teks berbahasa Jawa sukar ditemukan baik dari dunia daring maupun luring. Pencarian dokumen teks bahasa Jawa dengan menggunakan sistem temu kembali informasi yang tersedia didapatkan suatu hasil pencarian dengan dokumen terambil yang banyak (recall tinggi) sehingga hasil menjadi kurang akurat (precision rendah). pemeringkatan sistem temu kembali informasi dengan metode DICE Similarity agar user dibuat untuk mempermudah dalam melakukan pencarian dokumen teks berbahasa Jawa. Metode DICE Similarity akan bisa menghitung dan menampilkan dokumen yang ada pada database sehingga antar dokumen memiliki bobot yang berbeda untuk menentukan dokumen mana yang paling mirip (similar) dengan query. Dokumen teks yang terletak di posisi atas akan ditempati oleh dokumen dengan bobot tertinggi hasil pencarian metode DICE. Evaluasi hasil pencarian Sistem temu kembali informasi dilakukan dengan uji recall dan precision dengan model persepsi. Sistem temu kembali informasi mampu melakukan pencarian dokumen dan menampilkan hasil pencarian dokumen memiliki rata-rata recall 0,03 dan rata-rata precision 0,79

Kata-Kunci: Sistem Temu Kembali Informasi, Basa Jawa , Dice Similarity

ABSTRACT

Information retrieval systems for Javanese text documents are difficult to find both from the online and offline world. Search for Java language text documents using the available information retrieval system obtained a search result with a large number of retrieved documents (high recall) so that the results become less accurate (low precision). ranking information retrieval system with the DICE Similarity method so that users are made to make it easier to search Java text documents. The DICE Similarity method will be able to calculate and display documents in the database so that the documents have different weights to determine which documents are most similar (similar) to the query. The text document located at the top position will be occupied by the document with the highest weight of the DICE method search results. Search results evaluation The information retrieval system is carried out by recall and precision tests with perception models. The information retrieval system is able to search documents and display document search results having an average recall of 0.03 and an average of 0.79 precision

Keywords: Information retrieval system, Java base, Dice Similarity

1. PENDAHULUAN

1.1 Latar Belakang

Bangsa Indonesia adalah bangsa yang berbudaya dan memiliki ratusan etnis yang didalamnya juga memiliki ratusan bahasa daerah. Bahasa Jawa sebagai bahasa yang paling banyak digunakan di wilayah Indonesia setelah bahasa Indonesia, dewasa ini mulai banyak ditinggalkan oleh kebanyakan orang, terutama Bahasa Jawa (halus). 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 sedikit yang membahasa bahasa Jawa dan belum menggunakan atau belum menyediakan pencarian informasi menggunakan sistem temu kembali informasi khusus berbahasa Jawa. Bahasa daerah adalah bahasa yang terkait akan latar belakang etnis, suku, budaya, yang begitu kaya di Indonesia. Bahasa daerah mencerminkan identitas bangsa ini, cermin kita sebagai bangsa yang kaya akan budaya dan bahasa. Bangsa Indonesia memiliki

sekitar 700 lebih bahasa daerah, tetapi yang tercatat oleh Kementerian Pendidikan dan Kebudayaan (Kemendikbud) hanya sekitar 470 saja.

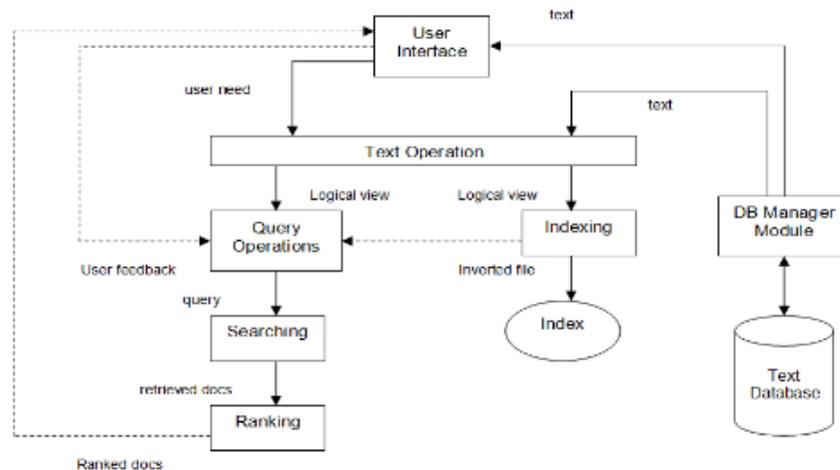
Ada 2 (dua) faktor utama yang menyebabkan bahasa Jawa (bahasa daerah pada umumnya) ditinggalkan oleh masyarakat, yaitu faktor internal dan faktor eksternal. Adapun Faktor internal yang dimaksud; 1) Melemahnya Sosialisasi dalam Keluarga, 2) Disorientasi Kurikulum Pendidikan, dan 3) Kurangnya Kesadaran Generasi Muda. Sedangkan Faktor eksternal yang menjadi penyebabnya yaitu; 1) Modernisasi dan Globalisasi, 2) Eksistensi Bahasa Asing di Indonesia, dan 3) Dominasi Kultural.

Pencarian informasi saat ini dilakukan dengan menggunakan Sistem temu kembali informasi atau sistem temu kembali informasi, user menuliskan query dan sistem temu kembali informasi akan menampilkan hasil pencarian. 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 pengguna (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.

2. METODE PENELITIAN

2.1. Dice Similarity

Proses Information Retrieval System (IRS) seperti pada gambar 1 menggunakan arsitektur yang sederhana. Sebelum dilakukannya proses temu kembali diperlukan pendefinisian database. Selanjutnya mengikuti tahapan proses; Dokumen-dokumen yang akan digunakan, Operasi yang akan digunakan dalam pencarian, dan model pengolahan teks (Baeza, 1999,h.9).



Gambar 1. The Process of Retrieving Information (Baeza, 1999,h.10)

Dice Similarity adalah metode untuk melihat tingkat kedekatan atau kesamaan (smilarity) term dengan cara pembobotan term. Dokumen dipandang sebagai sebuah vektor yang memiliki magnitude (jarak) dan direction (arah). Pada Vector Space Model, sebuah istilah direpresentasikan dengan sebuah dimensi dari ruang vektor. Relevansi sebuah dokumen ke sebuah query didasarkan pada similaritas diantara vektor dokumen dan vektor query. VSM memberikan sebuah kerangka pencocokan parsial adalah mungkin. Hal ini dicapai dengan menetapkan bobot non-biner untuk istilah indeks dalam query dan dokumen. Bobot istilah yang akhirnya digunakan untuk menghitung tingkat kesamaan antara setiap dokumen yang tersimpan dalam sistem dan permintaan user. Dokumen yang terambil disortir dalam urutan yang memiliki kemiripan, model vektor memperhitungkan pertimbangan dokumen yang

relevan dengan permintaan user. Hasilnya adalah himpunan dokumen yang terambil jauh lebih akurat (dalam arti sesuai dengan informasi yang dibutuhkan oleh user).

Dice similarity merupakan metode yang digunakan untuk menghitung tingkat kesamaan (similarity) antar dua buah objek. Untuk notasi himpunan dapat digunakan rumus (1):

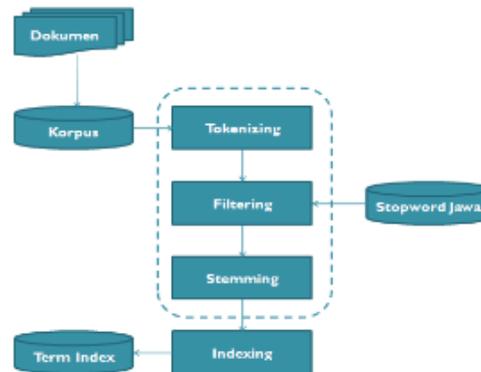
$$S_{Dice} = \frac{2 \sum_{i=1}^d P_i Q_i}{\sum_{i=1}^d P_i^2 + \sum_{i=1}^d Q_i^2} \quad (1)$$

dimana p dan q adalah dokumen yang berbeda. p_i adalah term i yang ada di dokumen p q_i adalah term i yang ada di dokumen q.

4 Proses perhitungan Dice Similarity sebelumnya melalui perhitungan term frequency (tf), Inverse Document Frequency (idf), term frequency Inverse Document Frequency (tfidf), Jarak query dan dokumen, Similaritas, dan Dice Similarity .

2.2. Stemmer Bahasa Jawa

Stemmer yang dipergunakan untuk mendukung STKI menggunakan metode rule base. Metode Rule Base Stemmer digunakan karena sifatnya yang fleksibel untuk digunakan sebagai stemmer berbagai macam bahasa dengan karakteristiknya yang lebih menekankan pada struktur morfologi suatu bahasa. Metode ini akan membuang (menghilangkan) Awalan (prefix), suffix (akhiran) dan infix (sisipan) dari term bentukan menjadi suatu term kata dasar. Stemmer diharapkan akan mampu mengurangi dimensi data (jumlah data terambil) sehingga nantinya akan meningkatkan performansi (akurasi) dari proses kategorisasi. Semakin sedikit dimensi data, maka akan semakin sedikit pula rule-rule untuk menentukan suatu dokumen terkategorisasi ke suatu kelas atau kategori tertentu sehingga bisa meningkatkan hasil kategorisasi. Pengurangan dimensi data dengan menggunakan stemming terjadi karena kata-kata yang memiliki kata dasar yang sama dikelompokkan menjadi satu atribut sehingga dimensi data bisa direduksi. Gambar 2 menunjukkan arsitektur stemmer jawa



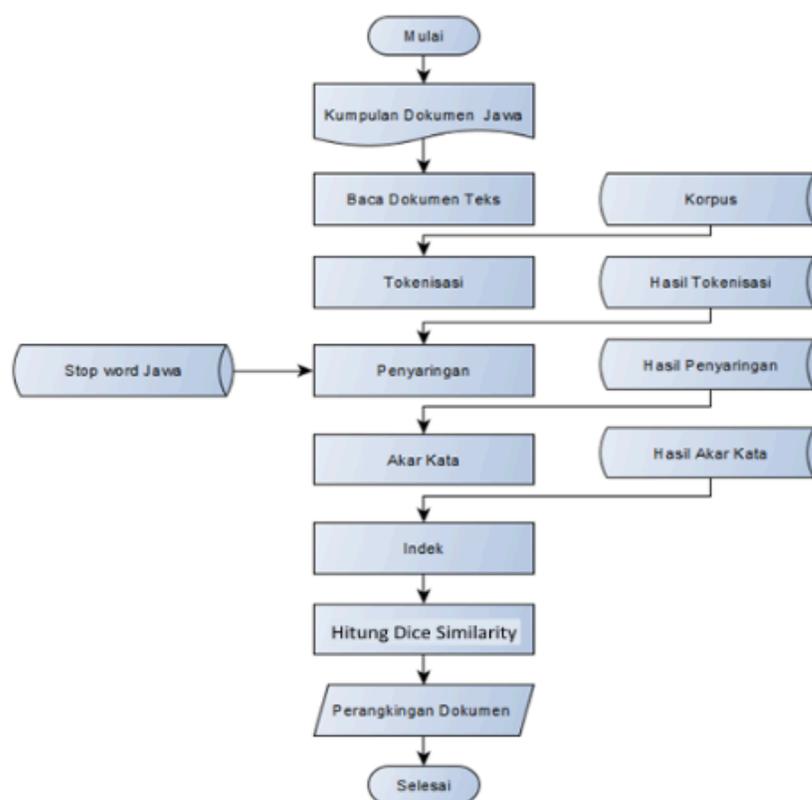
Gambar 2. Arsitektur Stemmer Jawa

3. HASIL DAN PEMBAHASAN

3.1. Sistem temu kembali informasi Dokumen Teks Bahasa Jawa

a. Flow Chart Sistem temu kembali informasi

Flowchart Sistem temu kembali informasi dokumen teks bahasa jawa diawali dengan input dokumen-dokumen kedalam korpus (gambar 3). Selanjutnya dokumen melalui proses preprosesing, dihitung bobotnya dan dibuat rankingnya berdasarkan bobot dokumen yang tertinggi.



Gambar 3. Diagram Alir STKI.

b. Tabel

Pada STKI ini menggunakan beberapa tabel untuk tempat meletakkan kumpulan data pada korpus, term-term hasil proses Tokenizing, Filtering dan Stemming. Selanjutnya untuk proses perhitungan dan pemeringkatan **Dice Similarity** digunakan tabel freq yaitu tabel yang berisi kumpulan term-term yang telah menjadi kata dasar. Berikut ini Rancangan tabel yang akan digunakan dalam STKI pada penelitian ini;

c. Tampilan Sistem temu kembali informasi Jawa

Pada Aplikasi interface ini akan ditampilkan kolom query yang bisa digunakan untuk memasukkan query oleh pengguna. Kotak button dengan label cari digunakan untuk memproses setelah query di input. Tombol button cari jika sudah diklik akan menampilkan abstraksi hasil pencarian

d. Implementasi Sistem temu kembali informasi Bahasa Jawa metode Dice Similarity

Dokumen abstrak di input dengan cara manual dengan format dokumen teks. Proses ini dilakukan dengan cara memasukkan abstrak-abstrak skripsi bahan kajian penelitian kedalam tabel korpus. Sebelum dimasukkan kedalam tabel, dibuat satu tabel dengan nama tabel korpus yang digunakan sebagai tempat data. Tabel korpus ini memiliki fiel-field id, judul, isi dan dokumen. Field id berisi urutan data penelitian didalam korpus yang tersusun sesuai dengan urutan input data. Field judul berisi judul skripsi. Field isi berisi abstrak skripsi dan field dokumen berisi nama dokumen dengan kode tertentu. Proses memasukkan dokumen ke dalam tabel korpus ini memerlukan waktu relative lama bergantung pada jumlah data yang akan di input kedalam tabel korpus (Tabel 1).

Tabel 1. Tabel Korpus

	Id	Judul	Isi	dokumen
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	846	Manfaat Daun Sirih	<p>Sinten ingkang mboten tepang kaliyan suruh? Saking simbah-simbah kita sedaya tiyang Indonesia kaliyan budaya "ninang suruh" piyambak nggadahi kalihi jenis. Enten suruh abrit, (jo, uga comeng. Masing-masing kagungan gina bonten. Sami nipi nanging mboten sami. Suruh piyambak sampun dangu panitados dening para leluhur kita sedaya konjuk dipundadosaken panjampen otomatis ingkang sanguh konjuk samukawis jenis sasadi. Nik uga ingkang nyiptake budaya "nyirih" punika piyambak konjuk ngiyataken untu uga gusti.</p> <p>Caramipun yaiku kaliyan nyampur godong suruh, apu, kapulaga, jambe, uga gambir. Lajeng, sedayanipun digulung ing godong suruh uga dipunglit utawi digosokkan ing untu ngantos wormi untu berubah. Saksampune punika, kemu kaliyan loya uga dipumbut. Pikanlukipun yaiku untu uga gusti ingkang sac kondisine. Padahal wanci zaman nyon doring enten prabot modern kados sekmenika.</p> <p>Bodah mendasar saking suruh-suruh ingkang tepang dados tanaman rambat kesebet yaiku wormi godong. Suruh abrit nggadahi wormi godong ingkang kabehlan. Semenara suruh (jo nggadahi wormi godong ingkang (jo segar. Somentara suruh comeng nggadahi wormi godong ingkang kecemengan. Ing antawis ketelu jenis suruh kesebet, enten kalihi ingkang paling populer. Raping kalihi ingkang populer punika yaiku gina godong suruh (jo uga abrit. Ing antaki niki badhe dipunembag kalihi jenis suruh kamentan: suruh (jo uga suruh abrit.</p>	BK1
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	848	Khasiat uga gina godong suruh (jo	<p>Godong suruh (jo nggadahi wormi (jo enem pajat. Suruh (jo tepang kaliyan nami latin Piper betle, ino. Suruh (jo kabet ibet keluarga piperaeaceae. Suruh (jo byasa merambat ing tanaman bonten. Nanging, suruh (jo uga sanguh merambat ing stti. Suruh (jo sugih badhe kandungan isah atsiri, feni propina, estragol, kavicol, hidroksikavicol, kavibetol, caryophyllene, allylpyronekatolol, cineole, cedrene, tanin, distase, pejah, terpenena, seskuiterpena, uga gendhis. Kandungan isah atsiri ing suruh (jo niki kalihi bebenera ing antawisipun dadosaken suruh (jo dados salah satunggala antijamur ingkang sae.</p> <p>Suruh (jo nggadahi gina dados punika:</p> <p>Ngampeni sakit untu Ngampeni gusti benzak</p>	BK2

1) Proses Tokenizing

Proses scanner dokumen korpus menggunakan format teks dilakukan dengan cara masuk kedalam dokumen korpus melalui perantara program php ke dalam database mysql. Proses scanner data dilakukan dengan cara scanner baris per baris, untuk tiap-tiap file naskah yang ada di dokumen. Tokenizing dimulai dengan memisahkan term-term yang ada pada dokumen korpus menjadi kumpulan term melalui proses scanner dengan dasar spasi. Selanjutnya term hasil proses Tokenizing di masukkan kedalam tabelawal dengan menyertakan field-field judul, term dan dokumen. Proses Tokenizing dilakukan dengan dua tahap yaitu tahap scanner term pada korpus kemudian term hasil scanner dimasukkan ke tabelawal dan tahap berikutnya adalah scanner term pada tabel awal dan menempatkan term hasil scanner di tabel kedua. Hasil scanner file pada proses (tabel 2)

Tabel 1. Tabel awal

judul	term	dokumen
Manfaat Daun Sirih	sinten	BK1
Manfaat Daun Sirih	ingkang	BK1
Manfaat Daun Sirih	mboten	BK1
Manfaat Daun Sirih	teping	BK1
Manfaat Daun Sirih	kaliyan	BK1
Manfaat Daun Sirih	suruh?	BK1
Manfaat Daun Sirih	saking	BK1
Manfaat Daun Sirih	simbah-simbah	BK1
Manfaat Daun Sirih	kita	BK1
Manfaat Daun Sirih	sedaya	BK1
Manfaat Daun Sirih	tiyang	BK1
Manfaat Daun Sirih	indonesia	BK1
Manfaat Daun Sirih	kaliyan	BK1
Manfaat Daun Sirih	budaya	BK1
Manfaat Daun Sirih	"ninang	BK1
Manfaat Daun Sirih	suruh"	BK1

2) Proses Filtering

Proses selanjutnya setelah proses Tokenizing adalah proses filtering. Proses filtering dilakukan untuk menghilangkan term-term yang tidak memiliki arti dengan menggunakan stopword list tala. Proses filtering adalah proses baca tabel kedua untuk diperiksa apakah semua term memiliki term-term yang termasuk dalam stopword list menurut tala. Jika dalam tabel kedua terdapat term-term yang termasuk dalam stopword, maka akan dilakukan penghilangan term-term tersebut. Hasil proses filtering selanjutnya dimasukkan dalam tabel freq (tabel 3)

Tabel 3. Proses Filtering

judul	term	dokumen
Manfaat Daun Sirih	sinten	BK1
Manfaat Daun Sirih	ingkang	BK1
Manfaat Daun Sirih	mboten	BK1
Manfaat Daun Sirih	teping	BK1
Manfaat Daun Sirih	kaliyan	BK1
Manfaat Daun Sirih	suruh?	BK1
Manfaat Daun Sirih	saking	BK1
Manfaat Daun Sirih	simbah-simbah	BK1
Manfaat Daun Sirih	kita	BK1
Manfaat Daun Sirih	sedaya	BK1
Manfaat Daun Sirih	tiyang	BK1
Manfaat Daun Sirih	Indonesia	BK1
Manfaat Daun Sirih	kaliyan	BK1
Manfaat Daun Sirih	budaya	BK1
Manfaat Daun Sirih	"ninang	BK1
Manfaat Daun Sirih	suruh"	BK1

3) Proses stopword Removal

Proses membuang **stopword (stopword removal)** dilakukan untuk menghilangkan term-term yang tidak memiliki arti dengan menggunakan **stopword jawa**. Proses ini dilakukan dengan cara scanner dan scanner tabel kedua

4) Proses Stemming

Proses stemming yang digunakan adalah proses stemmer menggunakan stemmer untuk bahasa Jawa ngoko berdasarkan stemmer bahasa Indonesia yang dibuat Tala. Proses stemming dengan menggunakan stemmer jawa melalui beberapa tahapan dan untuk mendukung proses ini juga digunakan stopword list jawa. Hasil akhir dari proses stemming adalah kumpulan term yang sudah menjadi kata dasar yang diinput dalam tabel freq. Proses stemming menghasilkan kumpulan term berupa kata dasar hasil scanner term pada tabel kedua. Proses stemming didukung stopword jawa yang digunakan untuk mengurangi term yang ada pada tabel kedua. Selanjutnya term hasil stemming di letakkan pada tabel freq

Tabel 3. Proses Stemming

judul	term	freq	freqpangkat
Manfaat Daun Sirih	sint	1	1
Manfaat Daun Sirih	ingkang	93	8649
Manfaat Daun Sirih	mbot	20	400
Manfaat Daun Sirih	tebang	4	16
Manfaat Daun Sirih	kaliy	35	1225
Manfaat Daun Sirih	suruh?	1	1
Manfaat Daun Sirih	saking	25	625
Manfaat Daun Sirih	simbah-simbah	1	1
Manfaat Daun Sirih	kita	16	256
Manfaat Daun Sirih	seday	18	324
Manfaat Daun Sirih	tiyang	19	361
Manfaat Daun Sirih	indonesi	11	121
Manfaat Daun Sirih	buday	5	25
Manfaat Daun Sirih	"ninang	1	1
Manfaat Daun Sirih	suruh"	1	1
Manfaat Daun Sirih	piyambak	7	49

5) Proses Indexing

Proses indexing dilakukan untuk mengambil atau meretrieve term-term yang ada pada tabel freq untuk selanjutnya diproses pada saat pencarian dilakukan oleh STKI. Proses perhitungan dilakukan langsung pada STKI saat query diproses oleh sistem. User memasukkan Kata Kunci (query) pada sistem temu kembali informasi, kemudian setelah kata kunci ditulis sistem temu kembali informasi akan melakukan pencarian query pada database dengan mengolahnya terlebih dahulu sesuai dengan arsitektur sistem temu kembali informasi menggunakan metode vector space model dan memberikan hasil pencarian.

6) Proses Perhitungan Dice Similarity

STKI metode Dice akan melakukan proses perhitungan dimulai dari menghitung tfidf, menghitung jarak query dan jarak dokumen, menghitung similaritas produk, dan menghitung bobot dokumen. STKI akan mengeksekusi query dari user dan akan mengolah query tersebut. Query yang di input oleh user selanjutnya akan dilakukan pencarian pada tabel freq kemudian dilakukan perhitungan pembobotan menggunakan metode **Dice Similarity**. Perhitungan dilakukan dalam sistem pencarian, sistem pencarian akan melakukan perhitungan kemudian akan menampilkan hasilnya. Hasil pencarian akan menampilkan nama dokumen di korpus, kemudian bobot similaritas dan disusun berdasarkan perankingan. Bobot terbesar akan menempati ranking teratas pada hasil pencarian.

Dice Similarity adalah metode untuk melihat tingkat kedekatan atau kesamaan (similarity) term dengan cara pembobotan term. Dokumen dipandang sebagai sebuah vektor yang memiliki magnitude (jarak) dan direction (arah). Hal ini dicapai dengan menetapkan bobot non-biner untuk istilah indeks dalam query dan dokumen. Bobot istilah yang akhirnya digunakan untuk menghitung tingkat kesamaan antara setiap dokumen yang tersimpan dalam sistem dan permintaan user. Dokumen yang terambil disortir dalam urutan yang memiliki kemiripan, model vektor memperhitungkan pertimbangan dokumen yang relevan dengan permintaan user. Hasilnya adalah himpunan dokumen yang terambil jauh lebih akurat (dalam arti sesuai dengan informasi yang dibutuhkan oleh user).

Dice similarity merupakan metode yang digunakan untuk menghitung tingkat kesamaan (similarity) antar dua buah objek. Untuk notasi himpunan dapat digunakan rumus (1):

$$S_{Dice} = \frac{2 \sum_{i=1}^d P_i Q_i}{\sum_{i=1}^d P_i^2 + \sum_{i=1}^d Q_i^2}$$

dimana p dan q adalah dokumen yang berbeda. p_i adalah term i yang ada di dokumen p q_i adalah term i yang ada di dokumen q.

e. Aplikasi STKI

Sistem temu kembali informasi Dice dirancang agar user mudah mengguru⁷kan dalam mencari dokumen yang relevan. Tampilan (interface) juga dirancang seperti sistem temu kembali informasi pada umumnya, sehingga siapapun usernya akan langsung mudah beradaptasi dalam menggunakan sistem temu kembali informasi. Prosedur menggunakan STKI ini sangat mudah, yaitu user hanya perlu menuliskan query atau kata kunci yang akan di cari pada kotak dialog kemudian setelah query di masukkan user tinggal mengklik tombol cari atau tekan enter.

Studi kasus pada aplikasi Sistem temu kembali informasi Dice ini menggunakan dokumen-dokumen Basa Jawa pada Majalah Online Penjebar Semangad yang terdapat pada 3 kategori yaitu; Kejawan, kebatinan, dan pasujarahan. Query yang dimasukkan pada Information Retrieval System adalah Studi kasus pada aplikasi sistem temu kembali informasi ini menggunakan dokumen-dokumen teks berbahasa jawa . Query yang dimasukkan pada sistem temu kembali informasi adalah keyword dengan 2 term yaitu "wayang jawi", 3 term "sejarah budaya wayang", "budaya wayang jawi" 4 term "sejarah budaya wayang wong". 5 term "Cerita sejarah budaya wayang wong".



Gambar 4 Aplikasi STKI Jawa



Gambar 5 Hasil Pencarian keyword

Hasil pencarian dokumen dengan keyword "wayang jawi", menunjukkan dokumen dengan bobot tertinggi adalah dokumen letak dokumen BK9 (bobot 0,001).

f. Pengujian recall dan precision

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

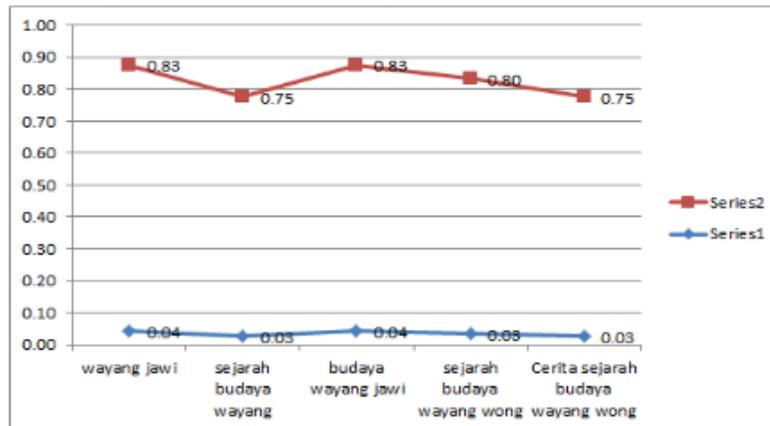
$$R = \frac{\text{Number of relevant items retrieved}}{\text{Total number of relevant items in collection}} \quad (2)$$

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

Tabel 1 Hasil Pengujian Recall dan Precision

No	Query	Recall	Precision
1	wayang jawi	0.04	0.83
2	sejarah budaya wayang	0.03	0.75
3	budaya wayang jawi	0.04	0.83
4	sejarah budaya wayang wong	0.03	0.80
5	Cerita sejarah budaya wayang wong	0.03	0.75

Hasil uji recall dan precision berdasarkan persepsi bisa dilihat pada gambar 5



Gambar 5 Diagram hasil uji Recall dan precision

4. KESIMPULAN DAN SARAN

4.1. Kesimpulan

- Sistem temu kembali informasi (information retrieval system) Dokumen Teks Bahasa Jawa mampu melakukan pencarian dokumen teks bahasa jawa dan menampilkan hasil pencarian dokumen teks berbahasa Jawa dengan disertai bobot tiap dokumen beserta letak dokumen dengan metode DICE Similarity.
- Hasil Uji recall dan precision sistem temu kembali informasi menunjukkan hasil pencarian dokumen teks memiliki rata-rata recall = 0,03 dan rata-rata precision = 0,79.

4.2. Saran

Stemmer Jawa masih perlu perbaikan untuk Proses stemming perlu diperbaiki karena hasil yang didapatkan masih belum bisa sepenuhnya membuat semua term kedalam bentuk term kata dasar dengan benar.

DAFTAR PUSTAKA

- Budi, I., Aji, R.F., 2006. Efektifitas Seleksi Fitur dalam Sistem Temu Kembali Informasi. Seminar Nasional Aplikasi Teknologi Informasi (SNATI), ISSN : 1907-5022.
- Bum, K.Y., 2010. An autonomous assessment system based on combined latent semantic kernels. Expert Systems with Applications: An International Journal , Volume 37 Issue 4.
- Kadir, A., 2001. Dasar Pemrograman Web Dinamis menggunakan PHP. Penerbit Andi. Yogyakarta.
- Khuat Thanh Tung , (2016) A Comparison of Algorithms used to measure the Similarity between two documents, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 Issue 4, April 2015
- Manning, C., Raghavan, P., 2007. An Introduction to Information Retrieval, Stanford. USA.
- Manoj Chahal, 2016. Information Retrieval using Dice Similarity Coefficient , International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 6, June 2016, ISSN: 2277 128
- Meadow, C.T., 1997. Text Information Retrieval Systems. Academic Press. New York.
- 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.
- R. Umamaheswari, K. Rajesh, 2014, Text Clustering Using Cosine Similarity and Matrix Factorization Cosine Similarity, International Journal of Research in Computer and Communication Technology, Vol 3, Issue 10, October - 2014
- Salton, G., 1989, Automatic Text Processing, The Transformation, Analysis, and Retrieval of information by computer. Addison – Wesley Publishing Company, Inc. USA.
- Vikas Thada , 2015. Comparison of Jaccard, Dice, Cosine Similarity Coefficient To Find Best Fitness Value for Web, Department of Computer Science and Engineering Dr. K.N.M University, Newai, Rajasthan, India
- Yates, R.B, 1999. Modern Information Retrieval, Addison Wesley-Pearson international edition, Boston. USA.