

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Desain Penelitian

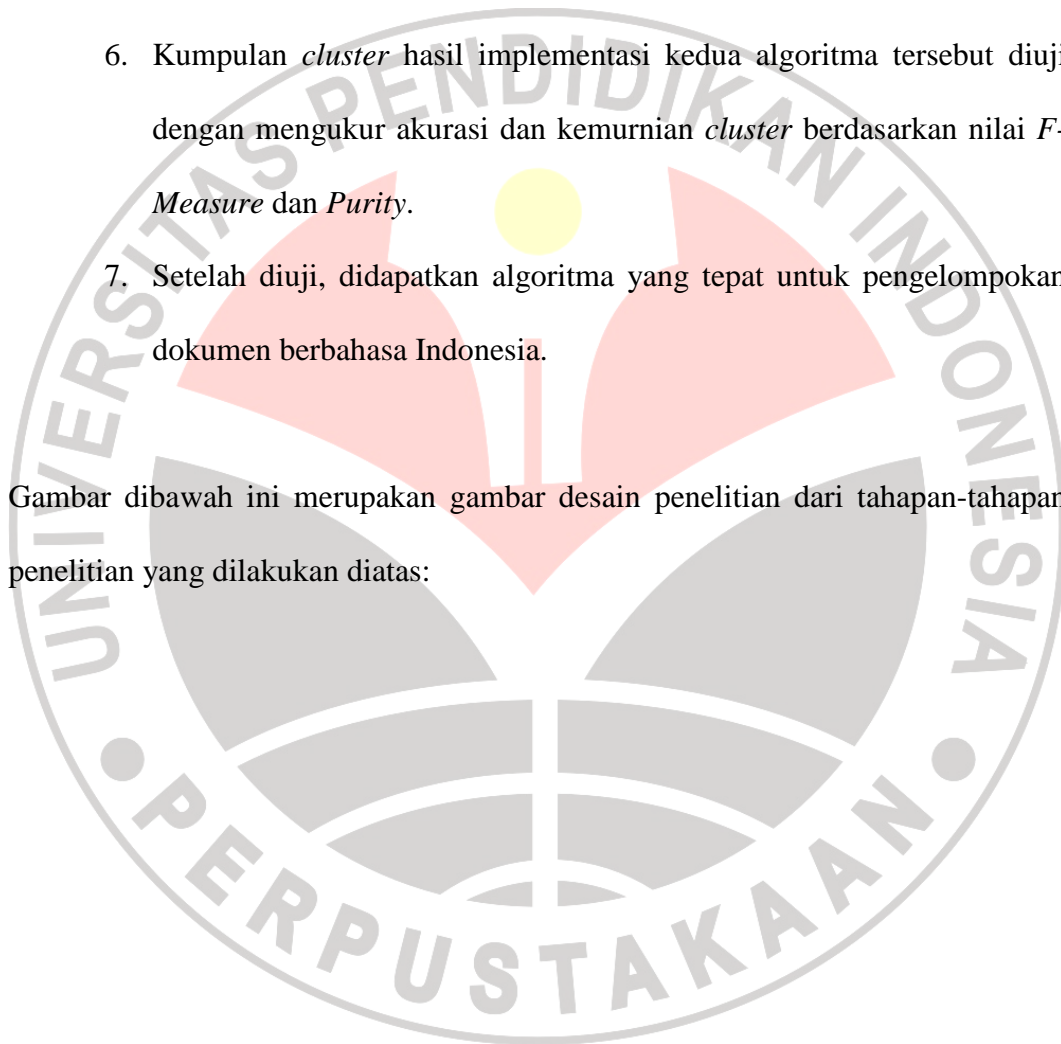
Desain penelitian adalah tahapan atau gambaran yang akan dilakukan peneliti dalam melakukan penelitian. Desain penelitian dibuat untuk memudahkan peneliti dalam melakukan tahap-tahap penelitian.

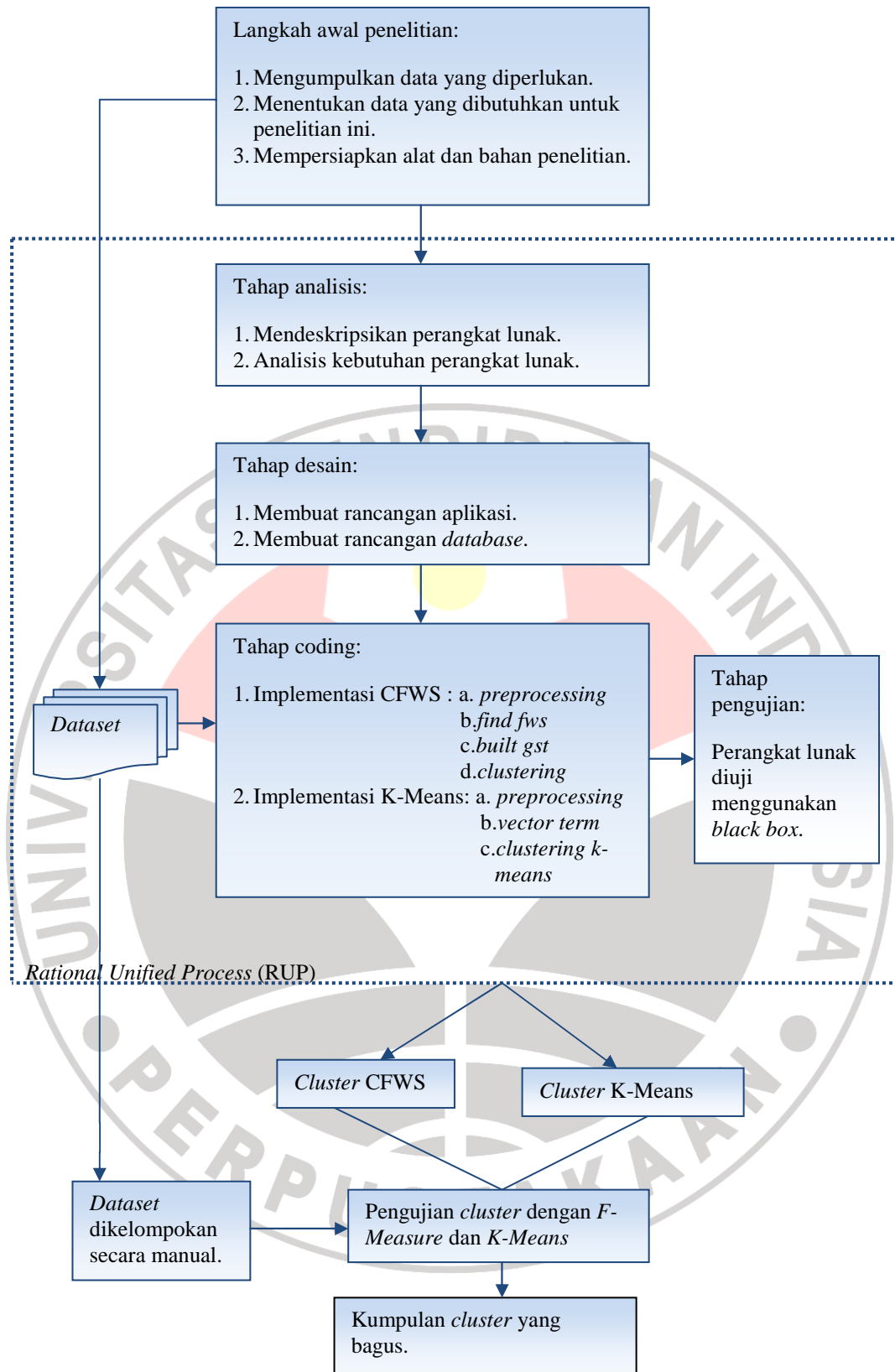
Tahapan-tahapan desain penelitian yang peneliti lakukan diantaranya adalah:

1. Mengumpulkan data yang diperlukan berupa *textbook*, *paper* dan karya ilmiah yang membahas algoritma-algoritma yang digunakan untuk pengelompokan dokumen. Dikumpulkan juga data berupa *paper* yang membahas pengujian *cluster* dan *preprocessing* teks.
2. Setelah dikumpulkan data yang diperlukan, ditentukan data yang akan dipakai pada penelitian ini yaitu berupa *textbook*, *paper* dan karya ilmiah yang membahas algoritma *Clustering Based On Frequent Word Sequences* (CFWS) dan *K-Means*.
3. Mempersiapkan alat dan bahan penelitian berupa komputer, *tools* bahasa pemrograman dan bahan penelitian berupa seratus dokumen yang diambil dari *website* kompas.

4. Setelah disiapkan alat dan bahan penelitian, dibangun perangkat lunak dengan metode *Rational Unified Process* (RUP).
5. Hasil dari pembangunan perangkat lunak pengelompokan dokumen adalah *cluster* hasil pengelompokan dokumen dengan CFWS dan *cluster* hasil pengelompokan dengan *K-Means*.
6. Kumpulan *cluster* hasil implementasi kedua algoritma tersebut diuji dengan mengukur akurasi dan kemurnian *cluster* berdasarkan nilai *F-Measure* dan *Purity*.
7. Setelah diuji, didapatkan algoritma yang tepat untuk pengelompokan dokumen berbahasa Indonesia.

Gambar dibawah ini merupakan gambar desain penelitian dari tahapan-tahapan penelitian yang dilakukan diatas:





**Gambar 3.1** Desain Penelitian

## 3.2 Metodologi Penelitian

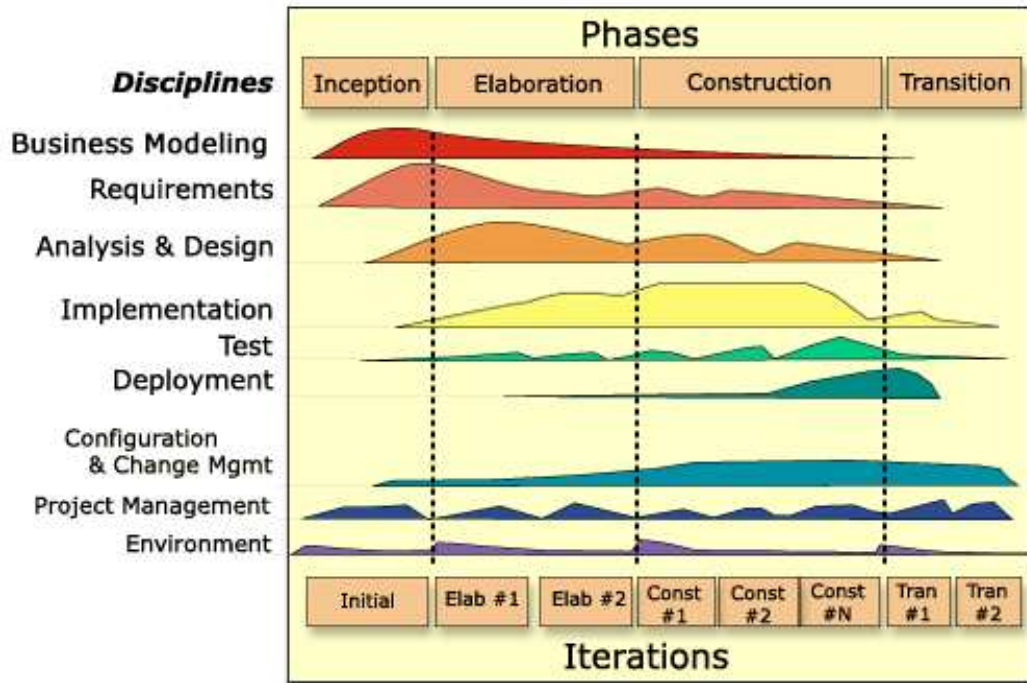
### 3.2.1 Metode Pengambilan Data

Metode pengambilan data yang digunakan pada penelitian ini adalah metode studi kepustakaan. Pada metode ini dilakukan eksplorasi dan studi literatur yaitu dengan mempelajari teknik *clustering*, algoritma CFWS, algoritma *K-Means*, pembangunan sistem pengelompokan dokumen melalui literatur-literatur seperti *textbook*, jurnal, karya ilmiah, *paper* dan sumber ilmiah lainnya yang didapat dari internet. Pada metode ini juga dilakukan eksplorasi dan studi literatur yang membahas pengujian *cluster* dan *preprocessing* teks.

### 3.2.2 Metode Pengembangan Perangkat Lunak

Metode pendekatan perangkat lunak yang digunakan untuk mengelompokkan dokumen adalah *Rational Unified Process* (RUP). *Rational Unified Process* merupakan proses rekayasa perangkat lunak yang melaksanakan tugas dan tanggung jawab diantara organisasi pengembang perangkat lunak berdasarkan pendekatan disiplin. RUP merupakan petunjuk untuk menggunakan *Unified Modeling Language* (UML) yang merupakan *tools* yang digunakan untuk membangun perangkat lunak pada penelitian ini.

Proses RUP dapat dilihat pada gambar berikut:



**Gambar 3.2** Rational Unified Process

(Sumber: Openia.2010 : 1 dalam [www.openia.com](http://www.openia.com))

Berdasarkan gambar diatas, terdapat 4 fase yaitu:

1. *Inception Phase* : Pada fase ini, dilakukan analisis kebutuhan perangkat lunak dan deskripsi perangkat lunak. Pada fase ini, dianalisis kebutuhan sistem, interaksi sistem, interaksi sistem dengan *user* dan analisis kelas-kelas yang akan diimplementasikan. Kebutuhan sistem pada penelitian ini yaitu sistem harus dapat melakukan *preprocessing* teks, sistem dapat melakukan proses mendapatkan *frequent word sequence*, sistem dapat membangun *generalized suffix tree*, sistem dapat melakukan *clustering*

dengan *k-mismatch*, sistem dapat mengubah dokumen menjadi *vector space model* dan sistem dapat melakukan *clustering* dengan *k-means*.

2. *Elaboration Phase*: Pada fase ini, dirancang kelas-kelas yang berkaitan dengan kebutuhan-kebutuhan sistem yang telah didefinisikan pada fase *inception*. Kelas-kelas tersebut adalah *preprocessing*, *getWords*, *fws*, *TreeNode*, *mapping*, *node*, *ClusterKmismatch*, *overlapping*, *ClusterKmeans*, *vectorTerm* dan dokumen. Pada fase ini, dirancang juga antarmuka perangkat lunak atau *interface* yang disesuaikan dengan kebutuhan fungsional yang telah ditentukan.
3. *Construction Phase*: Pada fase ini, semua desain yang berkaitan dikembangkan dan digabung menjadi perangkat lunak. Pada fase ini, dilakukan tahapan *coding* terhadap desain yang telah dirancang. Pada implementasi algoritma CFWS, *frequent word sequence* yang telah didapatkan, digabung berdasarkan batas ketidakcocokan kata. Apabila batas ketidakcocokan kata adalah satu maka dua *frequent word sequence* yang memiliki satu perbedaan kata dapat digabungkan. Untuk algoritma *k-means*, ditentukan *centroid* awal terlebih dahulu. Dokumen-dokumen yang memiliki kedekatan terhadap *centroid* tersebut digabung menjadi satu. Setelah itu, dihitung ulang *centroid* pada suatu *cluster* berdasarkan jumlah anggota pada *cluster* tersebut.
4. *Transition Phase*: Pada fase ini, perangkat lunak diuji dengan menggunakan *black box*. Pengujian dilakukan untuk semua fungsi yang

terdapat pada perangkat lunak. Jika ada kesalahan atau tambahan fitur dapat kembali ke tahap sebelumnya yaitu *Construction Phase*.

### 3.3 Alat dan Bahan Penelitian

#### 3.3.1 Alat Penelitian

Untuk membuat perangkat lunak, pada penelitian ini digunakan alat penelitian berupa perangkat keras dan perangkat lunak sebagai berikut:

##### 1. Perangkat keras

Perangkat keras yang digunakan adalah *notebook* TOSHIBA Portege M800 dengan spesifikasi sebagai berikut:

- a. *Processor* Intel Centrino 2.0 GHz
- b. RAM 1 GB
- c. *Harddisk* minimal 40 GB
- d. Monitor beresolusi 1280 *pixels*
- e. *Mouse* dan *keyboard*

##### 2. Perangkat lunak

Perangkat lunak yang digunakan adalah sebagai berikut:

- a. Sistem operasi: Microsoft Windows Vista Basic
- b. Bahasa pemrograman: Java (Java SE 6)
- c. *Tools* pemrograman: Netbeans 6.0.1
- d. *Database Management System* (DBMS): MySQL

Adapun *library* yang digunakan untuk membantu membangun perangkat lunak adalah sebagai berikut:

- a. MySQL Connector Java versi 5.1.6. *Library* yang digunakan untuk melakukan koneksi antara bahasa pemrograman Java dengan basis data MySQL.

Adapun daftar *stopword* berbahasa Indonesia yang digunakan pada saat *preprocessing text* diambil di situs berikut ini:

[http://fpmipa.upi.edu/staff/yudi/stop\\_words\\_list.txt](http://fpmipa.upi.edu/staff/yudi/stop_words_list.txt)

### **3.3.2 Bahan Penelitian**

Bahan penelitian yang digunakan adalah dokumen berupa berita berbahasa Indonesia yang didapat dari *World Wide Web* yang disalin ke dalam *file* berekstensi txt. Bahan penelitian ini diambil dari situs berita Kompas.

## **3.4 Implementasi Penelitian**

### **3.4.1 Prosedur Pengerjaan Penelitian**

Peneliti ini membutuhkan beberapa tahap pengerjaan dimulai dari pengumpulan data berupa dokumen berita berbahasa Indonesia, eksplorasi teori algoritma *K-Means* dan CFWS, pemahaman bahasa pemrograman Java, merancang perangkat lunak pengelompokan dokumen dan menguji akurasi dan kemurnian *cluster* hasil implementasi CFWS dan *K-Means* berdasarkan patokan *dataset* yang dikelompokkan secara manual.



Adapun urutan pengerjaan yang harus dilalui adalah sebagai berikut:

1. Pengumpulan data berupa dokumen berita berbahasa Indonesia dan menyalinnya pada *file* txt. Dikumpulkan juga data yang membantu proses *clustering* seperti daftar *stopword* berbahasa Indonesia.
2. Melakukan pengelompokan dokumen secara manual untuk digunakan sebagai patokan perhitungan akurasi dan kemurnian pada saat proses pengujian *cluster*.
3. Membuat rancangan kelas-kelas dengan Java. Dalam langkah ini, terlibat beberapa *tool* baik *software* maupun *library* yang digunakan untuk mengembangkan perangkat lunak.
4. Pembangunan *interface* aplikasi yang disesuaikan dengan fitur-fitur yang telah didefinisikan pada kebutuhan fungsional sistem yaitu:
  - a. *Preprocessing*
  - b. Menemukan *frequent word sequence*
  - c. Membangun *generalized suffix tree*
  - d. *Clustering* dengan *k-mismatch*
  - e. Mengubah dokumen menjadi *vector space model*
  - f. *Clustering* dengan *k-means*
5. Perancangan dan pembangunan *datastore* yang disesuaikan dengan kebutuhan fitur. Pada penelitian ini, dibuat 4 *datastore* yaitu *folder* untuk penyimpanan *file* hasil *preprocessing*, menemukan *frequent word sequence*, pengubahan dokumen menjadi *vector space model* dan *preprocessing* pada paket implementasi *k-means*.

### 3.4.2 Pengumpulan Data

Pengumpulan data pada penelitian ini meliputi teori-teori yang menunjang implementasi algoritma *K-Means* dan CFWS. Dikumpulkan juga teori-teori yang berhubungan dengan pengujian akurasi dan kemurnian *cluster* serta teori-teori yang berhubungan dengan *preprocessing* teks.

