**OTOMATISASI PERINGKASAN TEKS PADA DOKUMEN HUKUM MENGGUNAKAN METODE *LATENT SEMANTIC ANALYSIS***

SKRIPSI

Oleh:

MILLENIA RUSBANDI NIM. 1641720029



PROGRAM STUDI TEKNIK INFORMATIKA

JURUSAN TEKNOLOGI INFORMASI

POLITEKNIK NEGERI MALANG

**2020**

HALAMAN PENGESAHAN

**OTOMATISASI PERINGKASAN TEKS PADA DOKUMEN HUKUM MENGGUNAKAN METODE *LATENT SEMANTIC ANALYSIS***

Disusun oleh:

MILLENIA RUSBANDI NIM. 1641720029

**Proposal Skripsi ini telah diuji pada ………….**

**Disetujui oleh:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. | Penguji I | : | Pramana Yoga S, S.Kom., M.MT.  NIP. 19880504 201504 1 004 | ........................... |
|  |  |  |  |  |
| 2. | Penguji II | : | Dika Rizky Yunianto, S.Kom., M.Kom.  NIP. 19920606 201903 1 017 | ........................... |
|  |  |  |  |  |
| 3. | Pembimbing I | : | Imam Fahrur Rozi, S.T., M.T.  NIP. 19840610 200812 1 004 | ........................... |
| 4. | Pembimbing II | : |  | ........................... |

Mengetahui,

|  |  |
| --- | --- |
| Ketua Jurusan  Teknologi Informasi | Ketua Program Studi  Teknik Informatika |
| Rudy Ariyanto, S.T., M.CS  NIP. 19711110 199903 1 002 | Imam Fahrur Rozi, ST., MT.  NIP. 198406102008121004 |

# KATA PENGANTAR

Puji syukur Alhamdulillah penulis ucapkan kehadirat Allah SWT yang telah melimpahkan rahmat, taufik, dan hidayah-Nya sehingga laporan skripsi yang berjudul “” dapat selesai tepat waktu. Penyusunan laporan skripsi digunakan untuk memenuhi persyaratan kelulusan Program Diploma IV, Jurusan Teknologi Informasi. Program Studi Teknik Informatika, Politeknik Negeri Malang. Penulis menyampaikan terimakasih atas kontribusi dalam penyelesaian skripsi ini. Dengan kerendahan hati penulis menyampaikan terima kasih kepada:

1. Bapak selaku Direktur Politeknik Negeri Malang.
2. Bapak selaku Ketua Jurusan Teknologi Informasi Politeknik Negeri Malang.
3. Bapak selaku Ketua Program Studi Teknik Informatika Politeknik Negeri Malang.
4. Ibu selaku Dosen Pembimbing I Skripsi.
5. Bapak selaku Dosen Pembimbing II Skripsi.
6. Kedua orang tua dan keluarga
7. Rekan-rekan jurusan Teknologi Informasi khususnya Program Studi Teknologi Informasi Politeknik Negeri Malang yang telah banyak membantu.
8. Bapak yang telah membantu pada penelitian sebagai narasumber dan pengolahan data.
9. Seluruh pihak yang telah membantu dan mendukung lancarnya pembuatan laporan akhir dari awal sampai selesai.

# PERNYATAAN KEASLIAN

# DAFTAR ISI

[KATA PENGANTAR 3](#_Toc34990206)

[PERNYATAAN KEASLIAN 4](#_Toc34990207)

[DAFTAR ISI 5](#_Toc34990208)

[DAFTAR TABEL 7](#_Toc34990209)

[DAFTAR GAMBAR 8](#_Toc34990210)

[DAFTAR *SOURCE CODE* 9](#_Toc34990211)

[BAB 1 PENDAHULUAN 10](#_Toc34990212)

[1.1. Latar Belakang 10](#_Toc34990213)

[1.2. Rumusan Masalah 11](#_Toc34990214)

[1.3. Tujuan 11](#_Toc34990215)

[1.4. Batasan Masalah 11](#_Toc34990216)

[1.5. Manfaat 12](#_Toc34990217)

[1.6. Tata Cara Penulisan 12](#_Toc34990218)

[BAB 2 TINJAUAN PUSTAKA 14](#_Toc34990219)

[2.1. Penelitian Terdahulu 14](#_Toc34990220)

[2.2. Tanaman Kopi 16](#_Toc34990221)

[2.3. Sistem Pakar 22](#_Toc34990222)

[2.4. Metode Case Based Reasoning 23](#_Toc34990223)

[BAB 3 METODOLOGI PENELITIAN 26](#_Toc34990224)

[3.1. Studi Literatur 27](#_Toc34990225)

[3.2. Perancangan Sistem 27](#_Toc34990226)

[3.3. Implementasi 27](#_Toc34990227)

[3.4. Pengujian 28](#_Toc34990228)

[3.5. Analisa Hasil dan Kesimpulan 28](#_Toc34990229)

[BAB 4 ANALISA DAN PERANCANGAN 29](#_Toc34990230)

[4.1. Data (tampilkan contoh data, data hanya di tampilkan sebagian sisanya dimasukan ke lampiran, cara pengumpulan data) 29](#_Toc34990231)

[4.2. Kebutuhan Fungsional 29](#_Toc34990232)

[4.3. Kebutuhan Non-fungsional 29](#_Toc34990233)

[4.4. *Use Case* Diagram 29](#_Toc34990234)

[4.5. Use Case Pesifikasi (alur dari tiap use case yang ada di atas) 29](#_Toc34990235)

[4.6. Activity Diagram 29](#_Toc34990236)

[4.7. ERD (di tambah semua metode) 29](#_Toc34990237)

[4.8. Alur Metode Floyd Warshall (gambar flowchart, sub bab masing” dari proses di flowchart, rumus tinggal di panggil menggunakan persamaan dari BAB 2) 29](#_Toc34990238)

[BAB 5 IMPLEMENTASI DAN PENGUJIAN 30](#_Toc34990239)

[5.1. Implementasi *Source Code* 30](#_Toc34990240)

[5.1.1 sub bab use case 30](#_Toc34990241)

[5.2. Implementasi Tampilan (mock up dari aplikasi) 30](#_Toc34990242)

[5.3. Pengujian 30](#_Toc34990243)

[PENUTUP 31](#_Toc34990244)

[6.1. Kesimpulan (menjawab rumusan masalah, selanjtnya kesimpulan setelah pengejian poin poin dalam bentuk paragraf) 31](#_Toc34990245)

[6.2 Saran (kendala) 31](#_Toc34990246)

[DAFTAR PUSTAKA (minimal 10) 32](#_Toc34990247)

[LAMPIRAN 33](#_Toc34990248)

# DAFTAR TABEL

# DAFTAR GAMBAR

[Gambar 0.1 Bagan Metodologi Penelitian 11](file:///D:\File%20Kuliah\Semester%208\filing%20gudd%20lak%20a%20sudd\Laporan%20Skripsi.docx#_Toc34119670)

# DAFTAR *SOURCE CODE*

# BAB 1 PENDAHULUAN

## Latar Belakang

Selama periode tahun 2015–2017, jumlah kasus kriminalitas di Indonesia cenderung fluktuatif (tidak stabil). Berdasarkan data Kepolisian Republik Indonesia (Polri), total jumlah kasus kriminalitas pada tahun 2015 sebanyak 352.936 kasus. Pada tahun 2016, kasus semakin meningkat menjadi sebanyak 357.197 kasus. Sedangkan pada tahun 2017, mengalami penurunan menjadi 336.652 kasus (Badan Pusat Statistik, 2018). Dari data tersebut, rata-rata jumlah kriminalitas periode tahun 2015-2017 sebesar 348.928 kasus. Banyaknya jumlah kriminalitas di Indonesia akan berdampak pada jumlah dokumen hukum yang akan ditangani oleh aparat penegak hukum.

Dalam pemahaman dokumen hukum, aparat penegak hukum seperti advokat, hakim, maupun jaksa harus membaca keseluruhan dokumen dimana akan memerlukan waktu yang cukup lama. Kapasitas yang dimiliki manusia untuk mengonsumsi informasi sangat terbatas. Oleh karena itu, sangatlah penting untuk menyaring informasi yang hanya berguna dari jumlah data tidak terstuktur. Sehingga salah satu solusi yang dibutuhkan yaitu dengan membuat sebuah ringkasan pada dokumen hukum dimana dokumen tersebut berbentuk pdf. Dokumen hukum yang akan diringkas berupa nota pembelaan dan bagian yang diolah yaitu pendahuluan. Dilakukannya peringkasan pada bagian tersebut dikarenakan berbentuk paragraf dan dapat diringkas. Sehingga kalimat-kalimat yang tidak diperlukan akan dihilangkan.

Ringkasan adalah teks yang dihasilkan dari sebuah kumpulan teks yang mengandung informasi penting dengan panjang teks tidak lebih dari setengah teks aslinya (Radev, 2002). Dalam dunia komputer, peringkasan juga dikenal dengan Peringkasan teks Otomatis atau ATS (*automatic text summarization*). Peringkasan teks adalah pembuatan bentuk yang lebih singkat dari suatu teks dengan memanfaatkan aplikasi yang dijalankan dan dioperasikan pada komputer (Nugraha, 2008). Dengan adanya ringkasan tersebut, dapat mempermudah aparat penegak hukum dalam memahami dokumen tanpa harus membaca keseluruhan isi dokumen.

Pada penelitian yang telah dilakukan, umumnya menggunakan objek berupa artikel dan berita. Sedangkan untuk dokumen hukum, masih belum banyak dilakukan. Oleh karena itu, dalam penelitian ini mengusulkan peringkasan teks pada dokumen hukum dengan menggunakan metode latent semantic analysis (LSA) diharapkan dapat menghasilkan ringkasan kalimat dengan baik dan sesuai dengan isi dari dokumen hukum tersebut.

## Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah yang dapat diambil adalah sebagai berikut :

1. Bagaimanakah membaca dan mengubah dokumen hukum dalam format pdf ke dalam teks yang dapat diolah oleh sistem peringkasan otomatis.
2. Bagaimanakah pengimplementasian metode LSA (*Latent Semantic Analysis*) padaperingkasan dokumen hukum.
3. Bagaimanakah pengujian yang dilakukan untuk mengukur akurasi peringkasan dokumen otomatis.

## Tujuan

Berdasarkan latar belakang di atas, maka rumusan masalah yang dapat diambil adalah sebagai berikut :

1. Bagaimanakah membaca dan mengubah dokumen hukum dalam format pdf ke dalam teks yang dapat diolah oleh sistem peringkasan otomatis.
2. Bagaimanakah pengimplementasian metode LSA (*Latent Semantic Analysis*) padaperingkasan dokumen hukum.
3. Bagaimanakah pengujian yang dilakukan untuk mengukur akurasi peringkasan dokumen otomatis.

## Batasan Masalah

Agar penelitian ini dapat berjalan sesuai dengan rencana dan tujuan awal, maka penulis memberikan batasan-batasan masalah yaitu :

1. Data yang digunakan yaitu data berupa dokumen hukum pembelaan kasus pada Kantor Advokat Ahmad Riyadh Umar Balhmar, Surabaya, Jawa Timur, Indonesia.
2. Dokumen yang diringkas berformat pdf.
3. Bagian yang digunakan untuk peringkasan yaitu pada pendahuluan.
4. Peringkasan yang digunakan yaitu secara ekstraktif.
5. Sistem ini berbasis website.

## Manfaat

Manfaat yang diharapkan dalam penelitian ini yaitu hasil implementasi dapat membantu aparat hukum dalam melakukan peringkasan agar mempermudah dalam pemahaman dokumen hukum.

## Tata Cara Penulisan

Uraian dalam penulisan skripsi ini disusun dengan sistematika penulisan sebagai berikut :

BAB I : PENDAHULUAN

Bab ini berisikan uraian tentang hal-hal yang melatar belakangi masalah dilakukannya pembuatan penelitian dan yang menjadi dasar permasalahan. Terdiri atas latar belakang masalah, rumusan masalah, tujuan penelitian, metodologi penelitian, sistematika penulisan, dan penjadwalan kegiatan penelitian.

BAB II : LANDASAN TEORI

Pembahasan dalam bagian ini mengenai landasan teori yang menjadi referensi utama dalam melaksanakan penelitian.Teori-teori tersebut adalah teori mengenai “Sistem Pakar Diagnosa Hama Dan Penyakit Tanaman Kopi Dalam Upaya Pengendalian Menggunakan Metode Case Based Reasoning (Studi Kasus Perkebunan Kopi Ngantang)”.

BAB III : METODOLOGI PENELITIAN

Bagian ini berisi tahapan yang dilakukan untuk menyelesaikan masalah pada tugas akhir yang bersumber dari proses dalam perencanaan tugas akhir. Metodologi penelitian berisi uraian tentang metode pengambilan data, metode pengembangan sistem, fase-fase pengembangan sistem dan pengujian sistem.

BAB IV : ANALISIS DAN PERANCANGAN

Bab ini menjabarkan dan menguraikan tentang analisa dan perencanaan pembuatan keseluruhan aplikasi dari penelitian yang dilakukan, serta melakukan analisa hasil yang didapat. Serta lampiran user interface.

BAB V : IMPLEMENTASI

Bab ini menjelaskan tentang bagaimana sistem dibuat dan berjalan berdasarkan analisa dan perancangan yang telah dilakukan sebelumnya. Dimana sistem diharapkan mampu membuat “Sistem Pakar Diagnosa Hama Dan Penyakit Tanaman Kopi Dalam Upaya Pengendalian Menggunakan Metode Case Based Reasoning (Studi Kasus Perkebunan Kopi Ngantang)”.

BAB VI : PENGUJIAN DAN PEMBAHASAN

Bab ini berisikan tentang tampilan seperti form – form penginputan dan output dalam “Sistem Pakar Diagnosa Hama Dan Penyakit Tanaman Kopi Dalam Upaya Pengendalian Menggunakan Metode Case Based Reasoning (Studi Kasus Perkebunan Kopi Ngantang)”. Selain itu dilakukan juga pembahasan tentang analisa hasil yang diperoleh dari sistem yang dibuat.

BAB VII : KESIMPULAN

Bab ini dibagi menjadi dua sub bab, kesimpulan yang menjawab permasalahan yang dihadaopi dan saran yang berisikan solusi alternatif untuk permasalahan yang terjadi pada laporan akhir ini.

BAB VIII : DAFTAR PUSTAKA

Berisikan catatan semua sumber referensi yang digunakan dalam penulisan laporan akhir.

# BAB 2 TINJAUAN PUSTAKA

## Penelitian Terdahulu

Beberapa penelitian terdahulu yang memiliki keterkaitan dengan penelitian ini, untuk menentukan posisi penelitian. Adapun penelitian terdahulu yang peneliti maksud adalah:

Berdasarkan penelitian pada tahun 2014, Agustinus Widiantoro telah membangun sebuah aplikasi peringkasan dokumen berbahasa jawa secara otomatis menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*). Dari penelitian tersebut, hasil yang didapatkan menunjukan bahwa tingkat keakurasian sistem mencapai 64%. Tetapi, kelemahan dari sistem ini yaitu kurangnya daftar kata umum (*stopwords*) yang digunakan sehingga penulis menyarankan untuk melakukan penambahan daftar kata umum untuk meningkatkan akurasi (Widiantoro, 2014).

Pada tahun 2015, penelitian dilakukan oleh Geetha J.K. dan Deepamala N. dengan judul “*Kannada text summarization using Latent Semantic Analysis*”. Penelitian tersebut menggunakan Bahasa Kanada. Tingkat akurasi yang didapatkan yaitu 94% dan *precision* sebesar 80% (N & K, 2015).

Pada tahun 2017, penelitian dilakukan oleh Sohini Roy Chowdhury , Kamal Sarkar dan Santanu Dam dengan judul “*An Approach to Generic Bengali Text Summarization Using Latent Semantic Analysis*”. Penelitian tersebut menggunakan aksara Bengali. Teks yang diringkas diambil 10% dari teks asli. Nilai F-Score yang didapatkan yaitu 0.324347.

Penelitian berikutnya dilakukan oleh Tinaliah dan Triana Elizabeth pada tahun 2018. Peneliti menggunakan metode *Latent Semantic Analysis* dan *Jaro-Winkler Distance* dalam mendeteksi plagiarisme dokumen. Tingkat akurasi pada penelitian tersebut menghasilkan nilai plagiat mencapai 97,14% (Tinaliah & Elizabeth, 2018).

Berikut merupakan tabel 6.1 Penelitian terdahulu.

## *Text Mining*

*Text mining* memiliki definisi menambang data yang berupa teks dimana sumber data di dapatkan dari dokumen. Tujuan *text mining* adalah mencari kata-kata yang dapat mewakili isi dari dokumen dan dilakukannya analisa keterhubungan antar dokumen.

Jenis masukan (*input*) dari *text mining* berupa data tak terstruktur yang merupakan pembeda utama dari *data mining* dimana menggunakan data terstruktur atau basis data sebagai masukan. *Text mining* dapat dianggap sebagai proses dua tahap yang diawali dengan penerapan struktur terhadap sumber data teks dan dilanjutkan dengan ekstraksi informasi dan pengetahuan yang relevan dari data teks terstruktur ini dengan menggunakan teknik dan alat yang sama dengan *data mining*. Proses yang umum dilakukan oleh text mining, yaitu perangkuman otomatis, kategorisasi dokumen, penggugusan teks, dll (Nindito, 2016).

## Peringkasan Teks Otomatis

Peringkasan teks otomatis merupakan pembuatan rangkuman dari suatu teks secara otomatis dengan menggunakan serta memanfaatkan sistem peringkasan teks yang dijalankan pada komputer. Sebuah sistem peringkasan teks diberi *input* (masukan) berupa teks kemudian sistem akan memproses dengan melakukan peringkasan yang akan menghasilkan *output* (keluaran) berupa teks yang lebih singkat dari sumber teks aslinya (Hovy, 2005).Terdapat dua pendekatan peringkas teks yaitu:

1. Ekstraksi (*extractive summary*)

Pada teknik ekstraksi, sistem menyalin unit-unit yang dianggap paling penting dari sebuah teks dan diubah menjadi ringkasan. Unit-unit teks yang disalin dapat berupa klausa utama, kalimat utama, atau paragraf utama tanpa ada penambahan kalimat-kalimat baru yang terdapat pada dokumen aslinya.

1. Abstraksi (*abstractive summary*)

Teknik abstraksi menggunakan metode *linguistic* untuk memeriksa dan menafsirkan teks menjadi ringkasan. Ringkasan teks tersebut dihasilkan dengan cara menambahkan kalimat-kalimat baru yang merepresentasikan intisari teks sumber ke dalam bentuk yang berbeda dengan kalimat-kalimat yang ada pada teks sumber (Gupta & Lehal, 2010).

Pada penelitian ini, metode yang digunakan untuk melakukan peringkasan teks otomatis adalah dengan menggunakan teknik ekstraksi. Hasil dari ringkasan merupakan kalimat asli yang terdapat pada dokumen dan tidak mengalami penambahan kalimat.

## **Pre-Processing**

Pada tahapan ini, data tekstual akan diubah menjadi teks agar dapat diolah oleh sistem. Penelitian ini menggunakan dokumen sebagai inputan awal. Proses yang digunakan antara lain :

1. Pembentukan Kalimat

Pembentukan kalimat yaitu pemecahan teks dokumen menjadi kumpulan kalimat berdasarkan delimiter.

1. *Case Folding*

*Case folding* merupakan pengubahan huruf pada kalimat menjadi huruf kecil (*lowercase*) dan penghilangan karakter yang tidak valid seperti tanda baca.

1. *Tokenizing*

Pada proses ini, kalimat tersebut dipecah kembali menjadi beberapa kata tunggal penyusunnya.

1. *Stopword Removal*

*Stopword removal* adalah proses penghilangan kata-kata yang tidak merepresentasikan isi dokumen.

1. *Stemming*

*Stemming* adalah proses pengembalian kata tunggal yang memiliki imbuhan menjadi kata dasar.

## **TF-IDF**

Setelah dokumen diproses dengan cara preprocessing, tokenizing, filtering dan stemming, selanjutnya dilakukan proses pembobotan kata. Pada Metode ini pembobotan kata dalam sebuah dokumen dilakukan dengan mengalikan nilai TF dan IDF.

Term frequency (TF) adalah pengukuran yang paling sederhana dalam metode pembobotan. Pada metode ini, masing-masing term diasumsikan mempunyai proporsi kepentingan sesuai jumlah kemunculan dalam teks dokumen. Term frequency dapat memperbaiki nilai recall pada information retrieval, tetapi tidak selalu memperbaiki nilai precision (Tokunaga & Iwayama, 1994). Hal ini disebabkan term yang frequent cenderung muncul di banyak teks, sehingga term tersebut memiliki kekuatan

Inverse document frequency (IDF) adalah metode pembobotan term yang lebih condong (fokus) untuk memperhatikan kemunculan term pada keseluruhan kumpulan teks. Pada IDF, term yang jarang muncul pada keseluruhan koleksi teks dinilai lebih berharga. Nilai kepentingan tiap term diasumsikan berbanding terbalik dengan jumlah teks yang mengandung term tersebut (Tokunaga & Iwayama, 1994).

Term frequncy inverse document frequency (TF•IDF) adalah metode pembobotan yang menggabungkan metode TF dan IDF. Metode ini diusulkan oleh Salton sebagai sebuah kombinasi metode yang dapat memberikan performansi yang lebih baik, khususnya dalam memperbaiki nilai recall dan precision (Tokunaga & Iwayama, 1994). Berikut ini merupakan perhitungannya :

TF.IDF = TF \* log(N/DF) (7.1)

Keterangan :

TF : Jumlah term tersebut  
N : Total dokumen  
DF : Jumlah dokumen yang mengandung suatu term

## *Latent Semantic Analysis*

*Latent Semantic Analysis* (LSA) menurut bahasa terbagi atas beberapa kata yang penting yaitu *latent* dan *semantic*, *latent* yang memiliki arti tersembunyi atau sesuatu yang masih belum terlihat, sedangkan semantic berasal dari bahasa yunani “*semanticos*” yang berarti memberi tanda, penting atau cabang linguistik yang mempelajari arti dan makna dari suatu bahasa, kode atau jenis representasi lainnya.

Dari pengertian dapat ditarik kesimpulan bahwa, LSA adalah menguraikan atau menganalisa makna yang masih tersembunyi dari suatu bahasa, kode atau jenis representasi lainya, guna memperoleh informasi yang penting. Kesamaan kata dan kalimat diperoleh dengan cara menggunakan *Singular Value Decomposition* (SVD), di mana SVD mempunyai kapasitas untuk mereduksi *noise*, sehingga dapat meningkatkan hasil akurasi pada ringkasan (Peter & Kp, 2009).Peringkasan dokumen menggunakan metode LSA memiliki 3 tahapan yaitu : pembentukan matriks input dari dokumen untuk menampilkan kalkulasi, Singular Value Decomposition, dan penyeleksian kalimat.

# BAB 3 METODOLOGI PENELITIAN

Dalam metodologi penelitian, menjelaskan langkah-langkah yang dilakukan untuk merancang sistem. Metodologi penelitian yang dilakukan dalam penelitian ini melalui beberapa tahap, yaitu studi literature, pengumpulan data, analisa dan perancangan, implementasi sistem, uji coba sistem, dan kesimpulan. Tahapan dalam penelitian ini terdapat pada bagan berikut.

Studi Literatur

Pengumpulan Data

Analisa dan Perancangan

Implementasi

Pengujian dan Analisis

Pengambilan Kesimpulan

Gambar 31 Bagan

## Metode Pengambilan Data

Metode pengambilan data digunakan untuk mengumpulkan data-data yang dibutuhkan dalam proses penelitian. Pemgambilan data dengan cara antara lain studi literature dan *scraping* website

### **Studi Literatur**

Studi literatur digunakan untuk mencari informasi pustaka yang berkaitan dengan penelitian ini, yang diperoleh dari jurnal maupun penelitian sebelumnya meliputi metode – metode yang dilakukan dalam melakukan peringkasan dokumen.

### **Observasi**.

Dalam pengambilan data, Penulis langsung terjun ke lapangan Kantor Advokat Ahmad Riyadh Umar Balhmar Kota Surabaya untuk melakukan observasi.

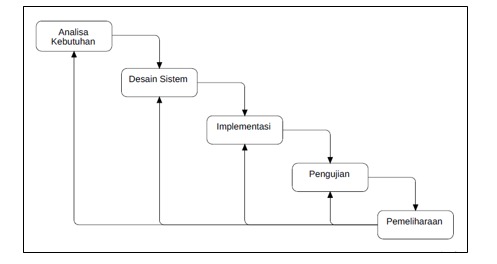
### **Wawancara**

Penulis melakukan kegiatan tanya jawab dengan pihak Advokat Kota Surabaya.

## Metode Pengembangan Perangkat Lunak

Pengembangan perangkat lunak dapat diartikan sebagai proses membuat suatu perangkat lunak baru untuk menggantikan perangkat lunak lama secara keseluruhan atau memperbaiki perangkat lunak yang telah ada. Agar lebih cepat dan tepat dalam mendeskripsikan solusi dan mengembangkan perangkat lunak   juga hasilnya mudah dikembangkan dan dipelihara, maka pengembangan perangkat lunak memerlukan suatu metodologi khusus. Metodologi pengembangan perangkat lunak adalah suatu proses pengorganisasian kumpulan metode dan konvensi notasi yang telah didefinisikan untuk mengembangkan perangkat lunak.

Pada penelitian ini, penulis menggunakan metode waterfall dalam pengembangan perangkat lunak. Berikut adalah tahapan – tahapan dalam metode waterfall :



Gambar 3.2 Metode Waterfall

Metode Waterfall memiliki tahapan-tahapan sebagai berikut :

1. Analisa Kebutuhan

Kondisi, kriteria, syarat atau kemampuan yang harus dimiliki oleh perangkat lunak untuk memenuhi apa yang disyaratkan atau diinginkan pengguna.

1. Desain Sistem

Tahapan perancangan sistem mengalokasikan kebutuhan-kebutuhan sistem baik perangkat keras maupun perangkat lunak dengan membentuk arsitektur sistem secara keseluruhan. Perancangan perangkat lunak melibatkan identifikasi dan penggambaran abstraksi sistem dasar perangkat lunak dan hubungannya.

1. Implementasi

Tahap ini merupakan tahap setelah dilakukannya perancangan sistem. Perancangan yang telah dibuat akan diterjemahkan menjadi kode pemrograman dengan menggunakan bahasa php berbasis website.

1. Pengujian Unit

Untuk pengujian unit digunakan *black box testing. Black Box Testing* berfokus pada spesifikasi fungsional dari perangkat lunak. Pengujian yang dilakukan masing-masing unit. Setelah integrasi seluruh sistem diuji untuk mengecek  setiap kegagalan maupun kesalahan.

1. Pemeliharaan

Pada tahap pemeliharaan sistem, aplikasi yang telah dijalankan akan di lakukan perbaikan jika terdapat kesalahan yang tidak ditemukan pada langkah sebelumnya. Perbaikan implementasi unit sistem dan peningkatan jasa sistem sebagai kebutuhan baru.

.

## Metode Pengolahan Data

Data berita tersebut perlu dilakukan pengolahan agar dapat dilakukan proses peringkasan teks otomatis. Data judul berita digunakan untuk uji kesamaan berita, sedangkan data konten berita akan dilakukan proses ekstraksi fitur. Ekstraksi fitur terdiri dari beberapa beberapa fitur yaitu kesamaan kata antar kalimat, kesamaan kata dengan judul, penggunaan huruf kapital, kelimat kutipan, penggunaan angka, dan panjang kalimat. Sedangkan untuk pengujian Relevansi dilakukan dengan membandingkan judul antar sumber data menggunakan metode *cosine similarity*, jika *score* hasil keterkaitannya melebihi nilai *threshold* maka dipilih konten yang memiliki bobot tertinggi pada proses ekstraksi fitur.

## Metode Pengujian

Untuk pengujian pada ringkasan dapat dilakukan dengan mencari nilai *recall*, *precision* dan *f-measure*. *Recall* ialah kemampuan untuk mengambil peringkat teratas yang sebagian besar relevan (benar). *Precision* adalah berapa banyak dokumen yang berhasil diambil oleh sistem, sedangkan untuk mengukur kualitas *recall* dan *precision* menggunakan *f-measure*.

(7.6)

(7.7)

(7.8)

## Analisa Hasil dan Kesimpulan

Proses pengambilan kesimpulan yaitu dengan cara melihat akurasi dari hasil peringkasan teks otomatis pada berita online.

# BAB 4 ANALISA DAN PERANCANGAN

(gambaran umum sistem)

## 4.1. Data (tampilkan contoh data, data hanya di tampilkan sebagian sisanya dimasukan ke lampiran, cara pengumpulan data)

## 4.2. Kebutuhan Fungsional

## 4.3. Kebutuhan Non-fungsional

## 4.4. *Use Case* Diagram

## 4.5. Use Case Pesifikasi (alur dari tiap use case yang ada di atas)

## 4.6. Activity Diagram

## 4.7. ERD (di tambah semua metode)

(ditambahkan UML class diagram, “kemungkina DFD”)

## 4.8. Alur Metode Floyd Warshall (gambar flowchart, sub bab masing” dari proses di flowchart, rumus tinggal di panggil menggunakan persamaan dari BAB 2)

# BAB 5 IMPLEMENTASI DAN PENGUJIAN

## 5.1. Implementasi *Source Code*

### 5.1.1 sub bab use case

## 5.2. Implementasi Tampilan (mock up dari aplikasi)

## 5.3. Pengujian

(Sesuai scenario pada bab 3)

# PENUTUP

## 6.1. Kesimpulan (menjawab rumusan masalah, selanjtnya kesimpulan setelah pengejian poin poin dalam bentuk paragraf)

## 6.2 Saran (kendala)

# DAFTAR PUSTAKA (minimal 10)

# LAMPIRAN