

SISTEM TEMU KEMBALI INFORMASI PADA DOKUMEN CERITA KISAH NABI

Muhammad Syukron Isnan Banani¹, Fatkhul Amin²

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

ABSTRAK

Mesin pencari saat ini sudah dapat membantu untuk menemukan dokumen. Akan tetapi dokumen yang terambil masih banyak jumlahnya. Sehingga memerlukan waktu yang lama untuk menemukan dokumen yang dibutuhkan. Kisah nabi adalah cerita tentang perjalanan orang yang menjadi pilihan Allah. Karena didalam kisah nabi banyak hal yang bisa diteladani tentang akhlak nabi, sifat-sifat para nabi dan budi pekerti sebagai pedoman hidup dalam ajaran umat muslim. Banyak kisah nabi yang dikenal masyarakat dianggap menjadi kendala karena terbatasnya sarana yang tersedia menemukan kisah nabi, oleh sebab itu dibangunlah suatu sistem temu kembali informasi (STKI) tentang kisah nabi yang menggunakan vector space model (VSM). Dengan VSM akan didapat hasil pencarian yang akurat.

Kata Kunci : *Sistem Temu Kembali Informasi, Vector Space Model, Kisah nabi*

1. PENDAHULUAN

Perkembangan teknologi informasi terutama yang berhubungan dengan kemudahan mengakses informasi, agar pengguna dapat memanfaatkan sumber data informasi dengan mudah, cepat, tepat dan efisien. Saat ini banyak orang telah meninggalkan proses penelusuran informasi secara manual yang membutuhkan waktu yang lama. Pada dasarnya sistem temu kembali informasi yang bertujuan untuk menyimpan informasi adalah sebuah kumpulan laporan yang tersimpan secara bersama-sama dalam satu tempat penyimpanan. mesin pencarian dapat digunakan di bidang edukasi islam, karena bagi orang tua untuk mendapatkan informasi yang berkualitas dan relevan sangat penting buat edukasi anak. Kisah nabi merupakan cerita islam yang tercantum dalam AlQur'an, karena di dalam kisah nabi banyak yang bisa diteladani tentang akhlak nabi, sifat-sifat para nabi dan budi pekerti sebagai pedoman hidup dalam ajaran umat muslim.

Masyarakat sekarang kurang sadar akan pentingnya pengetahuan tentang kisah para nabi, untuk memperoleh informasi tentang kisah nabi yaitu dengan bertanya kepada orang, mencari dalam buku, dan pencarian menggunakan sistem. Terkadang mengalami kendala yaitu jika bertanya kepada orang, terkadang tidak semua orang tahu kisah nabi, walaupun ada orang yang tahu kisah nabi semua. *vector space model* (VSM) akan memberikan hasil perolehan pencarian dokumen yang terpanggil tidak cukup banyak namun dokumen-dokumen yang peroleh tepat dan relevan sesuai dengan kata kunci yang diinputkan oleh user. *vector space model* (VSM) pada sistem temu kembali informasi ini sangat membantu dan menguntungkan user karena dengan hasil pencarian yang tepat dan relevan maka user hanya memerlukan waktu yang relatif lebih sedikit dan tidak membuang banyak waktu.

Mengatasi hal tersebut, maka dengan salah satu cara yang cepat dan tepat sesuai menurut kriteria yang dicari, dapat digunakan sistem informasi kisah nabi berbasis Sistem Temu Kembali Informasi (Information Retrieval).

2. TINJAUAN PUSTAKA

2.1 Sistem Temu Kembali Informasi

Sistem temu kembali informasi (*information retrieval system*) merupakan suatu sistem yang menemukan (*retrieve*) informasi yang sesuai dengan kebutuhan *user* dari kumpulan informasi secara otomatis. Prinsip kerja sistem temu kembali informasi jika ada sebuah kumpulan dokumen dan seorang *user* yang memformulasikan sebuah pertanyaan (*request* atau *query*). Jawaban dari pertanyaan tersebut adalah sekumpulan dokumen yang relevan dan membuang dokumen yang tidak relevan (salton, 1989).

Sistem temu kembali informasi akan mengambil salah satu dari kemungkinan tersebut. Sistem temu kembali informasi dibagi dalam dua komponen utama yaitu sistem pengindeksan (*indexing*) menghasilkan basis data sistem dan temu kembali merupakan gabungan dari *user interface* dan *look-up-table*. Sistem temu kembali informasi didesain untuk menemukan dokumen atau informasi yang diperlukan oleh *user*.

2.2 Tokenisasi

Tokenisasi merupakan proses pemisahan suatu rangkaian karakter berdasarkan karakter berdasarkan karakter spasi, dan mungkin pada waktu yang bersamaan dilakukan juga proses penghapusan karakter tertentu, seperti tanda baca. Sebagai contoh, kata-kata “*computer*”, “*computing*”, dan “*compute*” semua berasal dari *term* yang sama yaitu “*comput*”, tanpa pengetahuan sebelum dari morfologi bahasa Inggris. Token seringkali disebut sebagai istilah (*term*) atau kata, sebagai contoh sebuah token merupakan suatu urutan karakter dari dokumen tertentu yang dikelompokkan sebagai unit semantik yang berguna untuk diproses (Salton, 1989).

2.3 Filtering

Proses filtering menggunakan daftar *stopword* yang digunakan oleh Tala (2003), yang merupakan *stopword* bahasa indonesia yang berisi kata seperti: ada, yang, ke, kepada, dan lain sebagainya. Penelitian ini penulis menggunakan list *stopword tala* sebagai daftar listnya.

Stopword adalah kata umum (*common word*) yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. *Stopword* umumnya dimanfaatkan dalam *task information retrieval*, oleh google. Contoh *stop word* untuk bahasa Inggris diantaranya “yang”, “di”, “ke” (tala, 2003).

2.4 Stemming

Proses stemming digunakan untuk mengubah *term* yang masih melekat dalam *term* tersebut awalan, sisipan, dan akhiran. Proses stemming dilakukan dengan cara menghilangkan semua imbuhan (*affixes*) baik yang terdiri dari awalan (*prefixes*), sisipan (*infixes*), akhiran (*suffixes*) dan *confixes* (kombinasi dari awalan dan akhiran) pada kata turunan. *Stemming* digunakan untuk mengganti bentuk dari suatu kata menjadi kata dasar dari kata tersebut yang sesuai dengan struktur morfologi bahasa indonesia yang benar (Tala, 2003).

2.5 Vector Space Model

Vector Space Model (VSM) 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). 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*.

Bobot istilah yang akhirnya digunakan untuk menghitung tingkat kesamaan antara setiap dokumen yang tersimpan dalam urutan yang memiliki kemiripan, vektor memperhitungkan pertimbangan dokumen yang relevan dengan permintaan *user*. Hasilnya adalah himpunan dokumen yang terambil jauh lebih akurat (sesuai dengan informasi yang dibutuhkan) (Baeza, 1999).

Perhitungan Jarak query menggunakan persamaan dan dokumen, menggunakan persamaan.

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

Dengan $|q|$ adalah Jarak query, dan $W_{i,q}$ adalah bobot query dokumen ke-i, maka Jarak query ($|q|$) dihitung untuk didapatkan jarak query dari bobot query dokumen ($W_{i,q}$) yang terambil oleh sistem. Jarak query bisa dihitung dengan persamaan akar jumlah kuadrat dari query.

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

Dengan $|dj|$ adalah Jarak dokumen, dan $W_{i,j}$ adalah bobot dokumen ke-i, maka Jarak dokumen ($|dj|$) dihitung untuk didapatkan jarak dokumen dari bobot dokumen dokumen ($W_{i,j}$) yang terambil oleh sistem. Jarak dokumen bisa dihitung dengan persamaan akar jumlah kuadrat dari dokumen.

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

Dengan W_{ij} adalah bobot term dalam dokumen, W_{iq} adalah bobot query, dan $\text{Sim}(q, d_j)$ adalah similaritas antara query dan dokumen. Similaritas antar query dan dokumen atau inner product/ $\text{Sim}(q, d_j)$ 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.

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

Similaritas antara query dan dokumen atau $\text{Sim}(q,d_j)$ berbanding lurus terhadap jumlah bobot query (q) dikali bobot dokumen (dj) dan berbanding terbalik terhadap akar jumlah kuadrat q (|q|) dikali akar jumlah kuadrat dokumen (|dj|). 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

Koleksi dokumen pada VSM direpresentasikan sebagai sebuah matrik *term* dokumen (matrik *term frequency*). Setiap sel dalam matrik bersesuaian dengan bobot yang diberikan dari suatu *term* dalam dokumen yang ditentukan. Nilai nol berarti bahwa *term* tersebut tidak ada dalam dokumen.

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

dengan *idfi* adalah inverse document frequency, N adalah jumlah dokumen yang terambil oleh sistem, dan *dfi* adalah banyaknya dokumen dalam koleksi dimana term i di dalamnya, maka perhitungan *idfi* digunakan untuk mengetahui banyaknya term yang dicari (*dfi*) yang muncul dalam dokumen lain yang ada pada database (korpus).

2.7 Recall dan Precision

Hasil pencarian Sistem Temu Kembali Informasi ini bisa dinilai dari tingkat *recall* dan *precision* nya. *Precision* dapat dianggap sebagai ukuran ketepatan atau ketelitian, sedangkan *recall* adalah kesempurnaan. Nilai *precision* adalah proporsi dokumen yang terambil oleh sistem adalah relevan. Nilai *recall* adalah proporsi dokumen relevan yang terambil oleh sistem (Salton, 1989). Nilai *precision* bernilai anatar 0 sd 1. Sistem temu kembali informasi diharapkan untuk dapat memberikan nilai *recall* dan *Precision* mendekati 1. Pengguna rata-rata ingin mencapai nilai *recall* tinggi dan *prececision* tinggi, pada kenyataanya hal ini harus dikompromikan karena sulit dicapai (Salton, 1989).

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

dengan R adalah recall, maka bilai R didapatkan dengan membandingkan Total number of relevan items in collection. Recall adalah dokumen yang terpanggil dari STKI sesuai dengan permintaan user yang mengikuti pola dari STKI.

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

dengan P adalah precision. Nilai P didapatkan dengan membandingkan number of relevan items retrieved dengan Total number of items retrieved. Precision adalah jumlah dokumen yang terpanggil dari database relevan setelah dinilai user dengan informasi yang dibutuhkan.

2.8 Kisah Nabi

Menurut kamus besar bahasa indonesia, Kisah adalah cerita tentang perjalanan atau pelayaran seseorang dari suatu tempat ke tempat lain. Kisah juga cerita tentang kejadian dalam kehidupan, riwayat,

dongeng. Menurut kamus besar Indonesia, Nabi adalah orang yang menjadi pilihan Allah untuk menerima wahyu-Nya. Kisah Nabi adalah cerita tentang perjalanan orang yang menjadi pilihan Allah untuk menerima wahyu-Nya.

3. METODE PENELITIAN

Beberapa metode penelitian yang digunakan sistem temu kembali ini antara lain:

3.1 Objek Penelitian

Melakukan objek penelitian perlu menentukan objek penelitian terlebih dahulu. Hal tersebut dilakukan untuk mempermudah kegiatan objek penelitian yang dibahas pada penelitian ini adalah sekumpulan dokumen-dokumen cerita Nabi yang akan diolah nantinya. Dokumen-dokumen yang akan diolah merupakan dokumen yang ada di internet dan buku.

3.2 Metode Pengumpulan Data

Melakukan objek penelitian perlu menentukan objek penelitian terlebih dahulu. Hal tersebut dilakukan untuk mempermudah kegiatan objek penelitian yang dibahas pada penelitian ini adalah sekumpulan dokumen-dokumen cerita Nabi yang akan diolah nantinya. Dokumen-dokumen yang akan diolah merupakan dokumen yang ada di internet dan buku.

1. Sumber Data Primer

Metode yang digunakan dalam pengumpulan data secara langsung terhadap objek penelitian sebagai berikut :

a. Wawancara

Metode pengumpulan data dengan wawancara dilakukan dengan melakukan percakapan tanya jawab dengan narasumber. Dalam penelitian ini narasumber yang dimaksud adalah para orang tua beserta anak-anaknya untuk mendapatkan data. Data dapat berupa pemahaman cerita Nabi dari sudut pandang anak-anak dan orang tua.

b. Observasi

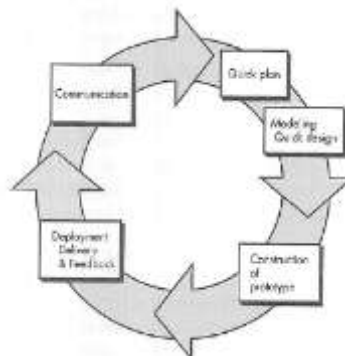
Peneliti melakukan pengamatan secara langsung terhadap beberapa orang tua untuk menceritakan cerita Nabi terhadap anaknya biar bisa mencontoh kisah Nabi tersebut sehingga dapat dijadikan data.

2. Sumber Data Sekunder

Metode pengumpulan data dengan studi kepustakaan, dilakukan dengan mencari cerita Nabi yang berupa buku-buku atau melalui media internet yang berkaitan dengan penelitian.

3.3 Metode Pengembangan Sistem

Metode Pengembangan Sistem yang digunakan 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.



Gambar 1. *Evolutionary Prototyping* (Pressman, 2005)

Adapun tahapan pembuatan perangkat lunak dengan prototype adalah sebagai berikut:

1. 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.

2. 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.

3. 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.

4. 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.

5. Penyerahan Kepada User dan Umpan Balik (Development Delivery dan 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.

a. Pengumpulan Kebutuhan

Kebutuhan yang diperlukan dalam pembuatan sistem ini adalah mengumpulkan dokumen-dokumen cerita nabi yang akan diolah nantinya. Dokumen-dokumen yang akan diolah merupakan dokumen yang diambil dari buku-buku dan website. Modul-modul yang digunakan adalah beranda/halaman utama, pencarian, about, daftar.

b. Perancangan sistem

Setelah kebutuhan sudah ditentukan, maka dilakukan perancangan sistem informasi cerita nabi. Perancangan antarmuka sistem menggunakan mysql, dan membuat sistem menggunakan bahasa pemrograman php.

4. HASIL DAN PEMBAHASAN

4.1 Analisis Sistem

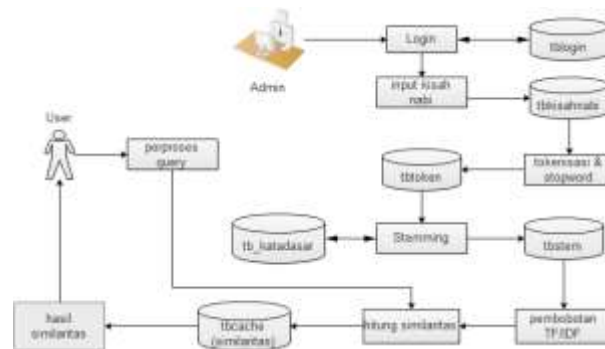
Analisis sistem dalam membangun aplikasi sistem temu kembali informasi memiliki tahapan-tahapan proses yang berurutan. Proses pertama adalah proses *tokenizing*, proses ini memecah dokumen berdasarkan karakter spasi sehingga diperoleh daftar kata-kata yang ada didalamnya. Setelah proses *tokenizing* selanjutnya proses *stopword*, proses ini mengambil kata-kata penting dari hasil token dan kemudian disaring dengan membuang kata-kata yang ada di daftar *stopword*. Kata-kata yang tersisa itu kemudian dihilangkan imbuhanannya melalui proses *stemming* sehingga didapatkan daftar kata dasar dari dokumen tersebut. Dokumen-dokumen yang telah diproses kemudian diurutkan berdasarkan bobot kemiripan masing-masing dokumen dengan *query user*.

Model proses temu kembali informasi yang penulis gunakan adalah Vektor Space Model. Dokumen dan query pada vektor space model, dipresentasikan sebagai vektor dalam ruang vektor yang disusun dalam indeks term, kemudian dimodelkan dengan persamaan geometri. Ada beberapa yang perlu diperhatikan pada model ruang vektor yaitu :

1. Menggunakan bobot index term.
2. Adanya vektor dokumen dari query.
3. Perhitungan cosine menentukan kesamaan dokumen query.

4.2 Arsitektur sistem

Arsitektur sistem digunakan untuk memperlihatkan langkah-langkah membangun sebuah aplikasi sistem temu kembali informasi dan menjelaskan alur kerja sistem secara sederhana dengan gambar yang sederhana yang berupa tahapan-tahapan preprocessing pada sistem temu kembali yang berhubungan antara setiap prosesnya.



Gambar 2. Arsitektur Sistem Temu Kembali Informasi

Arsitektur sistem yang ada pada gambar di atas dapat dijelaskan bawah pertama-tama admin diharuskan login terlebih dahulu. Jika proses login sudah dilakukan, maka proses selanjutnya admin menginputkan nama nabi yang disimpan dalam tabel *tbkisahnabi*. Data di tokenisasi dan dibuang stopwordsnya, disimpan ke dalam tabel *token*. Dilanjutkan dengan proses *stemming* hingga kata yang telah ditoken menjadi bentuk kata dasar, yang disimpan dalam tabel *tbstem*. Dilanjutkan dengan perhitungan bobot *keyword* setiap kata berdasarkan frekuensi, jumlah dokumen dan banyaknya dokumen dalam koleksi dimana *term* muncul di dalamnya

4.3 Perancangan *Unified Modelling Language* (UML)

Perancangan sistem dengan menggunakan *Unified Modeling Language* (UML) berfungsi untuk membuat model semua jenis aplikasi piranti lunak, dimana aplikasi tersebut dapat berjalan pada piranti keras, sistem operasi dan jaringan apapun yang ditulis dalam bahasa pemrograman. UML mendefinisikan bermacam-macam diagram. Diagram yang akan digunakan penulis untuk sistem temu kembali informasi kisah nabi adalah *use case* diagram dan *class* diagram.

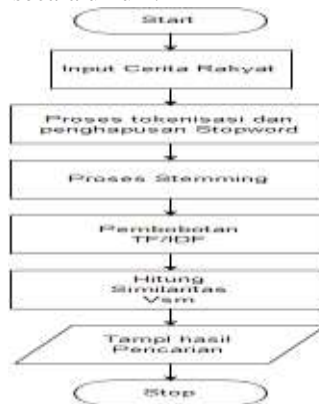
Use case diagram pada aplikasi ini dijelaskan ada dua actor yakni aktor admin dan aktor user. Sebelum masuk ke dalam menu admin, aktor admin diharuskan login terlebih dahulu. Setelah melalui proses login, admin dapat melakukan beberapa proses login, admin dapat melakukan beberapa proses diantaranya adalah menambah menu kisah nabi baru, menghapus kisah nabi yang sudah ada, melihat hasil tokenisasi, melihat hasil *stemming*, *keyword* dan bobot dari *keyword* tersebut, melihat vektor setiap dokumen dan melihat nilai kemiripan antara suatu *query* dengan dokumen-dokumen yang ada. Aktor user dapat langsung melakukan pencarian kisah nabi dengan cara memasukkan kata kunci yang diinginkan oleh user. Setelah itu user dapat mencari data atau dokumen kisah nabi tersebut sesuai dengan kata kunci yang telah dimasukkan oleh user sebelumnya. Gambar.3 Menunjukkan *use case* diagram pada sistem temu kembali informasi kisah nabi dengan menggunakan *Vector Space Model*(VSM).



Gambar 3. Use Case Diagram

4.4 Rancangan Flowchart

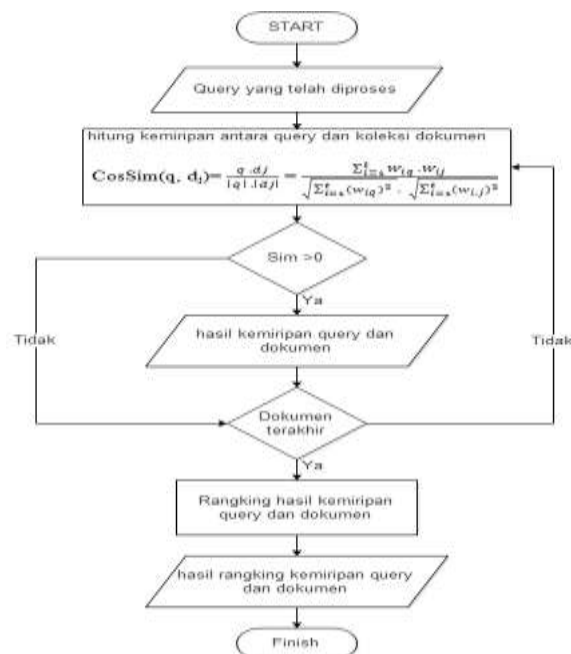
Flowchart menggambarkan bagian alur kerja dari proses sistem temu kembali informasi secara mendetail dan hubungan antar proses satu dengan proses yang lainnya. menggambarkan tentang alur kerja dari suatu sistem temu kembali informasi secara umum.



Gambar 4. Flowchart Sistem Temu Kembali Informasi

Dari flowchart gambar 4. dapat diterangkan, proses sistem temu kembali informasi dimulai dengan input *query* oleh *user*, kemudian *query* tersebut akan diteruskan dengan proses tokenisasi dan penghapusan stopword yaitu proses penghapusan tanda baca yang ada di dalam suatu dokumen, setelah melalui proses tokenisasi, dokumen akan dilakukan filtering yaitu pembuangan kata konjungsi dan kata yang tidak memiliki arti, diteruskan dengan *stemming* yaitu proses penghapusan imbuhan untuk mendapatkan kata dasar dari sebuah kata, diteruskan pembobotan *tf/idf* yaitu bobot yang diberikan dari suatu *term* dalam dokumen yang ditentukan, diteruskan menghitung similaritas VSM akan menampilkan hasil pencarian.

Sedangkan flowchart untuk menghitung VSM dapat dilihat pada gambar 5. Pada jelaskan proses hitung VSM yang merupakan proses untuk menghitung tingkat kedekatan *term* dengan cara pembobotan *term*. Pada prinsipnya perangkan pencarian dokumen ini didasarkan pada perhitungan jarak kemiripan dengan *cosinus* antara vektor dokumen dengan vektor *query*.



Gambar 5. Flowchart Menghitung VSM

4.5 Rancangan Database

Database Sistem Temu Kembali Informasi Kisah Nabi dibuat dengan nama kisahnabi yang di dalamnya terdapat beberapa tabel, diantaranya tbkisahnabi, tblogin, tbtokan, tbstem, tbkeyword, tbvektor, tbcache dan tb_katadasar. Adapun rancangan tabel dapat dijelaskan berikut ini :

1. Tabel Kisah Nabi

Tabel ini berfungsi untuk menyimpan kumpulan kisah nabi. Tabel 1. menunjukkan *field*, tipe data dan ukuran. Dari tabel 1. dapat dijelaskan bahwa Id dalam table *field* digunakan untuk penomoran database kisah nabi, Id_kisahnabi berisikan identitas dari kisah nabi, Judul digunakan untuk penulisan judul dari kisah nabi. Kisah nabi digunakan untuk potongan isi kisah nabi, Url digunakan untuk memasukan link yang terkait dari kisah nabi yang dicari.

Tabel 1. Tabel kisahnabi

FIELD	TIPE DATA	UKURAN
ID	INT	5
ID_KISAHNABI	VARCHAR	100
JUDUL	VARCHAR	200
KISAHNABI	TEXT	-
URL	TEXT	-

2. Tabel Token

Tabel ini untuk menyimpan kata-kata yang telah dipecah dari dokumen yang telah dipecah dari dokumen yang dibuat. Tabel 2. menunjukkan *field*, tipe data dan ukuran.

Dari tabel tbtokan ini dapat diterangkan bahwa Id dalam tbtokan yang digunakan untuk penomoran dari database, Id_kisahnabi digunakan untuk identitas dari judul kisah nabi, DocId digunakan untuk identitas nomor urut dari suatu dokumen dan frekuensi digunakan untuk mengetahui jumlah suatu kata tertentu yang ada dalam suatu dokumen.

Tabel 2. Tabel Token

FIELD	TIPE DATA	UKURAN
ID	INT	5
ID_KISAHNABI	VARCHAR	100
DOCID	INT	11
KATA	VARCHAR	30
FREKUENSI	INT	10

3. Tabel Stemming

Tabel ini untuk menyimpan kata-kata yang telah dipotong imbuhanannya sehingga kata-kata tersebut menjadi sebuah kata dasar. Tabel 3. menunjukkan *field*, tipe data dan ukuran. Dari tabel tbstem dibawah ini dapat dijelaskan bahwa Id dalam tbtokan digunakan untuk penomoran dari database, Id_kisahnabi digunakan untuk identitas dari judul kisah nabi, DocId digunakan untuk identitas nomor urut dari dokumen, *term* digunakan untuk menampilkan kata sebelum dilakukan proses *stemming* menghilangkan imbuhan yang ada dalam suatu kata agar menjadi kata dasar dan kata digunakan untuk menampilkan kata dari suatu kata setelah mengalami proses *stemming*.

Tabel 3. Tabel Stemming

FIELD	TIPE DATA	UKURAN
ID	INT	5
ID_KISAHNABI	VARCHAR	100
DOCID	INT	11
TERM	VARCHAR	30
KATA	VARCHAR	30

4. Tabel Keyword

Tabel ini untuk menyimpan kata dan kemudian menyimpan bobot setiap kata yang telah dilakukan proses token dan *stemming*. Tabel 4. menunjukkan *field*, tipe data dan ukuran. Tabel tbkeyword dapat diketahui bahwa Id dalam tbkeyword digunakan untuk penomoran dari database, Id_kisahnabi digunakan untuk identitas dari judul kisah nabi, DocId digunakan untuk identitas nomor urut dari suatu dokumen, kata digunakan untuk menampilkan kata setelah dilakukan proses tokenisasi dan *stemming*, frekuensi digunakan untuk mengetahui banyaknya

jumlah kemunculan suatu kata dalam suatu dokumen dan bobot digunakan untuk mengetahui nilai dari suatu kata dari tabel *keyword*.

Tabel 4. Tabel *Keyword*

FIELD	TIPE DATA	UKURAN
ID	INT	5
ID_KISAHNABI	VARCHAR	100
DOCID	INT	11
KATA	VARCHAR	30
FREKUENSI	INT	10
BOBOT	FLOAT	

5. Tabel Vektor

Tabel ini untuk menyimpan panjang vektor setiap dokumennya. Dari Tabel 5. dapat dijelaskan dimana DocId untuk identitas nomor urut dari dokumen dan *field* panjang untuk menampilkan hasil perhitungan vektor dari dokumen.

Tabel 5. Tabel Vektor

FIELD	TIPE DATA	UKURAN
DOCID	INT	15
PANJANG	FLOAT	

6. Tabel Cache

Tabel ini untuk menyimpan nilai kemiripan antara suatu *quer* dengan dokumen-dokumen yang ada. Tabel 6. menunjukkan *field*, tipe data dan ukuran. Tabel tbcache dapat dijelaskan Id untuk penomoran dari *query* inputan *user*, *query* digunakan untuk menampung inputan *user*, DocId digunakan untuk Id dokumen dan *value* digunakan untuk menampilkan hasil perhitungan dari kecocokan dokumen yang dicari oleh *user*.

Tabel 6. Tabel Cache

FIELD	TIPE DATA	UKURAN
ID	INT	5
QUERY	VARCHAR	100
DOCID	INT	11
VALUE	FLOAT	

7. Tabel Kata Dasar

Tabel ini untuk melakukan pencocokkan kata, jika kata yang di *input* sama dengan kata yang ada pada tabel *kt_dasar* maka tidak perlu dilakukan proses *stemming*. Tabel 7. menunjukkan *field*, tipe data, ukuran keterangan dari *tb_katadasar*.

Tabel 7. Tabel Kata Dasar

FIELD	TIPE DATA	UKURAN
ID_KTDSR	INT	10
KATADASR	VARCHAR	70
TIPE_KATADASAR	VARCHAR	25

4.6 Pengujian Sistem

1. Pengujian Sistem Dengan Satu *term*

Hasil penelitian dengan memasukkan kata kunci dengan term pada form yang tersedia menunjukkan bahwa sistem dapat berjalan dengan baik. Pengujian sistem dengan satu term untuk hasil pencarian “adam” didapatkan dokumen yang terambil sebanyak 27 dokumen kisah nabi.

Tabel 8. Pengujian satu term

No.	Nama Dokumen	No dokumen	Bobot
1.	Nabi Adam as dalam AlQuran	3	0.263955
2.	Adam A s ver 4	7	0.247776
3.	Adam A s ver 2	5	0.234098
4.	Adam A s ver 5	8	0.218372
5.	Adam A s ver 7	10	0.21257
6.	Adam A s ver 1	1	0.180159
8.	Adam A s ver 6	9	0.179309
9.	Perjalanan nabi Adam dan Hawa	4	0.178825
10.	Nabi Adam dan Siti Hawa pertama kali diciptakan	2	0.173405

2. Pengujian Sistem Dengan Dua *Term*

Hasil penelitian dengan memasukkan kata kunci dua term pada form yang tersedia menunjukkan bahwa sistem dapat berjalan. menunjukkan pengujian sistem dua term. Pengujian sistem dengan dua term untuk hasil pencarian “adam hawa” didapatkan dokumen yang terambil sebanyak 35 dokumen kisah nabi.

Tabel 9. Pengujian dua term

No.	Nama Dokumen	No dokumen	Bobot
1	Nabi Adam as dalam AlQuran	3	0.65222
2	Adam A s ver 5	8	0.5998
3	Adam A s ver 1	1	0.587105
4	Adam A s ver 2	5	0.550473
5	Perjalanan nabi Adam dan Hawa	4	0.529038
6	Adam A s ver 7	10	0.495394
8	Nabi Adam dan Siti Hawa pertama kali diciptakan	2	0.463913
9	Adam A s ver 4	7	0.446516
10	Adam A s ver 3	6	0.319721

3. Pengujian Sistem Dengan Tiga *Term*

Hasil penelitian dengan memasukkan kata kunci tiga term pada form yang tersedia menunjukkan bahwa sistem dapat berjalan. menunjukkan pengujian sistem tiga term. Pengujian sistem dengan tiga term untuk hasil pencarian “adam buah khuldi” didapatkan dokumen yang terambil sebanyak 46 dokumen kisah nabi.

Tabel 10. Pengujian tiga term

No.	Nama Dokumen	No dokumen	Bobot
1.	Adam A s ver 1	1	0.325067
2.	Adam A s ver 2	5	0.302315
3.	Adam A s ver 5	8	0.271756
4.	Nabi Adam as dalam AlQuran	3	0.241194
5.	Adam A s ver 4	7	0.211731
6.	Nabi Adam dan Siti Hawa pertama kali diciptakan	2	0.189945
8.	Adam A s ver 7	10	0.186765
9.	Perjalanan nabi Adam dan Hawa	4	0.178303
10.	Adam A s ver 6	9	0.165724

4. Pengujian Sistem Dengan Lebih Dari Tiga *Term*

Hasil penelitian dengan memasukkan kata kunci lebih dari tiga term pada form yang tersedia menunjukkan bahwa sistem dapat berjalan. menunjukkan pengujian sistem lebih dari tiga term. Pengujian sistem dengan tiga term untuk hasil pencarian “adam hawa buah khuldi” didapatkan dokumen yang terambil sebanyak 49 dokumen kisah nabi.

Tabel 11. Pengujian lebih dari tiga term

No.	Nama Dokumen	No dokumen	Bobot
1	Adam A s ver 1	1	0.491839
2	Adam A s ver 2	5	0.41301
3	Nabi Adam as dalam AlQuran	3	0.412151
4	Adam A s ver 5	8	0.406764
5	Perjalanan nabi Adam dan Hawa	4	0.329604
6	Nabi Adam dan Siti Hawa pertama kali diciptakan	2	0.298397
8	Adam A s ver 7	10	0.290682
9	Adam A s ver 4	7	0.27249
10	Adam A s ver 3	6	0.19796

5. Pengujian Sistem Dengan Keyword Tidak Relevan

Hasil penelitian dengan memasukkan kata kunci tidak relevan pada tabel tbkeyword menunjukkan bahwa sistem dapat berjalan.

6. Pengujian *Recal* Dan *Precision*

Studi kasus Sistem Temu Kembali Informasi (Information Retrieval) Pada Kisah Nabi dengan memasukan 1 term(Iblis, Qurban, Pasletina) untuk 2 term (Malaikat Maut, Kaum Aad, kaum Luth) selanjutnya 3 tem (Idris Pergi Surga, Hawa Makan Buah, Tulang Rusuk Adam) hasil pengujian recall dan precision

Tabel 12. Recall Dan Precision

No	Keyword	Recall Dan Precision							Ri	F
		Terambil	Relevan	Tidak Relevan	Total Dokumen	Relevan		Dok. Terambil		
						A	B			
1	Bila	12	11	1	251	11	211	12	0.12	0.97
2	Adam	27	24	3	251	24	251	27	0.10	0.89
3	Melakukan	74	20	54	251	20	251	79	0.08	0.27
4	Kafir	37	18	19	251	18	251	37	0.07	0.48
5	Abdial	132	57	115	251	57	251	132	0.15	0.24
6	manusia	51	50	1	251	50	251	51	0.20	0.88
7	diri peng	98	30	68	251	30	251	98	0.12	0.31
8	Hawa	113	33	78	251	33	251	113	0.09	0.08
9	Air	79	9	70	251	9	251	79	0.02	0.05
10	Raja	138	88	90	251	88	251	138	0.27	0.43

Nilai batas yang digunakan adalah 0.02 apabila nilai kemiripan dokumen yang terambil diatas 0.02 maka dokumen tersebut dapat dinyatakan relevan. Sebaliknya, apabila nilai kemiripan dokumen dibawah 0.02 maka dokumen yang terambil dianggap tidak relevan. Nilai rata-rata dari recall adalah 0.116 dan nilai dari precision adalah 0.469.

5. KESIMPULAN

Adapun kesimpulan pada penelitian ini, diantar adalah :

1. Hasil catatan pengujian waktu proses pencarian dokumen dari sistem temu kembali informasi kisah nabi adalah, pengujian dengan satu keyword membutuhkan waktu 0.22 detik. Pengujian dengan dua keyword membutuhkan 09.19 detik. Pengujian dengan tiga keyword membutuhkan 09.92 detik. Pengujian dengan lebih dari tiga keyword membutuhkan 11.49 detik.
2. Hasil dari rata-rata pengujian nilai *recall* didapat 0.116 dan nilai rata-rata dari *precision* 0.469 dari perbandingan antara nilai *recall* dan *precision* menunjukkan angka *recall* yang lebih rendah dari *precision*.

DAFTAR PUSTAKA

- [1] Agusta, L. (2009). *perbandingan algoritma stemming porter dengan algoritma Nazief & Adriani untuk Stemming Dokumen Tesk Bahasa Indonesia*, Konferensi Nasional Sistem dan Informatika, pp. 196-201.
- [2] Amin, Fatkhul., (2012). *Sistem Temu Kembali Informasi dengan Metode Vector Space Model*.
- [3] Hariyanto, B. (2004), *Rekaya Sistem Berorientasi Objek*, Cetak Pertama, Informati Bandung, Bandung.
- [4] Madcoms. (2011). *aplikasi web database dengan dreamweaver dan PHP-MySQL*, Andi, yogyakarta.
- [5] Munawar (2005), *pemodelan visual*, Graha Ilmu, Yogyakarta.
- [6] Nugroho, B. (2004), *PHP dan MySQL dengan Editor Dreamweaver MX*, Edisi pertama, Andi, Yogyakarta.
- [7] Pressman, (2005). *Software Engineering A Practicioners Approach*.
- [8] Salton, G. (1989), *Automatic Text Processing, The Transformation, Analysis, and Retrieval of information by computer*. Addison- Wesley Publishing Company, Inc. USA.
- [9] Raharja, D. T. Y. (2015). *Sistem Temu Kembali Informasi Dengan Metode Vector Space Model (Studi Kasus Abstraksi Skripsi Fti Unisbank)* (Doctoral dissertation, UNIVERSITAS STIKUBANK).
- [10] Sutarman, s.kom (2003). *Membangun aplikasi web dengan ip dan MySQL*. Graha Ilmu, Yogyakarta.
- [11] Tala, F.Z. (2003). *A Study of Stemming Effect on Information Retrieval in bahasa indonesia*, Institut for logic, Language and Computation Universiteit van Amsterdam the Netherlands.
- [12] Yates, R.B (1999), *Modern Information Retrieval*, addison Wesley-Pearson International editor, Boston, USA.