

SKRIPSI

**STEGANOGRAFI BERBASIS SUKU KATA BAHASA
INDONESIA**



MICHAEL THEODORE PANGESTU

NPM: 2012730048

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2015**

UNDERGRADUATE THESIS

INDONESIAN SYLLABLE-BASED STEGANOGRAPHY



MICHAEL THEODORE PANGESTU

NPM: 2012730048

**DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2015**

LEMBAR PENGESAHAN

**STEGANOGRAFI BERBASIS SUKU KATA BAHASA
INDONESIA**

MICHAEL THEODORE PANGESTU

NPM: 2012730048

Bandung, «tanggal» «bulan» 2015

Menyetujui,

Pembimbing Utama

Pembimbing Pendamping

**Thomas Anung Basuki, Ph.D.
Ketua Tim Penguji**

**«pembimbing pendamping/2»
Anggota Tim Penguji**

«penguji 1»

«penguji 2»

Mengetahui,

Ketua Program Studi

Thomas Anung Basuki, Ph.D.

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

STEGANOGRAFI BERBASIS SUKU KATA BAHASA INDONESIA

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal «tanggal» «bulan» 2015

Meterai

Michael Theodore Pangestu
NPM: 2012730048

ABSTRAK

«Tuliskan abstrak anda di sini, dalam bahasa Indonesia» Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Kata-kata kunci: «Tuliskan di sini kata-kata kunci yang anda gunakan, dalam bahasa Indonesia»

ABSTRACT

«Tuliskan abstrak anda di sini, dalam bahasa Inggris» Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Keywords: «Tuliskan di sini kata-kata kunci yang anda gunakan, dalam bahasa Inggris»

«kepada siapa anda mempersembahkan skripsi ini...?»

KATA PENGANTAR

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Bandung, «bulan» 2015

Penulis

DAFTAR ISI

| | |
|---|--------------|
| KATA PENGANTAR | xv |
| DAFTAR ISI | xvii |
| DAFTAR GAMBAR | xviii |
| DAFTAR TABEL | xix |
| 1 PENDAHULUAN | 3 |
| 1.1 Latar Belakang | 3 |
| 1.2 Rumusan Masalah | 4 |
| 1.3 Tujuan | 4 |
| 1.4 Deskripsi Perangkat Lunak | 4 |
| 1.5 Metode Penelitian | 5 |
| 2 LANDASAN TEORI | 7 |
| 2.1 Bahasa Indonesia | 7 |
| 2.1.1 Pemenggalan kata | 7 |
| 2.2 Pengolahan Teks | 8 |
| 2.2.1 Natural Language Processing | 8 |
| 2.2.2 Information Retrieval | 9 |
| 2.2.3 Teori Automata | 10 |
| 2.2.4 Teori Finite State Automata (FSA) | 11 |
| 2.2.5 Teori Deterministic FSA (DFSA) | 11 |
| 2.3 Steganografi | 12 |
| 2.3.1 Format Based Steganography[6] | 12 |
| 2.3.2 Linguistic Steganography | 14 |
| 3 ANALISIS | 15 |
| 3.1 Hipotesis | 16 |
| 3.1.1 Pemenggalan Kata | 16 |
| 3.1.2 <i>Database</i> sinonim | 21 |
| 3.1.3 <i>Stego-cover</i> | 22 |
| 3.1.4 <i>Embedding</i> | 22 |
| 3.1.5 <i>Extracting Information</i> | 24 |
| 3.2 Analisis Use Case | 24 |
| 3.2.1 Skenario Use Case | 25 |
| 3.2.2 Analisis Diagram Kelas | 26 |
| 3.2.3 Analisis Diagram Aktivitas | 27 |
| DAFTAR REFERENSI | 31 |

DAFTAR GAMBAR

| | | |
|-----|---|----|
| 2.1 | Kalimat awal (S0) | 13 |
| 2.2 | Kalimat setelah melalui proses <i>line-shift coding</i> (S1) | 13 |
| 2.3 | Hasil penumpukkan kalimat S0 dan S1 | 13 |
| 3.1 | Diagram Transisi DFSA tahap 1[1] | 17 |
| 3.2 | Diagram Transisi DFSA tahap 2[1] | 19 |
| 3.3 | Diagram Transisi DFSA tahap 3[1] | 20 |
| 3.4 | Use Case Diagram Perangkat Lunak Steganografi | 25 |
| 3.5 | Class Diagram perangkat lunak steganografi | 26 |
| 3.6 | Activity Diagram perangkat lunak steganografi saat menyisipkan | 28 |
| 3.7 | Activity Diagram perangkat lunak steganografi saat proses ekstraksi | 29 |

DAFTAR TABEL

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Tidak dapat dipungkiri bahwa teknologi membawa perubahan besar pada seluruh umat manusia di dunia. Dengan adanya internet yang merupakan salah satu produk dari kemajuan teknologi, jarak dan waktu bukanlah menjadi suatu masalah. Komunikasi antar benua, bahkan antar belahan dunia pun bukanlah masalah. Internet juga memudahkan pencarian segala informasi yang mulanya sulit ditemukan.

Namun segala sesuatu yang memiliki sisi positif, pasti ada sisi negatifnya. Informasi yang berserakan di internet dapat dilihat oleh siapa pun. Semua orang yang beraktivitas di internet dapat mengklaim dirinya sebagai siapa pun yang dikehendaki. Tidak terkecuali juga informasi pribadi yang seharusnya tidak diketahui sembarang orang, akhirnya dapat dibaca di internet. Hal ini memungkinkan orang yang tidak dikenali, dapat mengetahui nama lengkap, nama-nama kerabat, sampai alamat rumah yang seharusnya menjadi data pribadi dan tidak diketahui oleh sembarang orang. Semua informasi yang ada di internet merupakan informasi yang sengaja disebarkan atau pun yang secara tidak sadar ikut tersebar di internet.

Kriptografi merupakan salah satu cara yang dapat digunakan untuk menjaga kerahasiaan agar tidak sembarang orang dapat mengetahui suatu informasi. Dalam kriptografi ada dua istilah yang dikenal dengan enkripsi dan dekripsi, yang berarti menyandikan pesan dan memunculkan pesan asli. Proses enkripsi¹ merupakan konversi data ke dalam bentuk lain yang disebut *ciphertext*. *Ciphertext* tidak dapat dimengerti dengan mudah oleh semua orang, kecuali pihak yang berwenang. Dekripsi merupakan proses pemunculan kembali pesan asli dari *ciphertext*. Sebagian besar *ciphertext* merupakan tulisan yang tidak memiliki arti, sehingga hal ini akan dengan mudah menimbulkan kecurigaan pihak lain bahwa ada suatu informasi yang sebenarnya disembunyikan di dalam *ciphertext* tersebut.

Selain kriptografi, steganografi juga dapat digunakan untuk menjaga kerahasiaan informasi. Steganografi mengatasi permasalahan timbulnya kecurigaan pihak tidak berkepentingan pada saat membaca *ciphertext* yang biasanya merupakan kumpulan huruf dan angka yang tidak memiliki arti. Untuk orang awam yang melihat *ciphertext* mungkin tidak berarti apa-apa. Berbeda halnya jika yang melihat adalah orang yang mengerti kriptografi, dengan sendirinya akan diketahui bahwa ada informasi yang sengaja disembunyikan. Steganografi lebih berfokus pada penyembunyian informasi/pesan rahasia. Dengan menyembunyikan pesan rahasia di dalam pesan lain yang tidak bersifat

¹<http://searchsecurity.techtarget.com/definition/encryption>

rahasia, kecurigaan dari pihak tidak berkepentingan pun dapat diatasi.

Ada banyak alasan mengapa steganografi dibutuhkan. [2]Orang-orang yang ingin konseling tentang masalah yang sangat pribadi akan merasa aman. Polisi dapat berkomunikasi dengan *undercover agents* untuk menyergap komplotan orang jahat. Informasi pribadi terjaga dari eksploitasi teroris dan orang jahat.

Steganografi dapat menggunakan berbagai media sebagai media penyembunyian atau yang dikenal dengan istilah *stego-cover*, seperti contohnya media berupa gambar, video, suara, atau teks. Kebanyakan orang memakai gambar sebagai media, karena lebih mudah dalam teknik penyembunyiannya. Media teks masih belum banyak penggunaannya, karena harus memperhatikan cara penulisan, arti dari kalimat, dan lain-lain. Walaupun masih belum banyak yang menggunakannya, media teks merupakan media yang baik untuk menyembunyikan pesan karena kapasitasnya yang cukup besar. Ada berbagai cara menyembunyikan pesan di dalam teks, termasuk memanfaatkan bentuk dari teks itu sendiri (*format-based*) dan memanfaatkan susunan kalimat (*syntax-based*), tetapi belum banyak yang memanfaatkan suku kata.

1.2 Rumusan Masalah

Berikut beberapa rumusan masalah yang akan dibahas:

- Algoritma penyisipan mana yang tepat untuk diimplementasikan dengan memanfaatkan suku kata?
- Bagaimana cara implementasi algoritma pada perangkat lunak?
- Bagaimana performa yang dihasilkan perangkat lunak?

1.3 Tujuan

Berikut merupakan tujuan dari topik skripsi yang saya pilih:

- Menemukan algoritma yang mencakup seluruh/hampir seluruh kemungkinan (optimal).
- Menentukan seperti apa gambaran perangkat lunak yang akan dibuat.
- Mendapatkan performa yang baik dari perangkat lunak.

1.4 Deskripsi Perangkat Lunak

Perangkat lunak akhir yang akan dibuat memiliki fitur minimal sebagai berikut:

- Pengguna dapat menyisipkan pesan rahasianya ke dalam pesan lain yang tidak bersifat rahasia.
- Perangkat lunak akan menghasilkan dokumen stego setelah selesai melakukan proses penyisipan.
- Perangkat lunak dapat membuka pesan rahasia yang ada di dalam dokumen stego.

1.5 Metode Penelitian

Berikut adalah metode penelitian yang digunakan dalam pembuatan skripsi ini:

1. Melakukan studi literatur tentang pengolahan teks.
2. Melakukan studi literatur tentang pemenggalan kata pada Bahasa Indonesia dan jenis-jenisnya.
3. Mengumpulkan teks Bahasa Indonesia untuk dijadikan corpus.
4. Menganalisis kemunculan pola suku kata dari dokumen corpus.
5. Merancang algoritma pembangkitan cover untuk pesan rahasia.
6. Mengimplementasikan rancangan algoritma.
7. Melakukan pengujian.
8. Membuat dokumentasi skripsi.

BAB 2

LANDASAN TEORI

Dari topik yang ada, akan dipilah-pilah menjadi beberapa komponen penyusun. Pada judul tertera kata steganografi, maka pada bab ini akan dibahas mengenai dasar-dasar steganografi, istilah-istilah yang akan digunakan, dan beberapa contoh dari steganografi. Begitu pula dengan suku kata bahasa Indonesia, tentunya akan ada pengantar terlebih dahulu tentang bahasa Indonesia sebelum mengenalkan suku katanya. Selain unsur-unsur yang nampak jelas pada topik, ada juga beberapa elemen yang tidak nampak tetapi akan dibutuhkan, seperti beberapa teori tentang pemotongan kata, bahasa alami, dan pencarian temu yang tergabung menjadi pengolahan teks.

2.1 Bahasa Indonesia

[3] Bahasa Indonesia merupakan bahasa pemersatu Negara Indonesia. Dalam bahasa Indonesia dikenal bahasa tulisan dan bahasa lisan. Pada bahasa lisan dikenal istilah *fonem* yang merupakan satuan bahasa terkecil yang dapat membedakan arti. Dalam bahasa tulisan, *fonem* dilambangkan dengan huruf (alfabet). *Fonem* dibagi menjadi dua, yaitu vokal dan konsonan. Dalam bahasa Indonesia dikenal 5 vokal yaitu a, i, u, e, o, dan 25 konsonan yaitu b, c, d, f, g, h, j, k, l, m, n, p, q, r, s, t, v, w, x, y, z, kh, ng, ny, sy. Konsonan kh, ng, ny, dan sy merupakan contoh fonem yang terdiri atas dua huruf. Ada pula istilah diftong, yaitu gabungan 2 vokal yang membentuk kesatuan bunyi, di antaranya adalah au, ai, oi.

Bahasa Indonesia ditulis dengan menggunakan abjad. Abjad dalam bahasa Indonesia ada 52 huruf, yaitu 26 huruf (a sampai z) dan 26 huruf kapital (A sampai Z). Bahasa Indonesia juga memiliki 10 simbol untuk angka yaitu angka 0 sampai 9.

2.1.1 Pemenggalan kata

Pemenggalan kata menghasilkan suku kata. Setiap kata pada Bahasa Indonesia memiliki suku kata yang berbeda-beda. Berikut pengelompokkan berdasarkan jumlah suku katanya:

1. Terdiri dari satu suku kata, contoh: cat, bor, bom, dll.
2. Terdiri dari dua suku kata, contoh: pa-gi, ka-mu, dll.
3. Terdiri dari tiga suku kata, contoh: me-re-ka, ke-ma-ri, dll.
4. Terdiri dari empat suku kata, contoh: tu-na-wis-ma, da-sa-war-sa, dll.
5. Terdiri dari lima suku kata, contoh: pra-mu-ni-a-ga, dar-ma-wi-sa-ta, dll.

Bahasa Indonesia juga mengenal beberapa pola umum suku kata yaitu:

1. Pola V (vokal), contoh: **a**-nak, **i**-kan
2. Pola KV (konsonan vokal), contoh: **sa**-ku, **se**-la-ma
3. Pola VK, contoh: **an**-da, **am**-pun
4. Pola KVK, contoh: **lak**-sa-na, pe-**rak**
5. Pola KKV, contoh: **pra**-mu-ga-ri, **sas**-tra
6. Pola KKKV, contoh: ben-**trok**, kon-**trak**
7. Pola VKK, contoh: **ons**, **eks**
8. Pola KVKK, contoh: **pers**, kon-**teks**
9. Pola KKKVKK, contoh: kom-**pleks**
10. Pola KKKV, contoh: in-**stru**-men
11. Pola KKKVK, contoh: **struk**-tur

Untuk melakukan pemenggalan suku kata, telah dibuat aturan pemenggalan atau penyukuan kata:

1. Jika dua vokal berada di tengah kata, maka pemenggalan di antara dua vokal. Terdapat pengecualian untuk huruf diftong. Huruf diftong ai, au, dan oi tidak pernah dipisahkan sehingga pemenggalan kata tidak dilakukan di antara kedua huruf tersebut, contoh: main dipenggal menjadi ma-in.
2. Jika konsonan diapit dua vokal seperti kata anak, barang, maka pemenggalan sebelum huruf konsonan, contoh: a-nak, ba-rang.
3. Jika dua konsonan berurutan di tengah kata, pemenggalan dilakukan di antara huruf konsonan pertama dan kedua, contoh: sanjung menjadi san-jung.
4. Jika di tengah kata terdapat tiga konsonan atau lebih, pemenggalan dilakukan di antara konsonan pertama dan kedua, contoh: bentrok menjadi ben-trok.

2.2 Pengolahan Teks

2.2.1 Natural Language Processing

Natural Language Processing [4] atau disingkat menjadi NLP merupakan cabang ilmu *Artificial Intelligence* yang berfokus pada pengolahan bahasa alami. Bahasa alami merupakan bahasa yang digunakan manusia dalam berkomunikasi terhadap satu sama lain. Bahasa alami tidak dapat langsung diterima oleh komputer, sehingga harus diproses terlebih dahulu agar komputer dapat mengeksekusi perintah sesuai dengan keinginan *user*. Berikut beberapa area utama NLP:

- *Question Answering Systems (QAS)*

Komputer dapat menjawab pertanyaan dari *user*. Berbeda dengan *search engine* seperti *Google* misalnya yang menunggu masukan *keyword* dari *user*, dengan QAS *user* dapat langsung mengetikkan pertanyaannya dalam bahasa alami dalam berbagai bahasa.

- *Summarization*

Membuat ringkasan dari artikel, dokumen atau email secara otomatis. Dengan *auto-summarization*, *user* akan mendapatkan hasil ringkasan secara langsung. Ringkasan yang didapatkan telah berisi poin-poin penting dari keseluruhan dokumen.

- *Machine Translation*

Area NLP ini dapat menerima masukan berupa bahasa alami dan menghasilkan keluaran berupa terjemahan bahasa alami dalam bahasa lain (Inggris, Jerman, dll).

- *Speech Recognition*

Area yang merupakan cabang dari NLP yang cukup sulit karena mampu mengenali bahasa lisan. Biasanya merespon pada suara berupa perintah atau pertanyaan.

- *Document classification*

Salah satu cabang dari NLP yang paling banyak dipakai. Dapat menentukan dokumen apa termasuk jenis yang mana. Biasanya dipakai untuk filter *spam*, *movie classification*, dll.

NLP juga memiliki tiga aspek utama yaitu sintaks, semantik, dan pragmatik. Sintaks merupakan aturan tata bahasa, semantik menjelaskan arti dari suatu kalimat, dan pragmatik menjelaskan bagaimana pernyataan yang berhubungan dengan dunia. Selain itu juga yang terkait dengan NLP adalah morfologi, yang merupakan pengetahuan tentang kata dan bentukannya.

2.2.2 Information Retrieval

Information Retrieval atau IR merupakan proses mencari, menemukan lalu mengambil dokumen yang relevan dengan informasi yang dicari user. *Search engine* yang populer seperti *Google* merupakan contoh dari sistem IR. Kata kunci yang biasa diketikkan pada *search engine* merupakan *query* yang akan digunakan oleh IR untuk mencari dan ditampilkan pada layar. Berikut merupakan karakteristik dari sistem IR:

- *Corpus*. *Document Corpus* merupakan kumpulan teks dalam jumlah besar yang biasanya digunakan untuk analisis statistik atau untuk disimpan dan diproses lagi. *Document Corpus* memiliki bagian-bagian yang berbeda. Dalam sebuah corpus terdapat judul, sub-judul, dan paragraph.
- *Queries*. Merupakan kata kunci atas apa yang ingin dicari. *Query* bisa berupa kata kunci atau kata - kata yang harus berdekatan.
- *A result set*. Hasil keluaran dari proses IR yang dinilai sebagai dokumen yang relevan dengan *query*.
- *Presentation of the result set*. Menampilkan list judul dari dokumen yang telah diberi ranking.

2.2.3 Teori Automata

Teori automata[5] merupakan teori mengenai mesin-mesin abstrak dan memiliki hubungan yang dekat dengan teori bahasa formal. Dalam tulisan ini akan dipergunakan istilah *automaton* sebagai bentuk tunggal dan *automata* sebagai bentuk jamak. Teori automata adalah teori tentang mesin abstrak yang:

- bekerja secara sekuensial
- menerima input
- menghasilkan output

Pengertian mesin di sini buka hanya mesin elektronis/mekanis, tetapi juga perangkat lunak yang memenuhi ketiga kriteria tersebut. Seperti yang telah disampaikan sebelumnya, teori automata berhubungan erat dengan bahasa formal. Secara umum terdapat dua fungsi automata dalam hubungannya dengan bahasa, yaitu:

- sebagai pengenalan string dari suatu bahasa, dalam kasus ini bahasa sebagai masukan dari automata
- sebagai pengekstrak string dari suatu bahasa, dalam kasus ini bahasa sebagai keluaran dari automata

Hal yang akan ditekankan adalah poin pertama. Untuk mengenali string dari suatu bahasa, akan dimodelkan automaton yang memiliki komponen sebagai berikut:

- pita masukan, untuk menyimpan string masukan yang akan dikenali
- kepala pita, untuk membaca/menulis ke pita masukan
- *Finite State Controller* yang berisi status dan aturan yang mengatur langkah apa yang dilakukan oleh automaton berdasarkan status setiap saat dan simbol masukan yang dibaca oleh kepala pita
- pengingat, yang berfungsi sebagai tempat penyimpanan dan pemrosesan sementara

Setelah membaca string masukan dan melakukan langkah pemrosesan yang dibutuhkan, akan dihasilkan keputusan apakah string dapat dikenali. *Konfigurasi* merupakan suatu mekanisme untuk menggambarkan keadaan mesin pengenalan yang terdiri dari:

- status *Finite State Controller*
- isi pita masukan dan posisi kepalanya
- isi pengingat

Mesin pengenalan disebut *deterministik* bila di setiap konfigurasinya hanya ada satu kemungkinan yang dapat dilakukan mesin.

2.2.4 Teori Finite State Automata (FSA)

Tiap jenis automata memiliki keunikan yang membedakan fungsinya dengan automata lain. Berikut adalah sifat-sifat FSA:

- pita masukan hanya bisa dibaca, berisi string yang berasal dari suatu abjad
- setelah membaca satu simbol pada pita, kepala pita akan bergeser ke posisi simbol berikutnya
- kepala pita tidak bisa mundur
- memiliki sejumlah status, setiap saat FSA berada dalam status tertentu

Setiap FSA bisa diasosiasikan dengan sebuah diagram transisi berupa graf berarah. Setiap simpulnya mewakili tiap status pada FSA. Jika ada transisi dari status a ke b pada input i , maka ada busur dari a ke b dengan label i . Status awal ditandai dengan kata START dan status akhir ditandai dengan dua lingkaran. Jadi cara kerja suatu FSA dapat digambarkan dengan diagram transisi.

2.2.5 Teori Deterministic FSA (DFSA)

Sebelumnya telah disebutkan bahwa automaton dapat bekerja secara deterministik. Setiap bahasa reguler dapat dikenali oleh DFSA. Secara formal DFSA dapat dinyatakan dengan $(Q, \Sigma, \delta, q_0, F)$ di mana

- Q = himpunan berhingga status
- Σ = himpunan berhingga simbol masukan
- δ = fungsi transisi yang memetakan $Q \times \Sigma$ ke Q
- q_0 = status awal $q_0 \in Q$
- F = himpunan status akhir, $F \subseteq Q$

Cara kerja:

- pertama DFSA akan berada pada status q_0 , kepala pita berada pada simbol pertama pita,
- lalu kepala pita akan membaca simbol dari pita dan bergeser maju,
- untuk tiap simbol, DFSA akan berpindah status sesuai fungsi δ ,
- proses berakhir jika simbol masukan pita sudah habis,
- jika di akhir proses didapatkan status akhir, maka string masukan diterima dan bila tidak, maka string masukan ditolak.

2.3 Steganografi

Kriptografi[2] tradisional berhasil mengamankan pesan secara matematis. Aman secara matematis berarti pesan yang ada bisa saja dibuka, tetapi membutuhkan waktu yang sangat lama. Ketika pesan berhasil dibuka, informasinya sudah tidak lagi bermanfaat. Steganografi menyembunyikan informasi sehingga tidak dapat ditemukan. Kedua hal tersebut memiliki tujuan yang sama, namun melewati proses yang berbeda. Pada steganografi proses penyembunyian dan ekstraksi kembali informasi yang disembunyikan dikenal dengan istilah *embedding* dan *extracting information*. Sementara untuk informasi rahasia yang akan disembunyikan dikenal dengan istilah *secret*. Proses *embedding* memerlukan *secret* dan *stego-cover* sebagai media penyembunyiannya. Hasil dari proses *embedding* disebut dengan *stego-object*.

Untuk menyembunyikan informasi tentunya dibutuhkan suatu algoritma. Sebagian algoritma menyembunyikan informasi dengan menggunakan kunci untuk mengatur bagaimana penyisipan informasi. Sebagian lagi menyembunyikan informasi tidak menggunakan kunci, sehingga proses ekstraksinya pun tidak memerlukan kunci. Hal ini mirip dengan kriptografi, walaupun dilakukan dengan cara menyembunyikan informasi.

Steganografi teks menggunakan teks sebagai media untuk penyisipan (*embded*) informasi. Steganografi teks merupakan jenis yang paling jarang digunakan. Hal ini disebabkan sulitnya menyembunyikan teks dalam jumlah besar pada teks lain sebagai medianya. Walaupun demikian, teks sebagai media memiliki kapasitas lebih besar daripada suara, gambar, atau video sebagai medianya karena memiliki ukuran yang lebih kecil dan dapat menampung lebih banyak informasi. Steganografi teks dibedakan menjadi dua kategori, yaitu steganografi dengan memanfaatkan ilmu bahasa (*linguistic steganography*) dan steganografi yang memanfaatkan bentuk visual dari huruf (*format based steganography*).

2.3.1 Format Based Steganography[6]

Metode ini menggunakan format fisik dari teks seperti spasi, tinggi huruf, dan format fisik lainnya yang dapat dimanfaatkan untuk menyisipkan informasi. Metode ini secara umum mengubah teks yang ada untuk disembunyikan pesan rahasia. Spasi atau karakter yang tidak ditampilkan, kesalahan penulisan yang disengaja disembunyikan pada teks, mengubah ukuran tulisan merupakan contoh dari *format-based steganography*. Beberapa cara di atas seperti kesalahan penulisan dapat mengecoh pembaca yang menganggapnya wajar, namun jika memiliki dokumen asli, dapat dilakukan perbandingan antara dokumen yang asli dan dokumen hasil steganografi (*stego-object*), sehingga bagian-bagian yang diganti dapat terlihat.

The Word-Shift Coding

Mirip dengan *line-shift coding*, teknik ini menggeser kata secara horizontal untuk memberikan sebuah tanda. Pada ¹ diberikan contoh bahwa suatu teks dapat mengandung teks lain di dalamnya dengan menggunakan *line-shift coding*. Diberikan suatu kalimat S0:

¹<http://www.jjtc.com/stegdoc/sec202.html>


```
We explore new steganographic and cryptographic
algorithms and techniques throughout the world to
produce wide variety and security in the electronic web
called the Internet.
```

Gambar 2.1: Kalimat awal (S0)

lalu setelah disembunyikan suatu pesan dengan menggunakan *line-shift coding*, muncul kalimat S1.

```
We explore new steganographic and cryptographic
algorithms and techniques throughout the world to
produce wide variety and security in the electronic web
called the Internet.
```

Gambar 2.2: Kalimat setelah melalui proses *line-shift coding* (S1)

Dengan menumpukkan kedua kalimat S0 dan S1, didapatkan hasil sebagai berikut:

```
We explore new steganographic and cryptographic
algorithms and techniques throughout the world to
produce wide variety and security in the electronic web
called the Internet.
```

Gambar 2.3: Hasil penumpukkan kalimat S0 dan S1

Hasil ini dapat dicapai dengan melebarkan spasi sebelum kata *explore*, *the*, *wide*, dan *web*, juga menyempitkan spasi setelah kata *explore*, *world*, *wide*, dan *web* pada S1. Dengan demikian, kalimat yang mengandung *shifted-word* akan tetap memiliki arti yang sama.

Feature Coding

Feature coding memanfaatkan ciri yang ada dalam teks yang dapat diubah. Seperti garis lurus yang ada pada huruf b, d, h, k, dan lain-lain. Panjang dari garis tersebut dapat diubah tanpa membuat curiga pembaca pada umumnya. Dapat juga dilakukan substitusi pada beberapa kata dengan kata sinonimnya. Dengan itu kita dapat menyisipkan informasi 0 pada sinonim pertama dan 1 pada sinonim kedua. Cara ini harus melalui kesepakatan antara dua belah pihak tentang pasangan sinonimnya.

Inter-sentence Spacing

Cara ini akan menyandikan teks yang sudah dalam bentuk biner ke dalam teks dengan menyisipkan satu atau dua spasi setelah tiap kalimat. Misalkan satu spasi menyandikan 0 dan dua spasi menyandikan 1. Sayangnya cara ini memiliki beberapa masalah, yang pertama jika ingin menyisipkan teks yang memiliki ukuran yang besar, akan membutuhkan teks cover yang banyak pula karena 1 kalimat hanya dapat menyandikan 1 bit. Kedua, beberapa perangkat pengolah kata secara otomatis mengubah banyaknya spasi menjadi satu.

End of Line Spacing

Cara ini menyisipkan spasi pada akhir baris. Cara ini meningkatkan banyaknya informasi yang dapat disandikan dibandingkan cara sebelumnya. Sama seperti sebelumnya, cara ini dapat terganggu oleh beberapa program yang dapat menghapus spasi yang berlebihan dari teks dan cara ini tidak dapat menampilkan pesan rahasia dari *hard copy*.

Inter-word Spacing

Cara ketiga menggunakan spasi untuk menyandikan data yang menggunakan *right-justification*. Data disandikan dengan mengatur di mana spasi ekstra akan diletakkan. Satu spasi di antara kata menyandikan 0 dan dua spasi menyandikan 1. Cara ini memungkinkan beberapa bit dapat disembunyikan dalam satu baris.

2.3.2 Linguistic Steganography

Metode ini memanfaatkan *natural language processing* atau pemrosesan bahasa alami yang menyebabkan pesan dapat disembunyikan tanpa mengubah bahasa asli. Contohnya bisa dengan melakukan substitusi terhadap sinonim yang termasuk steganografi leksikal. Perubahan dari kalimat aktif ke pasif atau sebaliknya untuk contoh dari sintaksis steganografi. Pemanfaatan suku kata untuk menyisipkan pesan yang juga termasuk ke dalam steganografi linguistik yang akan menjadi bahasan utama dalam skripsi ini.

BAB 3

ANALISIS

Pada bab ini akan dibahas mengenai algoritma yang dipilih untuk menyembunyikan *secret message*. *Stego-cover* yang akan digunakan berupa *text file* yang tentu terdiri dari banyak kata. Algoritma yang dipilih akan memanfaatkan banyaknya suku kata dalam suatu kata. Notasi \sum akan digunakan untuk merepresentasikan banyaknya suku kata dalam suatu kata. Diketahui bahwa \sum dapat bernilai ganjil atau genap. Dua kemungkinan tersebut identik dengan bilangan biner. Dengan demikian diputuskan bahwa \sum yang bernilai genap akan merepresentasikan 0 dan \sum bernilai ganjil merepresentasikan 1. Dapat ditarik kesimpulan berarti 1 kata dapat diubah menjadi 1 bit bilangan biner. Jika sebuah *stego-cover* memiliki 700 kata, berarti *stego-cover* tersebut memiliki kapasitas sebesar 700 bit.

Setelah menemukan cara mengubah *stego-cover* menjadi bilangan biner, *secret message* juga harus diubah menjadi bilangan biner. Diketahui bahwa setiap karakter yang ditampilkan pada layar komputer, memiliki kode ASCII yang unik. Kode ASCII merupakan representasi numerik dari karakter seperti 'a', '1', atau '@'. Kode ASCII merupakan representasi numerik, berarti kode ASCII dapat diubah lagi menjadi bilangan biner. Sebagai contoh, jika sebuah *secret message* terdiri dari n karakter dan tiap karakter diubah menjadi 7 bit bilangan biner dengan memanfaatkan kode ASCII, maka dibutuhkan minimal $7 \cdot n$ bit *stego-cover* untuk dapat menyembunyikan *secret message* tersebut.

Setelah *stego-cover* dan *secret message* dapat diubah menjadi deret bilangan biner, *stego-cover* harus dapat diubah lagi sedemikian rupa agar $7 \cdot n$ deret biner pertamanya identik dengan deret biner *secret message*. Sehingga saat proses ekstraksi, deret biner dari *stego-cover* dapat diubah lagi menjadi karakter yang sesuai dengan *secret message*. Karena \sum dalam *stego-cover* dapat bernilai 0 atau 1, ada kemungkinan \sum tidak sesuai dengan angka biner dari *secret message* yang akan disembunyikan. Untuk mengatasi hal ini, kata dengan nilai \sum yang tidak sesuai, akan diganti dengan sinonim yang memiliki \sum yang sesuai.

Tentunya semua kata yang ada pada *stego-cover* harus memiliki sinonim agar setiap kata dapat memiliki \sum yang bernilai 0 atau 1. Tesaurus¹ merupakan buku referensi berupa daftar kata dengan sinonimnya. Namun karena format tesaurus yang memiliki banyak pengelompokkan kata, seperti kata "bobol" bisa berarti "ambrol"/"hancur", bisa juga berarti "buyar"/"tembus". Lalu dari kata "bobol" bisa jadi "membobol" yang juga bisa berarti "membobok", bisa juga berarti "mencuri". Karena format yang terlalu rumit untuk dapat dibaca secara otomatis oleh program, maka diputuskan untuk membuat *file database* yang lebih ringkas. *File database* ini berisi pasangan kata

¹<http://kbbi.web.id/tesaurus>

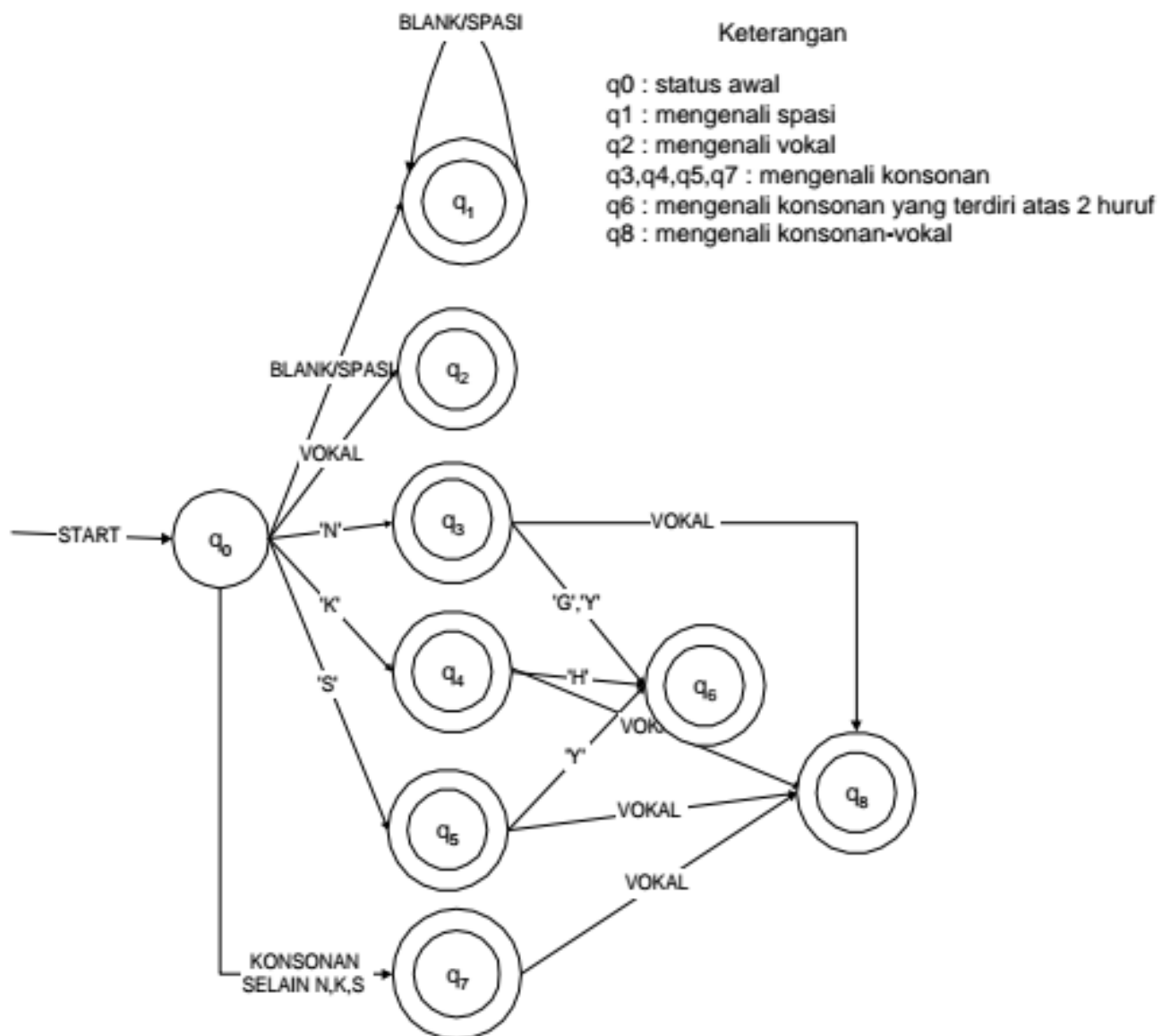
dengan sinonimnya pada tiap barisnya. Karena *file database* dibuat sendiri, maka daftar kata yang ada di dalamnya, masih terbatas, artinya jika pengguna menambahkan *stego-cover* baru, pengguna juga harus melengkapi sinonim yang belum terdaftar pada *file database*.

3.1 Hipotesis

3.1.1 Pemenggalan Kata

Untuk proses pemenggalan kata, akan menggunakan skripsi Frisca Sumarlin 2010 yang berjudul Aplikasi Pendongeng. Untuk mengenali suku kata dalam bahasa Indonesia, akan digunakan tiga tahap DFSA. Setiap masukan dari DFSA merupakan hasil DFSA tahap sebelumnya. Seperti yang telah dijelaskan, FSA memiliki prinsip kerja membaca hanya satu arah dan tidak bisa membaca mundur masukannya. Sedangkan untuk mengenali suku kata, untuk beberapa kasus kata dibutuhkan kemampuan untuk membaca mundur masukannya. Sebagai contoh, pada kata "anda", huruf ketiga adalah huruf konsonan sehingga pemenggalan suku kata dapat dilakukan pada saat membaca huruf ketiga.

Pada tahap 1 ini memang beberapa kata belum sesuai dengan aturan penyukuan kata yang benar, karena itu akan dilakukan tiga tahapan. Untuk diagramnya dapat dilihat pada Gambar [3.1](#).



Gambar 3.1: Diagram Transisi DFSA tahap 1[1]

Penjelasan tentang Diagram Transisi DFSA tahap 1 untuk mengenali spasi, vokal (V), konsonan yang terdiri dari 1 huruf (K), konsonan yang terdiri dari 2 huruf (KK), dan konsonan-vokal (KV) adalah sebagai berikut.

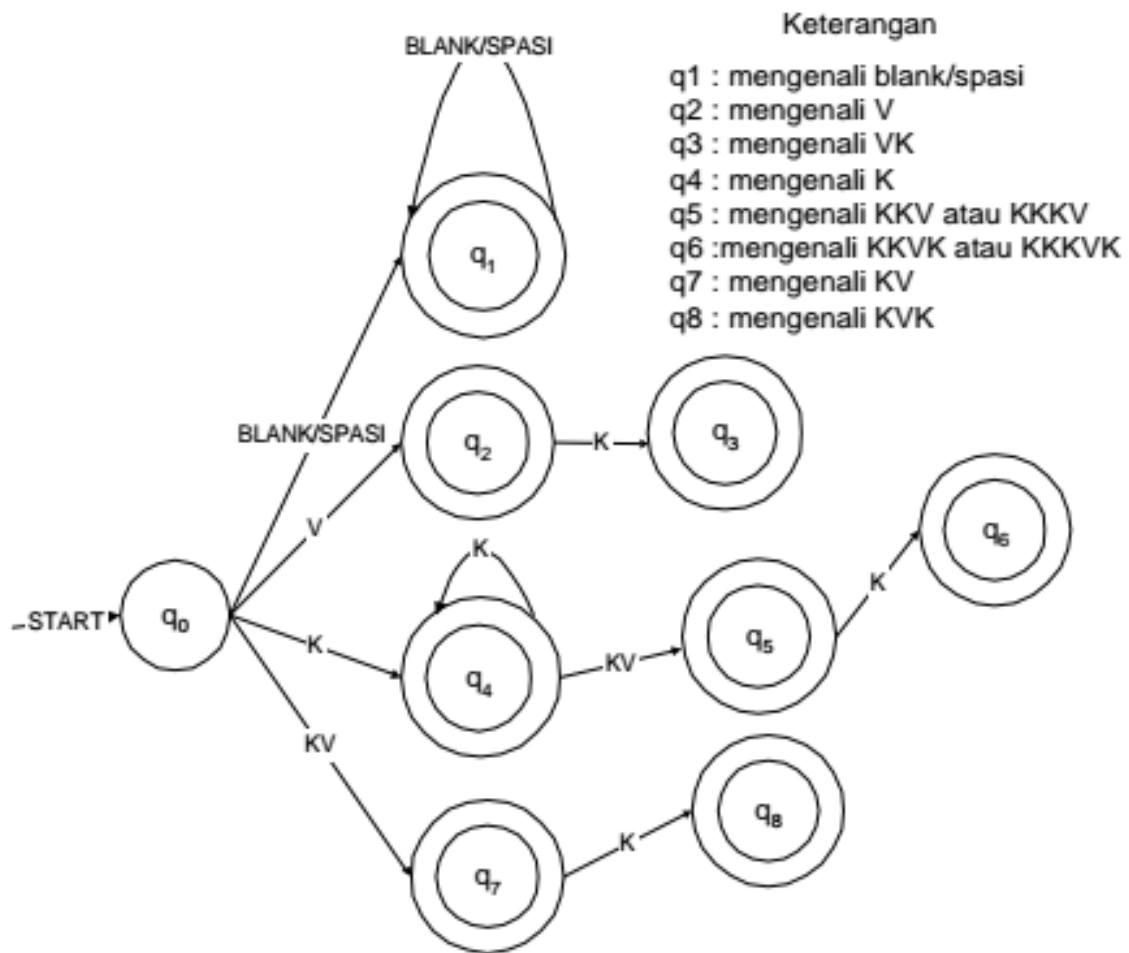
- Perpindahan dari status awal (q_0) ke q_1 adalah untuk mengenali spasi atau *string* kosong.
- Perpindahan dari q_0 ke q_2 adalah untuk mengenali huruf vokal.
- Perpindahan dari q_0 ke q_3 adalah untuk mengenali huruf 'N'.
- Perpindahan dari q_0 ke q_4 adalah untuk mengenali huruf 'K'.
- Perpindahan dari q_0 ke q_5 adalah untuk mengenali huruf 'S'.
- Perpindahan dari q_0 ke q_7 adalah untuk mengenali huruf konsonan selain 'N', 'K', dan 'S'.

- Perpindahan dari q_3 ke q_8 adalah untuk mengenali pola suku kata VK.
- Perpindahan dari q_3 ke q_6 adalah untuk mengenali pola suku kata KK (ng dan ny).
- Perpindahan dari q_3 ke q_8 adalah untuk mengenali pola suku kata KV.
- Perpindahan dari q_4 ke q_6 adalah untuk mengenali pola suku kata KK (kh).
- Perpindahan dari q_4 ke q_8 adalah untuk mengenali pola suku kata KV.
- Perpindahan dari q_5 ke q_6 adalah untuk mengenali pola suku kata KK (sy).
- Perpindahan dari q_5 ke q_8 adalah untuk mengenali pola suku kata KV.
- Perpindahan dari q_7 ke q_8 adalah untuk mengenali pola suku kata KV.

Sebagai contoh, jika melakukan pemenggalan kata hanya dengan menggunakan DFSA tahap 1 saja, kata "gondok" tidak dapat dipotong secara sempurna. Berikut langkah-langkahnya.

- Pada status awal (q_0) huruf ke-1 'g' akan diperiksa.
- Pindah ke q_7 karena huruf 'g' merupakan huruf konsonan selain huruf 'N', 'K', dan 'S'.
- Huruf ke-2 'o' diperiksa dan pindah ke q_8 karena huruf 'o' merupakan huruf vokal. Suku kata ke-1 "go" disimpan, lanjut ke huruf selanjutnya dan mulai lagi dari q_0 .
- Huruf ke-3 'n' diperiksa dan pindah ke q_3 karena huruf ke-3 adalah huruf 'n'.
- Huruf ke-4 'd' diperiksa dan tidak memenuhi syarat untuk ke q_8 ataupun q_6 . Suku kata ke-2 "n" disimpan. Huruf ke-4 'd' kembali ke q_0 .
- Pindah ke q_7 karena huruf 'd' merupakan huruf konsonan selain huruf 'N', 'K', dan 'S'.
- Huruf ke-5 'o' diperiksa dan pindah ke q_8 karena 'o' merupakan huruf vokal. Suku kata ke-3 "do" disimpan, lanjut ke huruf selanjutnya dan mulai lagi dari q_0 .
- Huruf ke-6 'k' diperiksa dan pindah ke q_4 karena huruf ke-6 adalah huruf 'K'. Suku kata ke-4 "k" disimpan. Tahap 1 selesai karena semua huruf telah melewati diagram transisi DFSA tahap 1.
- Didapatkan hasil pemenggalan suku kata "gondok" menjadi "go-n-do-k".

Dapat dilihat bahwa jika hanya menggunakan DFSA tahap 1, hasilnya masih belum sempurna. Pada tahap 1 hanya dikenali pola suku kata V, K, KV, dan KK. Selanjutnya, setelah melewati tahapan 1 DFSA, *outputnya* akan menjadi masukan untuk DFSA tahap 2. Pada tahap 2 ini harus dikenali V, KV, dan pengembangan dari ketiga pola yang dikenali pada tahap pertama. Pada tahap ini, pengembangan yang bisa dilakukan adalah V+K, K+KV, K+KV+K, K+K+KV, K+K+KV+K, K+KV+K, dan KV+K. Untuk digram transisi tahap kedua dapat dilihat pada Gambar 3.2.

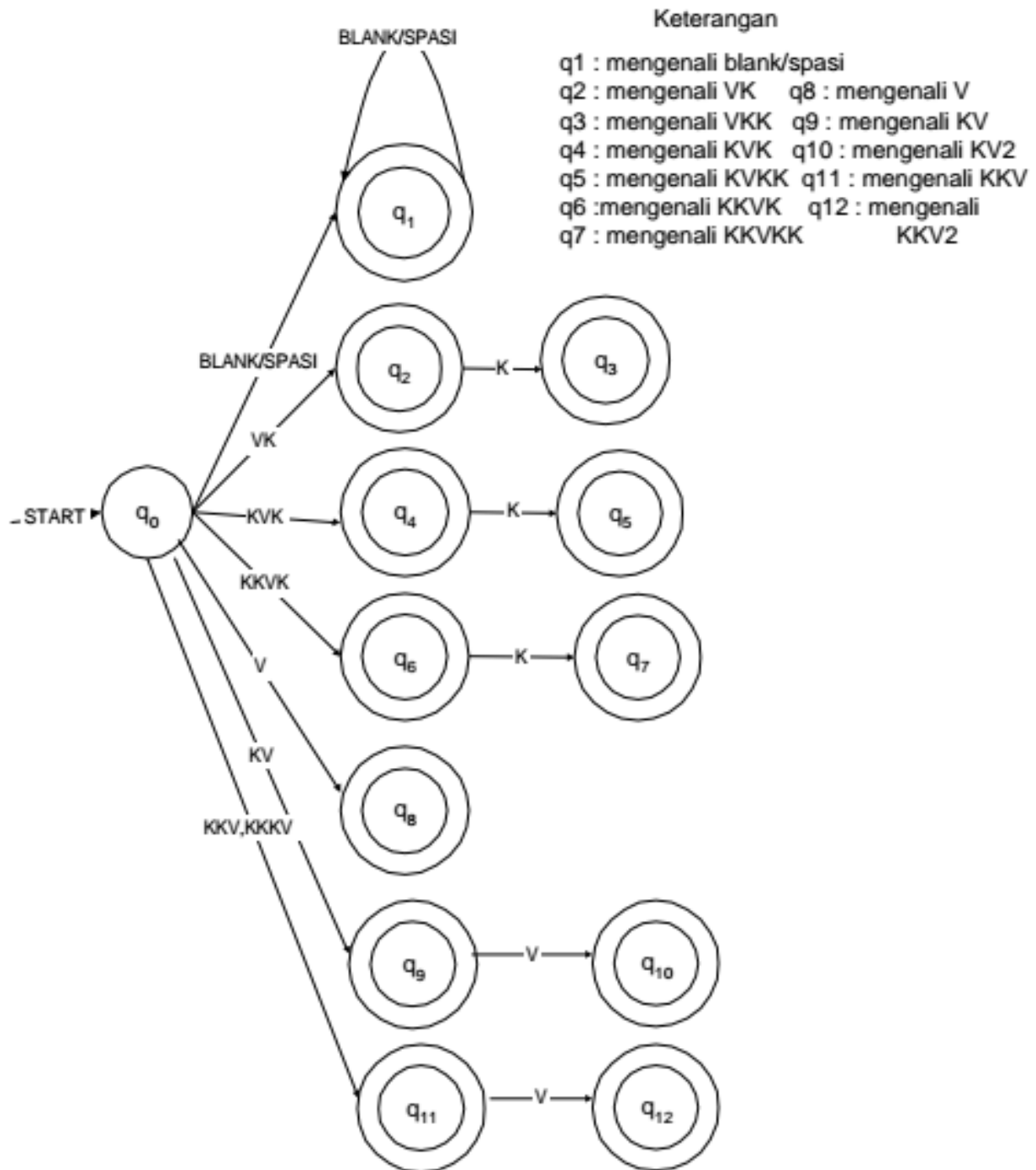


Gambar 3.2: Diagram Transisi DFSA tahap 2[1]

Penjelasan tentang Diagram Transisi DFSA tahap 2 untuk mengenali pola suku kata V, VK, K, KKV, KKKV, KKVK, KKKVK, KV, dan KVK adalah sebagai berikut.

- Perpindahan dari status awal (q_0) ke q_1 adalah untuk mengenali spasi atau *string* kosong.
- Perpindahan dari q_0 ke q_2 adalah untuk mengenali pola suku kata V.
- Perpindahan dari q_0 ke q_4 adalah untuk mengenali pola suku kata K.
- Perpindahan dari q_0 ke q_7 adalah untuk mengenali pola suku kata KV.
- Perpindahan dari q_2 ke q_3 adalah untuk mengenali pola suku kata VK.
- Perpindahan dari q_4 ke q_4 adalah untuk mengenali pola suku kata KK.
- Perpindahan dari q_4 ke q_5 adalah untuk mengenali pola suku kata KKV atau KKKV.
- Perpindahan dari q_5 ke q_6 adalah untuk mengenali pola suku kata KKVK atau KKKVK.
- Perpindahan dari q_7 ke q_8 adalah untuk mengenali pola suku kata KVK.

Pada tahap 2, masih ada tiga pola suku kata yang belum dapat dikenali, yaitu VKK, KVKK, dan KKVKK. Untuk itu masih dibutuhkan satu tahapan lagi untuk dapat mengenali pola-pola tersebut dan mengenali huruf diftong. Pada tahap ini hasil dari DFSA tahap kedua akan menjadi masukan bagi DFSA tahap terakhir yaitu tahap 3.



Gambar 3.3: Diagram Transisi DFSA tahap 3[1]

Penjelasan tentang Diagram Transisi DFSA tahap 3 untuk mengenali pola suku kata VK, V, VKK, KV, KVK, KVV, KVKK, KKV, KKVK, KKVKK, dan KKV2 adalah sebagai berikut.

- Perpindahan dari status awal (q_0) ke q_1 adalah untuk mengenali spasi atau *string* kosong.
- Perpindahan dari q_0 ke q_2 adalah untuk mengenali pola suku kata VK.
- Perpindahan dari q_0 ke q_4 adalah untuk mengenali pola suku kata KVK.
- Perpindahan dari q_0 ke q_6 adalah untuk mengenali pola suku kata KKVK.
- Perpindahan dari q_0 ke q_8 adalah untuk mengenali pola suku kata V.
- Perpindahan dari q_0 ke q_9 adalah untuk mengenali pola suku kata KV.
- Perpindahan dari q_0 ke q_{11} adalah untuk mengenali pola suku kata KKV atau KKKV.
- Perpindahan dari q_2 ke q_3 adalah untuk mengenali pola suku kata VKK.
- Perpindahan dari q_4 ke q_5 adalah untuk mengenali pola suku kata KVKK.
- Perpindahan dari q_6 ke q_7 adalah untuk mengenali pola suku kata KKVKK.
- Perpindahan dari q_9 ke q_{10} adalah untuk mengenali pola suku kata KVV.
- Perpindahan dari q_{11} ke q_{12} adalah untuk mengenali pola suku kata KKVV atau KKKVV.

Untuk dapat membagi suku kata, diperlukan automata yang dapat menerima masukan berupa kata dan keluaran berupa suku kata. *Finite State Transducer* merupakan *finite state machine* yang memiliki dua pita, yaitu pita masukan dan pita keluaran. Automata yang telah dirancang adalah DFSA, di mana keluaran yang dihasilkan adalah dikenali atau tidak dikenali. Dengan merubah keluaran dari DFSA tersebut menjadi suku kata untuk setiap sekuens karakter yang telah dikenali, maka DFSA yang telah dirancang dapat dimanfaatkan untuk membagi kata menjadi suku kata. Atas dasar itu, DFSA dipilih untuk melakukan penyukuan kata.

Pemenggalan suku kata akan dipakai untuk mendapatkan banyak suku kata kata-kata yang terdapat pada *stego-cover*, sehingga dapat dihitung banyak suku kata tiap katanya. Namun karena kumpulan *stego-cover* yang akan dipakai dapat berupa cerita pendek(cerpen) atau puisi, maka isi dari *stego-cover* tidak hanya berupa kata-kata alfabet saja, tetapi juga ada angka yang bisa berupa tanggal, kata serapan, dan lain-lain. Hal ini kemudian dapat menjadi hambatan, karena seperti yang kita ketahui bahwa angka tidak memiliki suku kata. Hambatan lainnya juga datang dari kata-kata dalam bahasa asing dan singkatan.

Setelah menemui hambatan-hambatan di atas, akhirnya diputuskan beberapa solusi. Pertama-tama, dari keseluruhan isi *stego-cover*, tanda baca akan diabaikan (kecuali tanda penghubung '-'), sehingga yang akan dipenggal hanyalah kata-katanya saja. Satu digit angka akan dihitung sebagai satu suku kata. Semua singkatan atau kata dalam bahasa asing, akan dilakukan pemenggalan kata apa adanya.

3.1.2 Database sinonim

Proses *embedding* akan memanfaatkan daftar sinonim yang ada pada *file database*. *File database* kata akan disimpan dalam sebuah *file* dengan ekstensi .txt. Pada *file database* satu per satu kata yang ada pada *stego-cover* memiliki sinonimnya sendiri. Pencarian sinonim pertama-tama akan

dicari di tesaurus. Tesaurus² merupakan buku referensi berupa daftar kata dengan sinonimnya. Kata-kata yang terdaftar di tesaurus adalah kata baku, sehingga jika pada *stego-cover* terdapat kata yang tidak dapat ditemukan pada tesaurus, diperlukan penanganan tersendiri.

Solusinya adalah dengan mengubah-ubah imbuhan pada awal atau akhir kalimat, namun tetap memperhatikan konteks kalimat tersebut. Sebagai contoh, pada *stego-cover* terdapat kata 'menggendongnya', namun kata tersebut tidak dapat ditemukan pada tesaurus, karena terdapat imbuhan. Sehingga untuk sinonimnya dapat digunakan kata 'menggendong'. Kedua kata tersebut memiliki jumlah suku kata yang berbeda, namun memiliki arti yang sama.

Terdapat beberapa kata yang memang tidak memiliki sinonim dengan jumlah suku kata yang berbeda. Kata 'lama' memiliki 2 suku kata, sedangkan sinonim yang ada pada Tesaurus adalah 'lamban' dan 'lelet'. Kedua sinonimnya memiliki \sum genap, tetapi yang dibutuhkan adalah \sum ganjil. Hal ini dapat diatasi dengan menyamakannya sebagai *typo* atau kesalahan pengetikkan. Contohnya dengan memasukkan kata 'lambaan' atau 'lamaa' sebagai sinonim dari kata 'lama'. Kata 'lambaan' dan 'lamaa', jika dipotong suku katanya dengan menggunakan pemotong suku kata yang telah dibuat, kedua kata tersebut memiliki \sum ganjil.

3.1.3 *Stego-cover*

Stego-cover yang telah dikumpulkan berupa puisi dan cerpen. Alasan dipilihnya cerpen dan puisi sebagai *stego-cover*, karena puisi umumnya menggunakan satu kata berulang-ulang, sehingga dapat mengurangi banyaknya kata yang disimpan pada *file database*. Sedangkan cerpen yang dipilih memiliki jumlah kata yang cukup banyak, ini menandakan kapasitas penyimpanan yang cukup besar.

Dengan dipilihnya cerpen dan puisi sebagai *stego-cover*, maka dibutuhkan skema komunikasi yang dapat menyamarkan artikel tersebut agar tidak menimbulkan kecurigaan oleh pihak lain. Skema komunikasi yang dipilih berupa komunikasi antara dua mahasiswa yang gemar mengirim puisi atau cerpen melalui *email*. *Email* yang dikirim mengandung *stego-object* yang merupakan hasil dari *embedding secret message* pada *stego-cover*. Dengan skema komunikasi seperti ini, jika ada pihak lain yang membaca *email* kedua mahasiswa tersebut, tidak akan mencurigai adanya kejanggalan karena kedua mahasiswa gemar berkirim puisi dan cerpen.

3.1.4 *Embedding*

Proses *embedding* akan memanfaatkan suku kata dan sinonim kata itu sendiri. Ide utamanya adalah dengan mengganti kata-kata tertentu yang ada dalam *stego-cover* dengan kata-kata yang telah disediakan pada *file database*, agar $7*n$ deret bilangan biner pertama pada *stego-cover* identik dengan deret bilangan biner *secret message*. *File database* berisi semua kata yang ada pada *stego cover* berpasangan dengan sinonimnya. Pasangan kata yang ada pada *database* merupakan sinonim dengan \sum yang berbeda (ganjil dan genap) dengan kata aslinya.

Penggantian kata pada *stego-cover* ditentukan dari ASCII *secret message* yang akan disembunyikan. ASCII³ (*American Standard Code for Information Interchange*) merupakan format yang paling umum untuk *file* teks yang ada di komputer dan internet. Kode ASCII yang akan dipakai

²<http://kbbi.web.id/tesaurus>

³<http://whatistechtarget.com/definition/ASCII-American-Standard-Code-for-Information-Interchange>

direpresentasikan dengan 7-bit angka biner, yang berarti ada 7 digit angka 0 atau 1. Pemilihan 7-bit ini dikarenakan kode ASCII pada bilangan desimal hanya ada dari 0 sampai 127, sehingga jika menggunakan 8-bit, angka pertama(paling kiri) akan selalu bernilai 0. Dapat disimpulkan bahwa untuk menyisipkan 7-bit dibutuhkan 7 kata, 8-bit dibutuhkan 8 kata. Untuk alasan memaksimalkan kapasitas penyisipan, maka diputuskan untuk merepresentasikan 1 karakter menjadi 7-bit.

Setelah *secret message* diubah menjadi deret bilangan biner dengan memanfaatkan kode ASCII, satu per satu digit kodenya akan dicocokkan dengan \sum pada *stego cover*. Kode digit pertama akan disisipkan pada kata pertama *stego-cover*, digit kedua disisipkan pada kata kedua, dst. Jika \sum pada *stego-cover* tidak sesuai dengan digit kodenya, maka kata tersebut akan diganti dengan sinonimnya yang ada pada *file database*. Pemberitahuan akan muncul jika ada kata yang sinonimnya tidak ditemukan pada *file database*.

Algoritma

Seperti yang telah dijelaskan sebelumnya, algoritma ini akan memanfaatkan kata yang ada pada *stego-cover* untuk menyisipkan kode ASCII dari pesan rahasia ditambah kode ASCII karakter '#' (untuk menandakan bahwa pesan rahasia telah berakhir). Proses penyisipan ini juga memanfaatkan *file database* untuk mengubah kata asli yang ada pada *stego-cover* dengan sinonim yang ada pada *file database* jika \sum tidak sesuai dengan kode ASCII pesan rahasia.

Algoritma *embedding* pesan rahasia pada *stego-cover* dengan memanfaatkan sinonim

1. Meminta input berupa pesan rahasia(*secret message*).
2. Tiap karakter pada *secret message* diubah menjadi kode ASCII 7-bit dan disimpan menjadi *secret code*.
3. Tambahkan deret biner 0100011, yang merupakan kode ASCII untuk karakter '#' untuk menandakan bahwa *secret message* telah berakhir.
4. Jika n adalah banyak karakter yang ada pada *secret message*, maka panjang *secret code* = $7*n+1$ (ditambah karakter '#').
5. Cari *stego cover* yang dapat menampung *secret code*. (Banyaknya kata yang ada dalam *stego cover* \geq panjang *secret code*).
6. Bangkitkan angka acak (r) dari 0 sampai banyaknya *stego cover* yang dapat menampung *secret code* - 1.
7. Buka *file* ke- (r) dan baca per kata dengan mengabaikan tanda baca dan spasi.
8. Untuk $7*n+1$ kata pertama, lakukan pemenggalan kata dan dapatkan nilai \sum .
9. Cocokkan nilai \sum dengan angka ke- n yang ada pada *secret code* secara berurutan. (\sum kata ke- n dicocokkan dengan angka ke- n *secret code*).
10. Jika pada tahap 9 didapatkan \sum yang tidak sesuai dengan angka ke- n *secret code*, lanjut ke tahap 11. Jika sesuai, ulangi tahap 9.

11. Cari sinonim kata tersebut pada *file database*. Kata yang diambil hanya kata dengan nilai \sum yang berbeda.
12. Dari kata-kata yang didapatkan dari tahap 11, akan dilakukan pengambilan satu kata secara acak.
13. Kata yang telah diambil dari tahap 12, akan menggantikan kata yang tidak sesuai sebelumnya.
14. Kembali ke tahap 9.

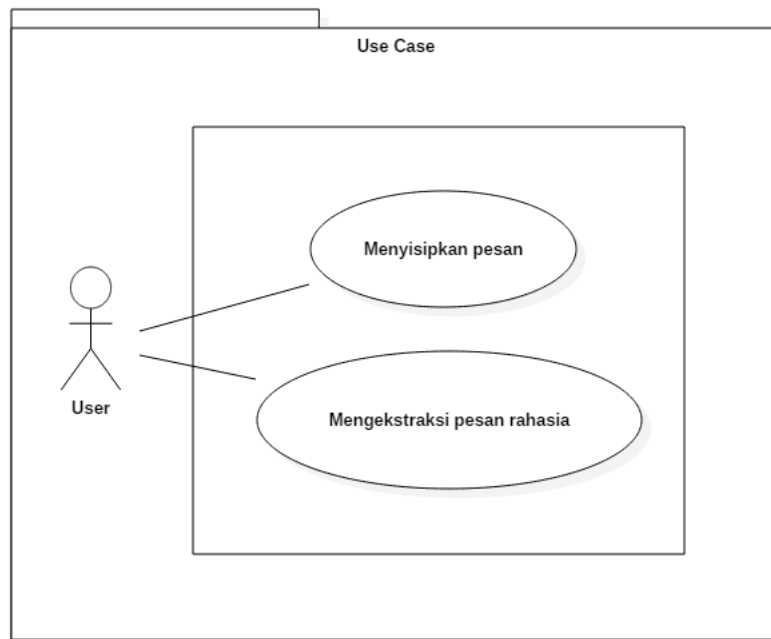
3.1.5 *Extracting Information*

Pada proses pengestraksian kembali pesan rahasia akan dibutuhkan input berupa *stego-object* yang merupakan hasil dari *embedding*. Perangkat lunak pertama-tama akan membaca secara keseluruhan inputnya, lalu melakukan pemenggalan kata sama seperti yang dilakukan pada saat proses *embedding*. Selesai melakukan pemenggalan kata, dari tiap kata akan didapatkan */sumnya*. Saat ini telah didapatkan deretan angka 0 dan 1 yang dinamakan *secret code*. Saat proses *embedding*, memang 1 karakter diubah menjadi kode ASCII 7-bit, namun *java* menyediakan fungsi untuk mengubah deret biner 8-bit menjadi karakter. Sehingga untuk setiap 7-bit pada *secret code* dapat ditambahkan angka 0 di paling kiri sebelum dijadikan *input* pada fungsi *java* tersebut.

Setelah selesai mengubah bit-bit yang ada menjadi karakter, hasilnya akan ditampilkan pada layar. Proses ekstraksi ini melibatkan semua kata yang ada pada *stego-object*, sehingga kata yang tidak disisipkan pesan rahasia pun ikut terekstraksi. Namun karena saat proses *embedding* telah ditambahkan kode ASCII dari karakter '#', maka *secret message* berakhir saat ditemukan karakter '#'.

3.2 Analisis Use Case

Diagram *usecase* perangkat lunak steganografi ini memiliki dua aktor, yaitu pengirim dan penerima. Pengirim berarti aktor yang akan melakukan *embedding* dan mengirimkan *stego-object*, sedangkan penerima berarti aktor yang menerima *stego-object* dan melakukan *extract information* untuk mendapatkan *secret message*. Berikut diagram *usecase* dari perangkat lunak yang akan dibangun.



Gambar 3.4: Use Case Diagram Perangkat Lunak Steganografi

Dari diagram di atas dapat dilihat dua *use case*, yakni:

1. **Menyembunyikan pesan**, pengirim mengetikkan pesan rahasia yang akan disembunyikan.
2. **Menambahkan *stego-cover***, pengirim menyalin teks sebagai *stego-cover*. Jika pengirim menambahkan *stego-cover*, maka pengirim wajib melengkapi sinonim yang belum terdaftar agar *stego-cover* dapat digunakan.
3. **Menambahkan sinonim**, pengirim mengetikkan kata beserta dengan sinonimnya pada masing-masing kotak yang telah disediakan.
4. **Memeriksa semua sinonim**, pengirim dapat memeriksa kata apa saja yang belum memiliki sinonim dengan menekan tombol yang telah disediakan.
5. **Mengekstraksi kembali pesan rahasia**, penerima akan memasukkan *stego-object* yang diterima.

3.2.1 Skenario Use Case

1. Menyisipkan pesan

- Nama: Menyisipkan pesan
- Aktor: Pengguna
- Deskripsi: Mengetikkan pesan rahasia yang akan disisipkan
- Prakondisi: -
- Tujuan: Menyisipkan pesan rahasia yang diketikkan ke dalam salah satu dokumen yang ada di dalam dokumen korpus

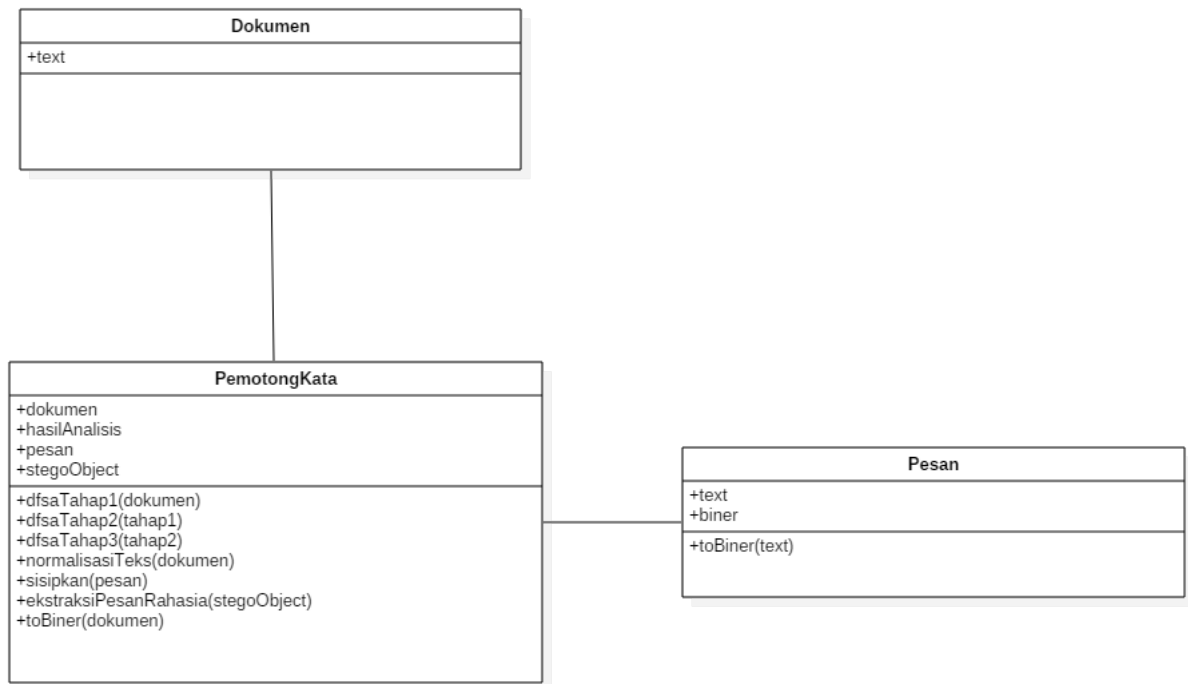
- Skenario:
 - (a) Pengguna mengetikkan pesan rahasia yang akan disisipkan
 - (b) Pengguna menekan tombol sisipkan
 - (c) Sistem lalu akan mengubah pesan yang diketik menjadi kode ASCII
 - (d) Sistem mencari dokumen yang cocok dengan kode ASCII pesan rahasia
 - (e) *Stego-object* sebagai hasil dari penyisipan akan muncul pada layar.

2. Mengekstraksi pesan rahasia

- Nama: Mengekstraksi pesan rahasia
- Aktor: Pengguna
- Deskripsi: Memasukkan *stego-object* yang didapatkan
- Prakondisi: Telah memiliki *stego-object* yang dihasilkan dari perangkat lunak ini
- Tujuan: Mengekstraksi pesan rahasia yang telah disisipkan sebelumnya
- Skenario:
 - (a) Pengguna menyalin *stego-object* yang didapat dari perangkat yang sama
 - (b) Pengguna menekan tombol Ekstrak
 - (c) Sistem akan melakukan penyukuan kata
 - (d) Masing-masing kata akan dijumlahkan suku katanya
 - (e) Sistem menyandikan jumlah suku kata yang genap menjadi 0 dan yang ganjil menjadi 1
 - (f) Sistem akan mengubah deret biner menjadi karakter yang sesuai dengan kode ASCII
 - (g) Pesan rahasia akan ditampilkan pada layar

3.2.2 Analisis Diagram Kelas

Diagram kelas perangkat lunak steganografi dapat dilihat pada [3.5](#)



Gambar 3.5: Class Diagram perangkat lunak steganografi

Keterangan atas diagram kelas akan dijelaskan sebagai berikut:

1. Kelas Dokumen

- Atribut text, untuk menampung teks dari dokumen berupa string

2. Kelas Pesan

- Atribut text, untuk menampung teks dari pesan berupa string
- Atribut biner, untuk menampung teks dari pesan berupa biner, tetapi dalam bentuk string
- Fungsi toBiner, memiliki parameter text. Fungsi ini akan merubah tiap karakter yang ditampung di atribut text menjadi kode ASCII dalam bentuk biner.

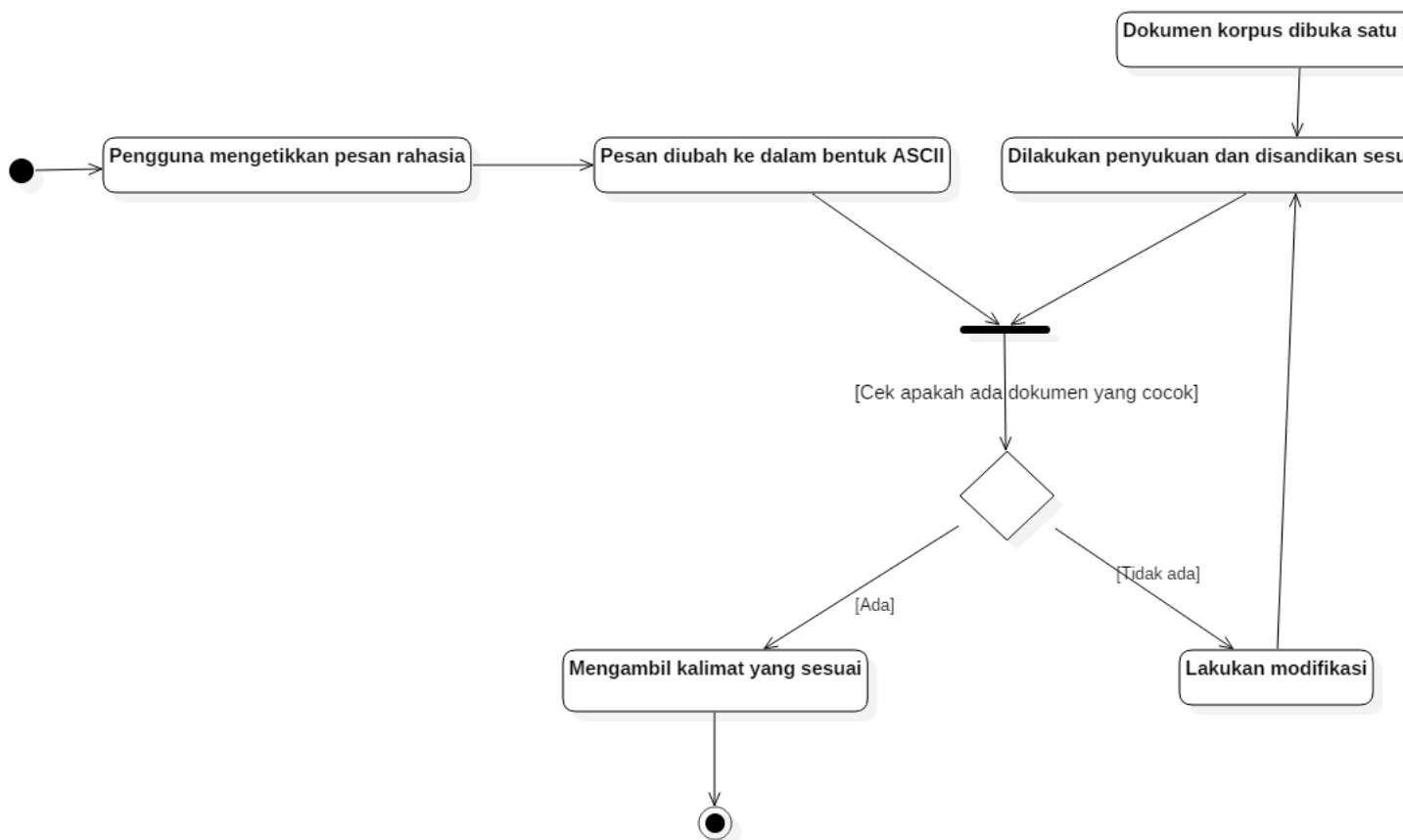
3. Kelas PemotongKata

- Atribut dokumen, untuk menampung isi dokumen
- Atribut hasilAnalisis, untuk menyimpan hasil dari penyukuan kata
- Atribut pesan, untuk menyimpan pesan rahasia yang akan disisipkan
- Atribut stegoObject, untuk menyimpan stego-object yang akan diekstraksi
- Fungsi dfsTahap1, memiliki parameter dokumen menghasilkan keluaran berupa hasil proses dfsa tahap pertama
- Fungsi dfsTahap2, memiliki parameter tahap1 yang merupakan keluaran dfsTahap1 dan mengeluarkan hasil proses dfsa tahap kedua

- Fungsi `dfsTahap3`, memiliki parameter `tahap2` yang merupakan keluaran `dfsTahap2` dan mengeluarkan hasil penyukuan kata dari `text`
- Fungsi `normalisasiText`, memiliki parameter dokumen dan menghasilkan keluaran berupa teks tanpa tanda baca dan simbol, sehingga siap untuk dipotong berdasarkan suku katanya
- Fungsi `sisipkan`, memiliki parameter pesan yang merupakan pesan rahasia yang diketikkan oleh pengguna dan akan mengeksekusi rangkaian penyisipan pesan
- Fungsi `ekstraksiPesanRahasia`, memiliki parameter `stegoObject` yang merupakan *stego-object* yang disalin oleh pengguna dari hasil penyisipan menggunakan perangkat lunak yang sama dan akan mengeksekusi rangkaian pengekstraksian kembali pesan rahasia
- Fungsi `toBiner`, memiliki parameter dokumen. Parameter dokumen merupakan dokumen yang telah dilakukan penyukuan, sehingga fungsi ini hanya menghitung jumlah suku kata untuk setiap katanya. Jumlah kata akan disandikan dengan 0 untuk jumlah genap dan 1 untuk jumlah ganjil

3.2.3 Analisis Diagram Aktivitas

Diagram aktivitas dari perangkat lunak dibagi menjadi dua. Diagram aktivitas saat menyisipkan pesan dapat dilihat pada [3.6](#)

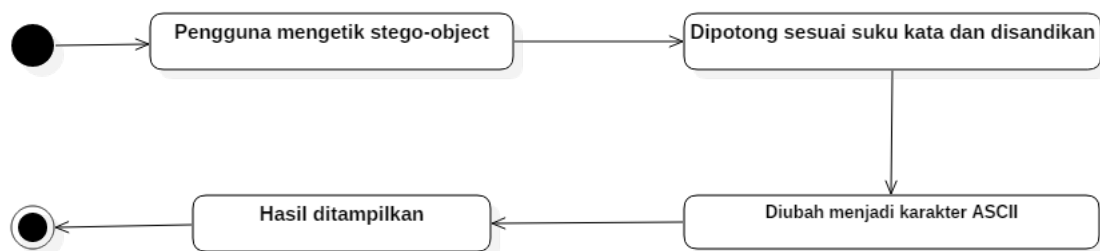


Gambar 3.6: Activity Diagram perangkat lunak steganografi saat menyisipkan

Dari diagram aktivitas dapat dijelaskan sebagai berikut:

1. Pengguna memasukkan input berupa string (pesan rahasia)
2. Pesan akan diubah ke bentuk ASCII
3. Dokumen korpus akan dibuka satu per satu
4. Untuk setiap dokumen yang dibuka akan dilakukan penyukuan kata dan dilakukan penyandian. Angka 1 untuk kata dengan jumlah suku kata genap dan angka 2 untuk kata dengan jumlah suku kata ganjil
5. Pesan rahasia yang telah diubah ke bentuk ASCII dan dokumen yang telah disandikan akan dibandingkan. Dari sini akan dihasilkan dua kemungkinan:
 - Jika ditemukan dokumen cocok dengan pesan rahasia, akan diambil kalimat yang sesuai
 - Jika tidak ditemukan dokumen yang cocok, dari dokumen yang paling mirip akan dilakukan modifikasi (perubahan sinonim atau perubahan kalimat aktif/pasif)
6. Jika telah didapatkan kalimat yang sesuai, kalimat akan ditampilkan dan program selesai

Diagram aktivitas ekstraksi dapat dilihat pada



Gambar 3.7: Activity Diagram perangkat lunak steganografi saat proses ekstraksi

Dari diagram aktivitas di atas dapat dijelaskan sebagai berikut:

1. Pengguna akan memberikan input berupa stego-object dari program yang sama
2. Sistem lalu akan menyukukan kata-kata yang ada. Setelah itu dilakukan penyandian, angka 0 untuk kata dengan jumlah suku kata genap dan angka 1 untuk jumlah ganjil
3. Selesai melakukan penyandian, hasilnya akan diubah menjadi karakter ASCII yang sesuai
4. Kini sudah didapatkan pesan dengan karakter ASCII, hasil akan ditampilkan pada layar

DAFTAR REFERENSI

- [1] T. A. Basuki *Pengenalan Suku Kata Bahasa Indonesia Menggunakan Finite-State Automata*, vol. 5, pp. 67–74, 2000.
- [2] P. Wayner in *Disappearing cryptography: Information hiding: Steganography and watermarking*, vol. 3, 2009.
- [3] Permendiknas in *EYD Terbaru (Sesuai Permendikbud No. 46 Tahun 2009)*, JENDELA PUBLISHING, 2009.
- [4] J. P. . A. Stubbs in *Natural Language Annotation for Machine Learning*, O'Reilly Media, Inc., 2013.
- [5] F. Sumarlin in *Sistem Text-To-Speech Untuk Membacakan Dongeng Berbahasa Indonesia*, 2014.
- [6] K. S. Hitesh Singh, Pradeep Kumar Singh *A Survey on Text Based Steganography*, 2009.