**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS TEKNOLOGI INFORMASI**

**INSTITUT TEKNOLOGI SEPULUH NOPEMBER**

**USULAN TUGAS AKHIR**

# **IDENTITAS PENGUSUL**

Nama : **Febri Fernanda**

NRP : **5108100136**

Dosen Wali : **Prof. Ir. Handayani Tjandrasa, M.Sc., Ph.D.**

# **JUDUL TUGAS AKHIR**

***“Rancang Bangun Aplikasi Pengelompokan dan Pemberi Rekomendasi Berita Lomba Online Menggunakan Klasifikasi Fuzzy Berbasis Kerangka Kerja Spring”***

# **LATAR BELAKANG**

Saat ini penyebaran berita lewat Internet berkembang dengan sangat cepat. Salah satunya adalah berita tentang lomba yang diselenggarakan oleh suatu organisasi atau institusi. Permasalahan yang muncul adalah pengguna tidak mudah mengunjungi situs yang menampilkan informasi lomba tersebut. Karena dengan hanya memanfaatkan *search engine,* situs yang dimaksud tidak mudah untuk ditemukan. Beberapa uji coba telah dilakukan, salah satunya yakni dengan memasukkan kata kunci “lomba fotografi” ke dalam pencarian Google. Hasil yang didapat adalah tampilnya puluhan situs yang memuat kata “lomba fotografi”. Sebagian besar memang tepat menampilkan situs-situs yang merupakan promosi lomba fotografi, namun ada juga situs yang hanya menampilkan artikel berita tentang lomba yang sudah dilaksanakan, atau situs yang berupa sebuah iklan, atau malah menampilkan situs ‘penipu’, yang hanya ingin mengeruk keuntungan dengan banyaknya *traffic* pengunjung yang mengunjungi situsnya. Untuk menangani masalah tersebut, saat ini telah bermunculan portal-portal yang khusus menampilkan informasi tentang lomba-lomba yang sedang diselenggarakan (penulis sedikitnya menemukan lima portal). Namun dari solusi tersebut, pengguna harus tetap membuka tautan pada portal satu persatu dan melakukan pencarian dan penyaringan jenis lomba yang sesuai dengan keinginan.

Pada tugas akhir ini, penulis akan mengajukan pengembangan suatu sistem yang dapat menangani masalah yang dipaparkan di atas. Sistem tersebut secara berkala akan dapat memperbarui dan mengelompokkan kategori data berita lomba dari berbagai portal lomba yang ada, dengan syarat portal tersebut memiliki fitur Rich Site Summary (RSS) *feed*. Fitur RSS *feed* mengijinkan pengguna untuk dapat berlangganan berita dari suatu portal. Sistem akan mengambil data melalui RSS *feed*  lalu mengolahnya ke dalam database sistem. Sistem juga akan dapat memberikan rekomendasi lomba yang mungkin diinginkan oleh pengguna. Rekomendasi didasarkan atas data profil pengguna dan sejarah pencarian pada sistem. Dengan adanya rekomendasi yang dikirim melalui email, pengguna akan mendapatkan kemudahan dalam mendapatkan informasi lomba yang diinginkan tanpa melakukan pencarian.

Fuzzy Similarity dan K-Nearest Neighbors (FSKNN) digunakan untuk melakukan pengelompokan kategori berita lomba. Metode tersebut mengelompokkan berita secara multilabel, sehingga sangat sesuai digunakan karena berita lomba bisa memiliki lebih dari satu label. Saat ini metode tersebut cukup dapat diandalkan karena optimasinya lebih baik dari metode lain, seperti Pohon Keputusan C4.5 dan Multi Label KNN[2]. Sebelum mengimplementasikan metode FSKNN, dibutuhkan pemrosesan teks untuk mengubah masukan awal yang berupa teks mentah menjadi sebuah struktur teks yang siap digunakan. Salah satu proses dalam pemrosesan teks adalah *stemming* , yang secara sederhana dapat diartikan merubah kata menjadi kata dasar[1]. Penggunaan lain dari FSKNN adalah untuk fitur pencarian dan fitur pemberi rekomendasi. FSKNN menghasilkan nilai Term Frequency(TF) dan nilai Inverse Document Frequency (IDF) yang nantinya digunakan untuk membangun sistem pencarian dan pemberi rekomendasi[8].

Pembangunan sistem menggunakan kerangka kerja (*framework)* akan bermanfaat bagi pengembang maupun bagi pengguna, karena dengan adanya standar arsitektur pembangunan sistem, maka perawatan dan pengembangan kembali sistem akan menjadi lebih mudah. Spring Framework akan digunakan dalam pembangunan sistem yang diajukan. Spring Framework merupakan salah satu kerangka kerja berbasis pemrograman java yang cukup populer dan handal[3].

Sistem yang diajukan akan dikembangkan berdasarkan pengembangan berbasis komponen. Manfaat dari pengembangan berbasis komponen adalah untuk mempermudah dan mempercepat pembangunan dan pengembangan kembali[4]. Komponen yang dibuat juga bisa digunakan oleh aplikasi lain. Komponen akan dibangun dengan memanfaatkan Java Class Library, sehingga terbentuk sebagai sebuah pustaka jar. Komponen yang akan dibangun meliputi komponen pemrosesan teks, komponen klasifikasi, dan komponen pemberi rekomendasi.

# **RUMUSAN MASALAH**

Rumusan masalah yang diangkat dalam Tugas Akhir ini adalah:

1. Bagaimana mendesain dan membangun komponen yang sesuai untuk aplikasi pengelompokan dan pemberi rekomendasi berita lomba online.
2. Bagaimana memanfaatkan Spring Framework sebagai kerangka kerja untuk membangun aplikasi pengelompokan dan pemberi rekomendasi berita lomba online.
3. Bagaimana mengimplementasikan algoritma Fuzzy Similarity K-Nearest Neighbors (FSKNN) ke dalam kode untuk mengelompokkan berita lomba secara multilabel.
4. Bagaimana memanfaatkan FSKNN dan Pembobotan TF-IDF untuk membangun sistem pencarian dan pemberi rekomendasi berita lomba.

# **BATASAN MASALAH**

Permasalahan yang dibahas dalam Tugas Akhir ini memiliki beberapa batasan, yakni sebagai berikut:

1. Komponen yang dibuat adalah komponen yang berbasiskan pemrograman java. Komponen yang dimaksud berupa pustaka java yang berekstensi .jar.
2. Bahasa yang digunakan dalam berita lomba adalah bahasa Indonesia.
3. Data profil pengguna didapat dari masukan profil pribadi dan sejarah *query* pencarian.
4. Penggunaan kerangka kerja Spring difokuskan pada pemanfaatan modul Spring Web Model View Controller (MVC), namun tidak menutup kemungkinan menggunakan modul Spring lainnya.

# **TUJUAN TUGAS AKHIR**

Tujuan dari tugas akhir adalah untuk merancang dan membangun sebuah aplikasi web yang dapat mengelompokkan dan memberikan rekomendasi berita lomba online secara otomatis berdasarkan pengembangan perangkat lunak berbasis komponen.

# **MANFAAT TUGAS AKHIR**

Manfaat dari pembangunan aplikasi yang diajukan pada tugas akhir adalah untuk memudahkan pengguna dalam mendapatkan informasi berita lomba yang diinginkan secara cepat. Dengan aplikasi tersebut, pengguna tidak perlu mencari tautan-tautan yang diinginkan secara manual, karena dengan otomatis sudah ditangani oleh fitur pencarian, pengelompokan,dan pemberi rekomendasi yang ada pada aplikasi. Pengguna akan ditautkan ke portal yang memuat berita lomba ketika pengguna sudah yakin untuk melihat informasi secara detail tentang lomba tersebut.

1. **TINJAUAN PUSTAKA**
2. **Tugas akhir yang terkait**

Ada beberapa tugas akhir yang terkait dengan system yang diajukan. Keterkaitan yang ada adalah kesamaan tema yang diangkat, yakni pengkategorian berita. Beberapa tugas akhir yang terkait adalah :

1. Klasifikasi Kategori dan Identifikasi Topik Pada Artikel berita berbahasa Indonesia.[12]

Secara ringkas, pada tugas akhir ini digunakan perhitungan *Likelihood* untuk mengklasifikasikan kategori dan *Cosine Similarity* untuk mengidentifikasikan topik dokumen.

Aplikasi yang dibangun mempunyai fitur utama mengelompokkan online berita menjadi hirarki dengan kategori yang lebih khusus dan menampilkannya dalam bentuk *snippet.* Data berita yang digunakan adalah portal berita umum seperti www.kompas.com, www.antaranews.com dan www.tempointeraktif.com.

Ada tiga inti proses pada tugas akhir ini :

* + - Representasi dokumen teks.

Proses ini melakukan ekstraksi kata kunci dengan menggunakan metode TF-IDF.

* + - Klasifikasi Kategori

Proses ini meliputi proses *training* dan proses klasifikasi.

* + - Identifikasi Topik

Proses ini melakukan identifikasi topik pada berita yang baru masuk.

1. Perancangan dan Pembangunan Perangkat Lunak Pembuat Berita Menggunakan RSS dengan Klasifikasi Bayesian Berbasis Kerangka Kerja JSF.[11]

Aplikasi yang dibangun pada tugas akhir ini adalah aplikasi pembuat berita online berbasis web dengan memanfaatkan kerangka kerja Java Server Faces. Berita yang didapat menggunakan pembaca RSS pada beberapa situs berita yang sudah ada. Berita dikelompokkan menggunakan metode klasifikasi Naïve Bayes yang dalam implementasinya menggunakan pustaka WEKA.

Beberapa proses yang ada pada lapisan servis aplikasi diantaranya :

* + - Proses mengunduh channel rss
    - Proses memperbarui berita pada channel rss
    - Proses memperbarui berita pada channel secara otomatis
    - Melakukan pengaturan konten berita

Beberapa perbedaan dari sistem yang akan dibangun dengan tugas akhir terkait di atas adalah :

* Metode klasifikasi yang dipakai dalam sistem adalah Fuzzy Similarity dan K-Nearest Neighbors (FSKNN). Metode tersebut merupakan metode yang didasarkan atas klasifikasi secara multilabel yang terbilang masih baru.
* Sistem yang akan dibangun lebih memfokuskan berita yang diklasifikasi, yakni berita lomba sedang pada kedua tugas akhir di atas berupa berita umum.
* Sistem diimplementasikan seperti sebuah mesin pencari. Pengguna dapat mencari dan melihat kategori berita lomba melalui sistem berdasarkan kata kunci.
* Sistem memiliki fitur unggulan berupa pemberian rekomendasi berita lomba yang diinginkan.
* Sistem dibangun berdasarkan pengembangan berbasis komponen. Komponen atau pustaka yang dibangun bisa dapat digunakan secara mudah oleh aplikasi lain, khususnya yang satu tema.

1. **Pengembangan Perangkat Lunak Berbasis Komponen[4][5][6]**

Component Based Software Development merupakan konsep dalam pengembangan perangkat lunak berbasis komponen. Pengembangan dengan konsep ini menganut prinsip **sistem sebagai hasil perakitan komponen-komponen, komponen sebagai entitas-entitas** reusable*,* danupgrading **sistem dengan memodifikasi atau mengganti komponennya.**

Konsep ini sangat erat kaitannya dengan teknologi berorientasi objek. Komponen tersusun atas kelas-kelas yang telah dibuat berdasarkan konsep OO. Secara umum proses yang terjadi dalam model ini adalah:

1. Kelas-kelas yang akan digunakan kembali diidentifikasi dengan menguji kelas tersebut dengan data yang akan dimanipulasi dengan aplikasi/perangkat lunak dan algoritma yang baru.
2. Kelas yang dibuat pada proyek sebelumnya disimpan dalam Class Library, sehingga bisa langsung diambil dari pustaka yang sudah ada. Jika ternyata ada kebutuhan kelas baru, maka kelas baru dibuat dengan metode berorientasi objek.
3. Perangkat lunak dibangun dengan kelas-kelas yang sudah ditentukan atau kelas baru yang dibuat.

Beberapa contoh komponen antara lain :

* *Widget Graphical User Interface*
* *Major “Reused” Functionality*
* Algoritma untuk *searching/sorting*/*classification*
  + Database *connection/querying*
* Aplikasi khusus

Komponen Perhitungan Akuntansi

* Komponen *Computational Fluid Dynamics*

Pembangunan perangkat lunak dengan menggunakan komponen yang sudah tersedia dapat menggunakan komponen COTS (*Commercial off-the-shelf*) yang bisa didapatkan dengan membeli komponen atau membangunnya sendiri secara internal. Component Based Software Engineering menekankan perancangan dan pembangunan perangkat lunak dengan menggunakan komponen perangkat lunak yang sudah ada. CBSE terdiri dari dua bagian yang terjadi secara paralel yaitu *Software Engineering* dan D*omain Engineering*.

* *Domain engineering* menciptakan model domain bagi aplikasi yang akan digunakan untuk menganalisis kebutuhan pengguna. Identifikasi, pembangunan, pengelompokan dan pengalokasikan komponen-komponen perangkat lunak supaya bisa digunakan pada sistem yang ada dan yang akan dating.
* *Software engineering* (*component-based development*) melakukan analisis terhadap domain model yang sudah ditetapkan kemudian menentukan spesifikasi dan merancang berdasarkan model struktur dan spesifikasi sistem, kemudian melakukan pembangunan perangkat lunak dengan menggunakan komponen-komponen yang sudah ditetapkan berdasarkan analisis dan rancangan yang dihasilkan sebelumnya hingga akhirnya menghasilkan perangkat lunak.

Pada tugas akhir ini, akan dibangun komponen yang dapat berdiri sendiri, mudah digunakan, dan mudah dikembangkan kembali. Komponen tersebut diharapkan bisa digunakan oleh berbagai aplikasi dan platform, sehingga dapat menghemat sumber daya, waktu, dan biaya dalam pembangunannya.

Jenis komponen yang akan dibangun berupa pustaka jar, karena merupakan jenis pustaka yang umum digunakan dalam berbagai aplikasi berbasis java. Pembangunan pustaka memanfaatkan Java Class Library pada Netbeans IDE 7.0.1. Dalam implementasinya, komponen yang akan dibangun bisa mempunyai kemungkinan memanfaatkan pustaka jar yang sudah ada. Untuk menangani masalah tersebut, penulis akan memanfaatkan teknologi One-JAR. One-JAR mampu membungkus aplikasi java bersama pustaka jar yang mendukungnya dalam sebuah file jar. Walaupun begitu, beberapa pustaka dari luar juga diperlukan, diantaranya adalah RSS Reader dan JavaMail.

1. **Spring Framework[3][7]**

Spring adalah kerangka kerja pengembangan aplikasi berbasis java yang handal dan populer. Spring dianggap sebagai *standard stack development* untuk pembuatan aplikasi berbasis Java EE. Spring muncul karena spesifikasi (Enterprise Java Beans) EJB yang memaksakan pengembangan komponen harus mengikuti aturan EJB agar dapat berjalan dalam aplikasi server SJAS dan Jboss. Dengan Spring, pengembangan komponen dapat dilakukan dengan teknik pemrograman yang lebih sederhana. Keunggulan utama dari Spring adalah adanya teknik *Inversion of Control*(IoC). *Inversion of Control* (IOC), atau yang disebut juga sebagai *Dependency Injection* (DI), adalah pendekatan pengembangan perangkat lunak di mana objek atau kerangka kerja (seperti Spring Framework) yang terpisah bertanggung jawab untuk menciptakan dan "menyuntikkan (*injecting*)" objek ke objek lain yang mempunyai ketergantungan terhadapnya. Hal ini akan menghasilkan kode yang *loosely coupled* (mempunyai sedikit ketergantungan terhadap objek lain) sehingga mudah untuk  
diuji dan digunakan kembali.

Spring framework memiliki beberapa modul yang memudahkan pengembang untuk membangun aplikasi. Modul utama yang akan digunakan dalam pengembangan sistem yang diajukan adalah Spring MVC. Spring MVC akan membentuk kerangka kerja berbasis Model, View dan Controller. Kerangka kerja tersebut memisahkan lapisan elemen pembangunan menjadi elemen Model, elemen View, dan elemen Controller. Pemisahan tersebut akan menjadi standard dalam pengkodean sehingga memudahkan dalam perawatan dan pengembangan ulang.

Spring dapat dengan mudah diintegrasikan dengan berbagai macam ORM dan library database seperti Hibernate, Ibatis dan sebagainya. Dalam pembangunan sistem, Spring Framework akan diintegrasikan dengan ORM Hibernate.

1. **Rich Site Summary (RSS)**

RSS merupakan fitur yang umumnya ada pada setiap situs. Fungsi utama dari fitur ini adalah memberikan rangkuman singkat mengenai artikel atau postingan baru dari situs. Fitur tersebut berguna bagi pengguna untuk berlangganan artikel tanpa harus menuju tautan situs yang bersangkutan. Biasanya fitur pembaca RSS ada pada *add on* atau *plugin* w*eb browser*. Struktur RSS secara umum terdiri dari judul, deskripsi, tanggal dibuat, dan URL. Data-data dari struktur itu akan disimpan dalam database sistem secara periodik untuk kemudian diproses.

Sistem mendapatkan data RSS melalui Rome Library. Rome Library menyediakan fitur untuk mendapatkan struktur data RSS dari situs. Rome Library cukup handal, *up to date* dengan versi RSS saat ini, dan mudah diimplementasikan.

1. **Pemrosesan Teks[1]**

Pemrosesan teks adalah proses pengelolaan teks untuk dapat digunakan lebih lanjut. Pada tugas akhir kali ini, pemrosesan teks digunakan untuk klasifikasi data, pencarian data, dan untuk mendapatkan nilai rekomendasi. Proses-proses pada pemrosesan teks yang akan digunakan adalah :

1. Penghapusan Token

Bagian ini bertujuan untuk menghapus semua token yang ada pada teks. Token bisa berupa tanda baca, tanda penghubung, spasi, angka dan semacamnya.

1. *Case folding*

Bagian ini bertujuan untuk mengubah semua huruf menjadi huruf kecil.

1. Penghapusan *Stopwords*

Bagian ini bertujuan untuk menghapus semua kata yang merupakan *stopwords. Stopwords* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*. Contoh *stopwords* adalah “yang”, “dan”, “di”, “dari” dan seterusnya.

1. *Stemming*

Bagian ini memuat proses pengembalian berbagai bentukan kata ke dalam suatu representasi yang sama. Dalam arti yang sederhana, stemming bertujuan untuk mendapatkan kata dasar, walau pada implementasinya tidak selalu seperti itu. Algoritma stemming bahasa Indonesia yang akan digunakan adalah *non deterministic stemmer*[9] atau *enhanced* *confix stripping approach*[10].

1. **Klasifikasi Fuzzy Similarity and K-Nearest Neighbors (FSKNN)[2]**

FSKNN merupakan algoritma klasifikasi teks secara multilabel yang cukup baru dipublikasikan (tahun 2011). Algoritma ini menyempurnakan algoritma Multi Label KNN(MLKNN) dengan perbaikan di sisi komputasi.

Secara umum, algoritma ini terdiri dari dua bagian :

* Pelatihan

Proses training digunakan untuk mengelompokkan dokumen kedalam kategori-kategori yang mungkin masih terjadi ketidaksesuaian pada kategori aslinya. Selain itu, proses pelatihan membentuk desain yang digunakan dalam pencarian kategori suatu dokumen yang belum diketahui kategorinya. Hasil pelatihan digunakan sebagai acuan untuk menentukan klasifikasi pada dokumen yang sebenarnya.

* Mengelompokkan data training kedalam kategori-kategori

Tahap ini bertujuan untuk mendapatkan derajat keanggotaan suatu dokumen terhadap kategori.

1. Menghitung nilai keanggotaan setiap term pada setiap kategori.
2. Menghitung fuzzy similaritysetiapdokumen terhadap setiap kategori.
3. Cari nilai keanggotaan setiap term terhadap setiap dokumen
4. Menghitung derajat keanggotaan dokumenpada setiap kategori

Untuk setiap dokumen, apabila nilai derajat keanggotaan lebih dari *threshold*, maka dokumen tersebutmerupakan anggota pada kategori yang dimaksud.

1. Membentuk himpunan pencarian pada setiap dokumen.

* Menghitung nilai *prior* dan *likelihood*

1. Menghitung nilai *prior*
2. Menemukan k-nearest neighbours dokumen pada himpunan pencarian yang telah terbentuk.
3. Menghitung nilai *likelihood*

* *Testing*

1. Menghitung derajat keanggotaan dokumen pada kategori
2. Pengecekan derajat keanggotaan dokumen pada kategori

Jika lebih dari *threshold*, maka dokumen tersebut diklasifikasikan ke dalam kategori tersebut.

1. Membentuk himpunan pencarian dari dokumen tersebut.
2. Menemukan k-nearest neighbor dokumen pada himpunan pencarian yang

telah terbentuk.

1. Menghitung nilai *label vector* dokumen untuk setiap kategori, jika nilai *label vector* adalah 1 maka dokumen merupakan anggota kategori , sebaliknya jika nilai 0 maka dokumen tidak termasuk ke dalam kategori.

Secara singkat Fuzzy Similarity Measure (FSM) digunakan untuk mengelompokkan pola pelatihan ke dalam kategori(*cluster*). Kemudian hanya dokumen yang terlatih pada *cluster* yang nilai fuzzy similaritynya diatas *threshold* dipertimbangkan dalam pencarian k tetangga terdekat(KNN) untuk dokumen. Sebuah dokumen baru (*unseen document*) diberi label berdasarkan k tetangga terdekatnya menggunakan nilai maksimum dari sebuah perkiraan posteriori(*posteriori estimate*). Hasil penelitian menunjukkan bahwa metode FSKNN dapat bekerja lebih efektif daripada metode C4.5, LinearSVM, dan MLKNN.

Pada implementasinya, akan disediakan beberapa berita lomba yang mewakili setiap kategori untuk dijadikan *data training*. *Data training* itu dijadikan acuan untuk proses klasifikasi berita bagi masukan berita baru.

1. **Sistem Pemberi Rekomendasi[8]**

Sistem pemberi rekomendasi yang diajukan menggunakan metode pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*) dan Cosine Similarity.

* TF-IDF adalah pembobotan term berdasarkan frekuensinya dalam suatu dokumen. TF-IDF merupakan pembobotan yang umum digunakan dalam berbagai permasalahan *text mining*, sehingga implementasinya cukup mudah karena banyak referensinya.
* Cosine Similarity adalah suatu formula untuk dapat mengecek tingkat kemiripan suatu vektor. Vektor yang dihitung kemiripannya didasarkan vektor yang dibentuk dari pembobotan TF-IDF.

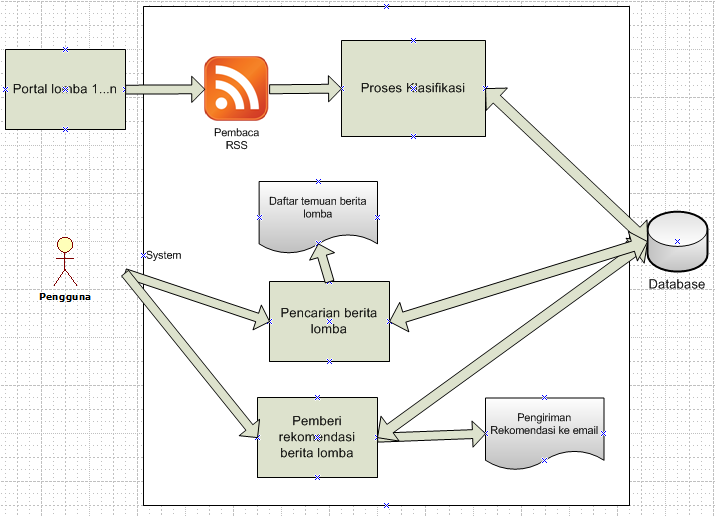
Mekanisme sistem pemberi rekomendasi adalah:

* Sistem mengambil data profil dan sejarah pencarian pengguna pada sistem.
* Data tersebut dibentuk sebagai sebuah *query*.
* Dari *query* tersebut akan dicari nilai dokumen terdekatnya berdasarkan cosine similarity. Klasifikasi FSKNN dimanfaatkan untuk mempercepat komputasi sistem ini, karena dapat memfokuskan penemuan rekomendasi lomba berita pada kelompok tertentu saja.
* Nilai dokumen yang lebih dari *threshold* akan dijadikan sebagai berita yang direkomendasikan kepada pengguna.

# **RINGKASAN TUGAS AKHIR**

Sistem yang diajukan adalah sistem pengelompokan dan pemberi rekomendasi berita lomba online berbasis komponen. Sistem terbentuk dari susunan beberapa komponen yang akan dibangun sendiri dan beberapa menggunakan komponen yang sudah ada.Komponen atau library yang akan dibangun pada tugas akhir ini disesuaikan pada kebutuhan pada aplikasi pengelomspokkan dan pemberi rekomendasi otomatis berita lomba online. Untuk menentukan jenis komponen yang akan dibangun diperlukan pengetahuan mengenai deskripsi dan kebutuhan fungsional dari sistem.

Sistem merupakan system berbasis web. Pengguna dapat menjalankan aplikasi sistem melalui w*eb browser*. Gambaran sistem dijelaskan secara singkat pada gambar berikut.



**Gambar 1. Diagram Blok Sistem**

Sistem membaca data berita lomba melalui RSS dari beberapa portal berita lomba. Data tersebut kemudian diolah dan diklasifikasi kemudian dimasukkan database sistem. Pengguna mempunyai dua fitur utama, yakni pencarian berita lomba dan pemberi rekomendasi berita lomba. Kedua fitur tersebut melakukan proses berdasarkan database yang telah terbentuk. Pencarian berita lomba berdasaran kata kunci dan kategori. Keluaran dari pencarian adalah daftar temuan, sedangkan keluaran dari pemberi rekomendasi adalah daftar rekomendasi yang bisa dilihat pada tampilan aplikasi atau dikirim melalui email. Khusus untuk fitur pemberi rekomendasi, pengguna harus sudah terdaftar keanggotaannya pada sistem.

Sistem merupakan sistem berbasis web dengan dua aktor yang terlibat di dalamnya. Kedua aktor tersebut adalah pengguna dan admin, yang masing-masingnya mempunyai kebutuhan fungsional seperti yang dijelaskan pada gambar di bawah ini.



**Gambar 2. Diagram Kasus Pengguna**

Penjelasan diagram kasus pengguna :

* Mendaftar keanggotaan

Pengguna bisa mendaftar keanggotaannya pada sistem. Dengan mendaftar, pengguna bisa menggunakan semua fitur yang disediakan, semisal berlangganan rekomendasi berita lomba. Pendaftaran bisa melalui Facebook API atau mengisi formulir secara manual.

* Melihat rekomendasi berita lomba

Pengguna akan mendapatkan rekomendasi lomba yang mungkin diinginkan. Rekomendasi ini didasarkan pada beberapa faktor, antara lain data profil pengguna dan sejarah *query* pencarian.

* Mengubah data profil pribadi

Pengguna bisa mengisi dan mengubah data pribadinya, yang meliputi nama, umur, pekerjaan, kegemaran, kesukaan, lokasi tinggal dan sebagainya. Data-data tersebut digunakan sebagai salah satu faktor yang akan dihitung untuk mendapatkan rekomendasi berita lomba.

* Melakukan pencarian berita lomba

Pengguna dan admin bisa mencari lomba yang diinginkan menggunakan fitur pencarian yang ada di aplikasi. Beberapa penyaringan juga akan diberikan sistem untuk mempermudah pencarian, seperti penyaringan tanggal dibuat dan jenis kategori. Dari hasil pencarian, pengguna bisa langsung menuju ke tautan dari lomba yang diinginkan.

* Melihat hasil klasifikasi lomba

Pengguna dan admin bisa mendapatkan daftar lomba berdasarkan kategori lomba. Pengelompokan berita lomba ke dalam kategori diklasifikasikan oleh sistem.

* Memperbarui data berita lomba dari portal

Admin dapat memperbarui database sistem dari database portal beritasecara manual. Dalam fungsi tersebut, data-data baru yang berasal dari berbagai database server portal yang didapat melalui RSS Reader API akan dimasukkan ke dalam database sistem untuk kemudian diproses.

* Menghapus data berita

Admin bisa menghapus data berita lomba yang ada pada database sistem.. Penghapusan digunakan untuk kasus data lomba yang tanggalnya sudah *expired*.

* Mengubah daftar portal berita lomba

Admin dapat menambah atau mengurangi daftar portal berita lomba yang dilanggan.

Dari pengetahuan tentang aplikasi yang telah dipaparkan di atas, dirumuskan komponen, yang berupa pustaka, yang dapat dibangun dalam menyusun aplikasi tersebut. Pustaka yang akan dibangun antara lain :

1. Pustaka Pemrosesan Teks Bahasa Indonesia

Pustaka ini berguna untuk melakukan semua pemrosesan teks yang digunakan untuk membangun aplikasi. Fitur yang ada pada komponen ini meliputi :

* Penghapusan token

Pustaka mampu menghapus token yang ada dalam suatu teks. Daftar token telah tersedia dalam properti pustaka. Properti tersebut dapat ditambah atau dikurangi. Masukan dan keluaran dari fitur ini bisa berupa string, objek, atau struktur data primitif seperti array dan list. Masukan dan keluaran yang dinamis bertujuan agar fitur ini dapat digunakan dengan mudah oleh pengembang aplikasi lain.

* Menghilangkan *stopwords*

Fitur ini berguna untuk menghilangkan kata-kata yang tidak perlu dari teks. Daftar *stopwords* yang ada pada bahasa Indonesia telah dipasangkan ke dalam properti dari pustaka. Data tersebut dapat diubah sesuai keinginan pengguna. Perubahan dapat memaksimalkan proses pemrosesan teks. Masukan dan keluaran fitur ini akan dibuat sedinamis mungkin.

* *Case folding*

*Case folding* merupakan perubahan huruf pada teks menjadi huruf kecil semua. Proses ini berguna untuk menstandarkan huruf.

* *Stemming*

Secara umum, *stemming* dapat dikatakan proses untuk mendapatkan kata dasar. Proses ini sangat penting dan dapat dimaanfaatkan dalam segala permasalahan yang mencakup pemrosesan teks, semisal pada temu kembali informasi, semantik dan sebagainya. Dinamisasi yang didapat dari fitur ini, selain tentu masukan dan keluarannya, adalah algoritma stemming itu sendiri. Ada banyak algoritma stemming bahasa Indonesia dan itu masih berkembang sampai saat ini. Sehingga jika ada algoritma yang dirasa lebih baik, pengembangan kembalinya mudah.

Pengguna dapat memanfaatkan semua fitur yang ada pada pustaka atau hanya memanfaatkan beberapa saja. Pemanfaatan tersebut disesuaikan dengan kebutuhan. Pustaka juga menyediakan fitur utama berupa kombinasi dari fitur-fitur yang dipunyai, bisa kombinasi dari dua fitur atau bahkan semua fitur. Pada aplikasi yang diajukan penulis, dinamisasi yang mungkin bisa terjadi adalah perubahan bahasa yang digunakan.

1. Klasifikasi Fuzzy Similarity dan K-Nearest Neighbors (FSKNN)

FSKNN merupakan algoritma yang cukup baru untuk pengkategorisasian teks secara multilabel. Algoritma tersebut dapat digunakan dalam banyak permasalahan pemrosesan teks. Dengan adanya komponen Klasifikasi FSKNN, maka pembangunan aplikasi serupa yang memanfaatkan klasifikasi bisa lebih menghemat waktu dan biaya. Pada aplikasi yang diajukan penulis, dinamisasi yang mungkin adalah perubahan algoritma klasifikasi karena algoritma klasifikasi akan terus berkembang.

1. Sistem pemberi rekomendasi

Saat ini, sistem pemberi rekomendasi merupakan sistem yang sering kita temui, seperti pada aplikasi jejaring sosial yang sering kita gunakan. Banyak sekali fungsi dari pemberi rekomendasi, salah satunya untuk menampilkan iklan yang sesuai pada pengguna. Pembuatan komponen ini bisa dimanfaatkan oleh berbagai aplikasi yang menyediakan fitur pemberi rekomendasi. Sistem pemberi rekomendasi yang akan dibuat adalah berdasarkan metode pengembangan TF-IDF dan Cosine Similarity. Pada aplikasi yang diajukan penulis, dinamisasi yang mungkin adalah perubahan algoritma dari sistem pemberi rekomendasi.

Komponen-komponen yang dibuat adalah komponen yang bersifat independen, sehingga tidak tergantung pada komponen lain. Namun antar komponen bisa saling memanfaatkan fitur-fitur yang dimilikinya.

# **METODOLOGI**

1. Penyusunan proposal tugas akhir

Tahap awal dalam pengerjaan tugas akhir ini adalah penyusunan proposal tugas akhir. Pada proposal ini, penulis mengajukan gagasan mengenai perancangan dan pembangunan aplikasi pengelompokan dan pemberi rekomendasi berita lomba online berbahasa Indonesia menggunakan klasifikasi fuzzy dan memanfaatkan Spring Framework sebagai kerangka kerja.

1. Studi literatur

Pada tahap ini dilakukan pencarian, pengumpulan, penyaringan, pembelajaran dan pemahaman literatur yang diperlukan dalam pembangunan sistem yang penulis ajukan. Literatur yang dibutuhkan berhubungan dengan :

1. Analisa daftar portal berita lomba yang dapat digunakan untuk menguji coba sistem.
2. Rekayasa Perangkat Lunak berbasis Komponen (*Component based* *software Development*).
3. Pembangunan komponen berbasis pustaka jar pada bahasa java menggunakan Java Class Library.
4. Spring Framework MVC sebagai kerangka kerja berbasis java untuk pengembangan aplikasi web.
5. Pemrosesan teks berbahasa Indonesia, yang antara lain meliputi penghapusan stop words dan stemming.
6. Metode Fuzzy Similarity dan K-Nearest Neighbors untuk mengelompokkan berita lomba.
7. Metode TF-IDF dan Cosine Similarity sebagai metode untuk pencariaan dan pemberi rekomendasi berita lomba.
8. Teknologi-teknologi yang mendukung pembangunan komponen dan aplikasi, seperti One-Jar, Rome API dan sebagainya.

Literatur yang digunakan dalam pengerjaan tugas akhir ini berasal dari artikel, paper, materi kuliah dan juga buku referensi yang terkait.

1. Analisis dan Perancangan Sistem

Analisis dan perancangan kebutuhan sistem dilakukan untuk merumuskan solusi yang tepat dalam pembuatan aplikasi serta kemungkinan yang dapat dilakukan dalam pengimplementasian rancangan tersebut. Tujuan lainnya adalah untuk merumuskan solusi cara mendesain dan membangun komponen-komponen yang digunakan dalam pembuatan aplikasi. Di samping itu, permasalahan teknologi yang digunakan juga dibahas pada fase ini.

1. Implementasi

Implementasi merupakan tahap dalam membangun perangkat lunak, dimana hasil dari fase desain perangkat lunak diwujudkan ke dalam bentuk kode program. Pembangunan perangkat lunak menggunakan IDE Netbeans, sedangkan bahasa pemrograman yang dipakai adalah Java. Guna membantu kemudahan dalam membangun sistem, penulis menggunakan kerangka kerja Spring.

1. Uji Coba dan Evaluasi (internal)

Pada tahap ini, perangkat lunak yang sudah dibuat akan diuji coba dan di evaluasi secara berkala. Uji coba perangkat lunak bertujuan untuk menemukan kesalahan-kesalahan, seperti bug code, bug fungsionalitas dan sebagainya, pada kode program sedini mungkin sehingga dapat diperbaiki sesegera mungkin. Uji coba yang dilakukan pada tahap ini masih uji coba yang dilakukan oleh penulis sendiri, sehingga bersifat internal.

1. Uji coba eksternal

Pada tahap ini, aplikasi yang telah dibuat akan diuji cobakan pada beberapa sample pengguna. Untuk kemudahan evaluasi, pengguna yang dibidik adalah mahasiswa. Dari uji coba yang dilaksanakan pengguna, akan didapat nilai kepuasan. Semakin tinggi nilai kepuasan, maka aplikasi ini sudah siap dikenalkan kepada umum.

1. Penyusunan buku tugas akhir

Tahap terakhir dari proses pengerjaan tugas akhir adalah penyusunan laporan atau biasa disebut buku tugas akhir. Buku tugas akhir tersebut berisi seluruh dokumentasi mengenai proses pembangunan tugas akhir, yang mencakup dari perancangan sampai uji coba dan evaluasi dari perangkat lunak yang telah dibuat. Sistematika penulisan dalam penyusunan buku tugas akhir secara umum adalah sebagai berikut :

1. Pendahuluan
2. Latar Belakang
3. Rumusan Masalah
4. Batasan Tugas Akhir
5. Tujuan
6. Metodologi
7. Sistematika Penulisan
8. Tinjauan Pustaka
9. Desain dan Implementasi
10. Uji Coba dan Evaluasi
11. Kesimpulan dan Saran
12. Daftar Pustaka

# **JADWAL PENGERJAAN TUGAS AKHIR**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Tahapan | 2012 | | | | | | | | | | | | | | | | | | | | |
| Februari | | | | Maret | | | | April | | | | Mei | | | | | Juni | | | |
| Penyusunan Proposal |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Studi Literatur |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Analisa dan Perancangan Sistem |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Implementasi |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Uji coba dan Evaluasi(internal) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Uji coba eksternal |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Penyusunan Buku |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

# **DAFTAR PUSTAKA**

[1] Wibisono, Yudi and Khodra, Masayu Leylia. Clustering Berita Berbahasa Indonesia.

[2] Jiang,Jung-Yi , Tsai,Shian-Chi, Lee,Shie-Jue. “FSKNN: Multi-label text categorization based on fuzzy similarity and k nearest neighbors” . 2011.

[3] Prita,Anggik L. Y. , Samopa,Febriliyan, Wibowo,Radityo Prasetianto. ”Pengembangan Aplikasi Manajemen Proyek Perangkat Lunak Berbasis Spring : Modul Komunikasi dan Bug Tracker”.2011.

[4] Cai,Xia and Lyu,Michael R. and Wong,Kam-Fai. “Component-Based Software Engineering : Technologies, Development Frameworks, and Quality Assurance Schemes”. 2000.

[5] “Model dan Rekayasa Perangkat Lunak”. [online]. (http://agustina31.wordpress.com/2008/07/31/model-dan-proses-rekayasa-perangkat-lunak/ , diakses tanggal 28 Februari 2012)

[6] “Belajar Instan, Component Based Software Engineering” . [online]. (http://egadioniputri.wordpress.com/2010/03/14/belajar-instan-component-based-software-engineering/ , diakses tanggal 28 Februari 2012)

[7] Johnson ,Rod and Friends. “Spring Framework Reference Documentation”.

[8] Chen,Ting , Han,Wie-Li , Wang,Hai-Dong and friends. “Content Recomendation Sistem Based on Private Dynamic User Profile”. 2007.

[9] Purwarianti,Ayu.” A Non Deterministic Indonesian Stemmer”.2011.

[10] Mahendra, I Putu Adhi Kerta. 2008. “Enhanced Confix Stripping Stemmer And Ants Algorithm For Classifying News Document In Indonesian Language”. The 5th International Conference on Information & Communication Technology and SystemsISSN 2085-1944.

[11] Dwi Nugroho,Fajar. Perancangan dan Pembangunan Perangkat Lunak Pembuat Berita Menggunakan RSS dengan Klasifikasi Bayesian Berbasis Kerangka Kerja JSF.2012

[12] Rachmania Kusumaagama Fuddily, Aini. Klasifikasi Kategori dan Identifikasi Topik Pada Artikel berita berbahasa Indonesia.2011