

Hak Cipta Diindungi Undang-Undang

1. Diarangi mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB II

LANDASAN TEORI

2.1 *Information Retrieval (IR)*

Information Retrieval (IR) merupakan prosedur pengambilan kembali data yang relevan dari koleksi data tidak terstruktur yang memenuhi tuntutan pengguna. Metode pencarian berbasis kata kunci yang digunakan untuk mengambil file teks yang relevan dari file teks yang diindeks itu relevan untuk pengguna (Khan, 2014). Istilah *information retrieval (IR)* digunakan untuk menggambarkan proses temuan materi (biasanya dokumen) bersifat tidak terstruktur (biasanya teks) yang memuaskan informasi. Kebutuhan dari dalam koleksi besar (biasanya disimpan di komputer) (Nkweteyim, 2014).

Information retrieval (IR) adalah bidang yang berkaitan dengan organisasi dan pengambilan informasi berbasis pengetahuan yang berfokus terutama pada informasi tekstual, namun dengan bertambahnya jumlah multimedia dan juga *database* yang lebih kompleks, sifat *IR* telah berubah. Sistem pencarian informasi dirancang untuk mengambil dokumen atau informasi yang dibutuhkan oleh komunitas pengguna. Ini harus membuat informasi yang benar tersedia bagi pengguna yang tepat. Dengan demikian, sistem pencarian informasi bertujuan untuk mengumpulkan dan mengatur informasi di satu atau lebih bidang subjek untuk memberikannya kepada pengguna segera setelah diminta. (Onwuchekwa & Jegede, 2011)

Information Retrieval (IR) adalah proses pengumpulan data, disimpan, dan dicari untuk tujuan penemuan pengetahuan sebagai respons terhadap permintaan pengguna (*query*). Pada proses tersebut melibatkan berbagai tahap yang dimulai dengan mewakili data lalu diakhiri dengan mengembalikan informasi yang relevan kepada pengguna. Adapun tahapannya meliputi penyaringan, pencarian, pencocokan dan operasi pemeringkatan. Tujuan utama dari sistem pencarian informasi adalah untuk menemukan informasi yang relevan atau dokumen yang memenuhi kebutuhan informasi pengguna.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2.1.1 Model IR (*Information Retrieval*)

Dalam penelitian (Roohparvar, 2015), terdapat 4 (empat) model yang bisa digunakan untuk *information retrieval*, antara lain model *boolean*, *Probabilistic Model*, *Vector Space Model*, dan *Inference Network Model*. Adapun penjelasannya sebagai berikut:

1. *Boolean Model*

Model *boolean* adalah model pengambilan informasi pertama dan mungkin juga model yang paling dikritik. Model dapat dijelaskan dengan memikirkan istilah *query* sebagai definisi yang tidak ambigu dari sekumpulan dokumen. Misalnya, istilah *query* ekonomi secara sederhana mendefinisikan kumpulan semua dokumen yang diindeks dengan istilah ekonomi. Model *boolean* memungkinkan penggunaan operator aljabar *Boolean*, *AND*, *OR* dan *NOT*, untuk formulasi *query*, tetapi memiliki satu kerugian utama: sistem *boolean* tidak dapat memberi peringkat pada daftar dokumen yang dikembalikan.

2. *Probabilistic Model*

Karakteristik yang paling penting dari model probabilistik adalah usahanya untuk menentukan peringkat dokumen berdasarkan kemungkinan relevansi mereka dengan *query*. Dokumen dan pertanyaan diwakili oleh vektor biner $\sim d$ dan $\sim q$, setiap elemen vektor menunjukkan apakah atribut atau istilah dokumen terjadi dalam dokumen atau permintaan, atau tidak.

3. *Vector Space Model*

Vector Space Model dapat dikarakterisasi dengan baik oleh upaya untuk menentukan peringkat dokumen dengan kemiripan antara *query* dan setiap dokumen. Dalam *Vector Space Model* (VSM), dokumen dan *query* direpresentasikan sebagai vektor dan sudut antara keduanya. vektor dihitung menggunakan fungsi cosinus kesamaan.

4. *Inference Network Model*

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Dalam model ini, pengambilan dokumen dimodelkan sebagai proses inferensi dalam jaringan inferensi. Sebagian besar teknik yang digunakan oleh sistem IR dapat diimplementasikan di bawah model ini.

2.1.2 Arsitektur *Information Retrieval*

Secara umum, *Information Retrieval* dibangun dengan dua tahapan, yaitu melakukan *preprocessing* terhadap *database* dan metode untuk menghitung relevansi antara dokumen-dokumen dalam *database* yang sudah dilakukan *preprocessing* sebelumnya dan *query* pengguna (Cios, 2007). Pada tahapan *preprocessing*, sistem yang berurusan dengan dokumen *semistructured* biasanya memberikan tag tertentu pada *term* atau bagian dari dokumen, sedangkan pada dokumen tidak terstruktur proses ini dilewati dan membiarkan *term* tanpa imbuhan tag. *Query* yang dimasukkan pengguna dikonversi sesuai aturan tertentu untuk mengekstrak *terms* penting yang sejalan dengan *term* yang sebelumnya yang telah diekstrak dari dokumen dan menghitung relevansi antara *query* dan dokumen berdasarkan pada *term* tersebut. Sebagai hasilnya, sistem mengembalikan suatu daftar dokumen terurut sesuai nilai kemiripannya dengan *query* pengguna.

2.2 Pembangunan *Index*

Untuk mempercepat proses pencarian dokumen, salah satu cara yang perlu dilakukan adalah dengan menggunakan *index* yang merupakan tugas pokok dalam *information retrieval*. Salah satu proses untuk mendapatkan kata-kata yang digunakan dalam teks tertentu adalah proses *tokenization*, yaitu dimana dokumen teks dibagi menjadi aliran kata-kata yang terpisah kemudian dengan menghapus semua tanda baca dan karakter non-teks lain dengan spasi tunggal (Hotho, 2005).

2.3 *Preprocessing* Dokumen

Preprocessing dokumen atau biasa disebut *text preprocessing* merupakan tahapan untuk mempersiapkan teks menjadi data yang akan diolah di tahapan berikutnya. Inputan awal pada proses tersebut adalah berupa teks dokumen dan

Hak Cipta Diindungi Undang-Undang

membagi dokumen menjadi beberapa tahapan seperti *case folding*, *tokenization*, *stopword removal*, dan *stemming* (Kadhim, 2014).

2.3.1 Case Folding

Case folding merupakan tahapan proses mengubah semua huruf dalam *document text* menjadi huruf kecil, serta menghilangkan karakter selain huruf a-z dan dianggap sebagai delimiter seperti tanda koma, titik koma, atau titik dua, dan lain-lain.

2.3.2 Tokenization

Salah satu tahapan dalam pembuatan *index* adalah *Tokenization* yaitu dengan cara memotong *string* kalimat dalam teks dokumen menjadi kata tunggal. Menurut Kadhim (2014), tokenisasi dalam pengertiannya tidak hanya memisahkan *string* unit pemrosesan dasar, tetapi juga interpretasi dan grup token terisolasi untuk membuat token level yang lebih tinggi.

2.3.3 Stopword Removal

Stopword didefinisikan sebagai kata yang tidak relevan sehubungan dengan subjek utama dari *database*, meskipun kata tersebut sering muncul dalam dokumen dan ini akan mempengaruhi hasil pencarian karena kata-kata yang sering muncul tersebut memiliki frekuensi yang tinggi. Maka dari itu *stopword* atau kata yang tidak relevan tersebut harus dihilangkan dengan proses *stopword removal* agar hasil pencarian lebih meningkat sesuai dengan *query* yang diinginkan pengguna. Contoh daftar *stopword*: dan, pada, di, belum, atau, apa, kenapa, dimana, mana, kalau, mungkin, bila, lagi, apabila, walaupun, bagaimanapun, meskipun, agar, jika, untuk, akan, daripada, karena, juga, dengan, adalah, dapat, kah, dan sebagainya.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.3.4 Stemming

Proses *stemming* diperlukan untuk meningkatkan performa pada sistem *information retrieval*. Setiap kata tentunya ada yang memiliki kata imbuhan dan *stemming* berperan untuk mengubahnya menjadi kata dasar. Bahasa Indonesia memiliki imbuhan yang terdiri dari kombinasi sebagai berikut:

Prefiks1 + Prefiks2 + Kata Dasar + Sufiks3 + Sufiks2 + Sufiks1

Contohnya adalah mem+per+malu+kan yang merupakan kata dasar dari kata “malu”.

Proses *stemming* pada penelitian ini menggunakan Algoritma Nazief & Adriani yang merupakan algoritma *stemming* untuk teks berbahasa Indonesia yang mempunyai tingkat keakuratan yang lebih baik dengan algoritma lainnya (Agusta, 2009).

2.3.4.1 Algoritma Nazief & Adriani

Algoritma Nazief & Adriani mengacu pada aturan morfologi bahasa Indonesia yang mengelompokkan imbuhan, yaitu imbuhan yang diperbolehkan atau imbuhan yang tidak diperbolehkan. Pengelompokan ini termasuk imbuhan di depan (awalan), imbuhan kata di belakang (akhiran), imbuhan kata di tengah (sisipan) dan kombinasi imbuhan pada awal dan akhir kata (konfiks).

Berikut ini adalah langkah-langkah yang dilakukan oleh algoritma Nazief & Adriani (Agusta, 2009):

1. Cari kata yang akan di-*stemming* dalam kamus. Jika ditemukan maka diasumsikan bahwa kata tersebut adalah *root word*, maka algoritma berhenti.
2. *Inflection suffixes* (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”) dihilangkan.

Jika berupa partikel (“-lah”, “-kah”, “-tah” atau “-pun”) maka langkah ini diulangi lagi untuk menghapus *possesive pronouns* (“-ku”, “-mu”, atau “-nya”), jika ada.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3. *Derivation suffixes* (“-i”, “-an” atau “-kan”) dihilangkan. Jika kata ditemukan di kamus, maka algoritma berhenti. Jika tidak maka ke langkah 3a.
 - a. Jika “-an” telah dihapus dan huruf terakhir dari kata tersebut adalah “k”, maka “-k” juga ikut dihapus. Jika kata tersebut ditemukan dalam kamus maka algoritma berhenti. Jika tidak ditemukan maka lakukan langkah 3b.
 - b. Akhiran yang dihapus (“-i”, “-an” atau “-kan”) dikembalikan, lanjut ke langkah 4.
4. *Derivation prefix* dihilangkan. Jika pada langkah 3 ada *sufixes* yang dihapus maka pergi ke langkah 4a, jika tidak pergi ke langkah 4b.
 - a. Periksa kombinasi awalan-akhiran yang tidak diijinkan. Jika ditemukan maka algoritma berhenti, jika tidak pergi ke langkah 4b.
 - b. For $i = 1$ to 3, tentukan tipe awalan kemudian hapus awalan. Jika *root word* belum juga ditemukan lakukan langkah 5. Jika sudah maka algoritma berhenti. Catatan: jika awalan kedua sama dengan awalan pertama maka algoritma berhenti.
5. Jika semua langkah telah selesai tetapi tidak juga berhasil maka kata awal diasumsikan sebagai *root word*. Proses selesai.

2.4 Pembobotan Kata

Setiap *term* yang telah di-*index* diberikan bobot sesuai dengan struktur pembobotan yang dipilih, apakah pembobotan lokal, global atau kombinasi keduanya. Jika menggunakan pembobotan lokal maka, pembobotan *term* diekspresikan sebagai *tf* (*term frequency*). Namun, jika pembobotan global yang digunakan maka, pembobotan *term* didapatkan melalui nilai *idf* (*inverse document frequency*). Beberapa aplikasi juga ada yang menerapkan pembobotan kombinasi keduanya yaitu, dengan mengalikan bobot lokal dan global ($tf \cdot idf$) (Bintana, 2012).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Term Frequency

Empat cara yang dapat digunakan untuk memperoleh nilai *term frequency* (*tf*), yaitu:

- a. *Raw term frequency*. Nilai *tf* sebuah *term* diperoleh berdasarkan jumlah kemunculan *term* tersebut dalam dokumen. Contoh kasus dimana *term* muncul sebanyak dua kali dalam suatu dokumen maka, nilai *tf term* tersebut adalah 2.
- b. *Logarithm term frequency*. Hal ini untuk menghindari dominasi dokumen yang mengandung sedikit *term* dalam *query*, namun mempunyai frekuensi yang tinggi. Cara ini menggunakan fungsi logaritmik matematika untuk memperoleh nilai *tf*.

$$tf = 1 + \log(tf) \quad (2.1)$$

- c. *Binary term frequency*. Hanya memperhatikan apakah suatu *term* ada atau tidak dalam dokumen. Jika ada, maka *tf* diberi nilai 1, jika tidak ada diberi nilai 0. Pada cara ini jumlah kemunculan *term* dalam dokumen tidak berpengaruh.

Augmented Term Frequency. Nilai *tf* adalah jumlah kemunculan suatu *term* pada sebuah dokumen, sedangkan nilai *max(tf)* adalah jumlah kemunculan terbanyak sebuah *term* pada dokumen yang sama.

$$tf = 0.5 + 0.5 \times \frac{tf}{\max(tf)} \quad (2.2)$$

Inverse Document Frequency

Inverse document frequency (*idf*) digunakan untuk memberikan tekanan terhadap dominasi *term* yang sering muncul di berbagai dokumen. Hal ini diperlukan karena *term* yang banyak muncul di berbagai dokumen, dapat dianggap sebagai *term* umum (*common term*) sehingga tidak penting nilainya. Pembobotan akan memperhitungkan faktor kebalikan frekuensi dokumen yang mengandung suatu *term* (*inverse document frequency*).

$$idf(t) = \log\left(\frac{N}{df(t)}\right) \quad (2.3)$$

Hak Cipta Diindungi Undang-Undang

Keterangan:

N : jumlah total dokumen

$idf(t)$: *Inverse document frequency* atau jumlah dokumen dalam *corpus* yang mengandung *term t*.

Pada penelitian ini akan menggunakan *tf-idf* secara bersamaan dengan mengalikan keduanya dengan persamaan berikut:

$$(tf-idf)t_j = tf(t).idf(t) \quad (2.4)$$

Menurut (Arifin, 2002) penggunaan *tf* (*term frequency*) saja hanya mampu mendukung proporsi jumlah dokumen yang dapat ditemukan kembali oleh proses pencarian dokumen pada sistem, sedangkan proporsi jumlah dokumen yang ditemukan dan dianggap relevan untuk kebutuhan pengguna akan lebih meningkat bila vektor bobot tersebut menggunakan *term* yang jarang muncul pada koleksi dokumen. Hal ini dapat dilakukan dengan menghitung *idf* (*invers document frequency*). Sehingga kriteria *term* yang paling tepat adalah *term* yang sering muncul dalam dokumen secara individu, namun jarang dijumpai pada dokumen lainnya.

2.5 *Vector Space Model (VSM)*

Vector space model merupakan model paling sederhana dan paling produktif. Model vektor ini merepresentasikan *term* yang digunakan baik oleh dokumen maupun oleh *query* (Arifin, 2002). Dalam *Information Retrieval*, kemiripan antar dokumen didefinisikan berdasarkan representasi *bag-of-words* dan dikonversi ke suatu *vector space model (Vector Space Model)*. Pada *vector space model*, setiap dokumen dan *query* pengguna direpresentasikan oleh suatu vektor *n*-dimensi. Tiap dimensi pada vektor tersebut diwakili oleh satu *term*. *Term* yang digunakan biasanya berpatokan pada *term* yang ada pada *query*, sehingga *term* yang ada pada dokumen tetapi tidak ada pada *query* biasanya diabaikan (Mustaqim, 2008).

1. Diarangi mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

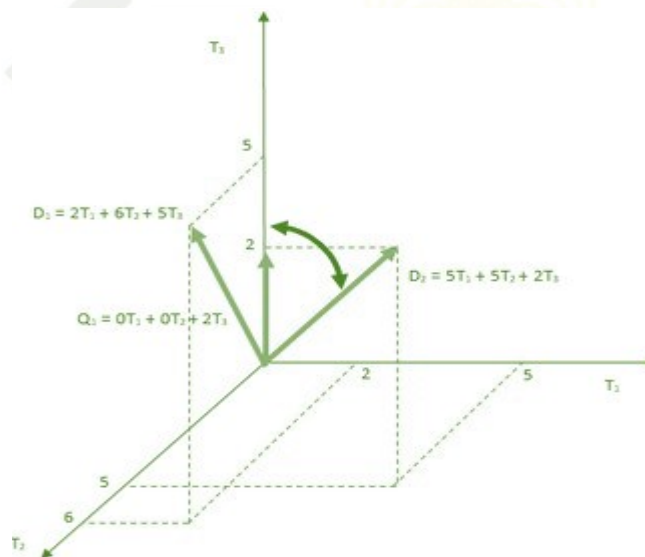
- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Pada setiap dokumen di dalam *database* dan *query* pengguna direpresentasikan oleh suatu vektor multi-dimensi. Dimensi yang sesuai dengan jumlah *term* pada dokumen yang terlibat pada *vector space model* adalah:

1. *Vocabulary* merupakan kumpulan semua *term* berbeda yang tersisa dari dokumen setelah *preprocessing* dan mengandung *term index*. Kumpulan *term* ini membentuk suatu ruang vektor.
2. Setiap *term* i di dalam dokumen atau *query* j , diberikan suatu bobot (*weight*) bernilai real w_{ij} .
3. Dokumen dan *query* diekspresikan sebagai vektor t dimensi $d_j = (w_{1j}, w_{2j}, \dots, w_{tj})$ dan terdapat n dokumen didalam koleksi, yaitu $j = 1, 2, \dots, n$.

Berikut merupakan contoh dari *vector space model* tiga dimensi untuk dua dokumen dan satu *query* pengguna, dan tiga *term*:



Gambar 2.1 Contoh Vector Space Model dengan Dua Dokumen, serta Query Pengguna (sumber: Cios. 2007)

2.5.1 Cosine Similarity

Salah satu ukuran kemiripan teks yang populer digunakan pada *VSM* untuk pencarian dokumen adalah *cosine similarity* (Krzysztof J. Cios, dkk. 2007). Konsep dari *cosine similarity* yaitu menghitung nilai cosinus sudut antara dua vektor yaitu jika diberikan dokumen yang dipresentasikan oleh vektor d_j dan

Tabel 2.2 Penelitian Terkait

NO	Tahun	Peneliti	Judul	Hasil	Perbedaan
1	2009	Agusta, Ledy.	Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief & Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia.	Pada penelitian ini penulis melakukan perbandingan algoritma <i>stemming</i> Nazief & Adriani dengan algoritma Porter terhadap 30 dokumen teks yang bervariasi dan hasilnya algoritma Nazief & Adriani lebih tinggi presisinya dari algoritma Porter yaitu sebesar 95.9% Untuk Nazief &	Pada penelitian ini melakukan komparasi algoritma <i>stemming</i> Nazief & Adriani dan algoritma Porter dengan dokumen teks yang bervariasi sedangkan penulis melakukan <i>stemming</i> hadits dengan algoritma <i>stemming</i> Nazief & Adriani dan pencarian dengan menggunakan metode VSM (Vector Space Model)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

				Adriani sedangkan algoritma Porter hanya sebesar 89.7%.	
2	2014	Nesdi E. Rozanda, Arif Marsal, Kiki Iswanti	Rancang Bangun Sistem Informasi Hadits Menggunakan Teknik Temu Kembali Informasi Model Ruang Vektor	Pada penelitian ini melakukan pencarian hadits dengan menggunakan Algoritma VSM (<i>Vector Space Model</i>) dan menghasilkan hasil uji dari tujuh <i>query</i> dengan perhitungan precision dan recall, dan didapatkan hasil rata-rata precision 65% dan menghasilkan recall rata- rata 0,97.	Penelitian ini menggunakan data pencarian hadits tertentu, sedangkan penulis melakukan pencarian menggunakan data hadits yang diambil dari kitab Al- Lu`lu` wal Marjan karya syaikh Muhammad Fuad Abdul Baqi dengan menggunakan metode yang sama.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2011	Heru Adi Darmawan, Tutut Wuriyanto, Akh. Masturi	Rancang Bangun Aplikasi Search Engine Tafsir Al-Qur'an Menggunakan Teknik Text Mining Dengan Algoritma VSM (Vector Space Model)	Pada penelitian ini melakukan pencarian dengan metode <i>VSM</i> terhadap teks tafsir al-Qur'an dengan hasil uji coba terhadap 35 <i>query</i> dimana nilai rata-rata tingkat keakuratan terhadap hasil pencarian dengan nilai 73.8%.	Pada penelitian ini melakukan pencarian tafsir Al-Qura'n dengan proses <i>stemming</i> menggunakan algoritma <i>Enhanced Confix Stripping Stemmer</i> sedangkan penulis melakukan pencarian hadits dengan proses <i>stemming</i> menggunakan algoritma Nazief & Adriani dengan metode yang sama.
4	2008 Elvia Budianita	Studi Temu Kembali	Dari hasil studi ini diperoleh	Pada penelitian ini

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

		<p>Informasi (<i>Information Retrieval</i>) Dengan Model Ruang Vektor</p>	<p>kesimpulan bahwa semakin banyak suatu istilah muncul dalam dokumen maka bobot suatu istilah semakin kecil, begitu pula sebaliknya. Dokumen yang berhasil ditemukan sistem (ruang vektor) bersifat <i>partial match</i> (agak relevan), dan frekuensi istilah <i>quert</i> pada setiap dokumen mempengaruhi nilai <i>similarity coefficiance</i>. Selain itu, kemampuan sistem <i>information retrieval</i> dengan model</p>	<p>menggunakan metode <i>vector space model</i> namun tidak menggunakan tahapan <i>stemming</i> sedangkan penulis menggunakan tahapan <i>stemming</i></p>
--	--	---	--	---



				ruang vektor rata-rata berada pada kelas interval kesesuaian sedang (0.34-0.66).	

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



UIN SUSKA RIAU