

CHAPTER 3

RESEARCH METHODOLOGY

This project can be classified with few steps below:

1. Planning

Define what problem will be resolved with the program. In here the problem topic is data mining. Searching from many source and references for the problem. After looking for the problem, then detemined news clustering as the project topic.

2. Analysis

Analyse the news clustering problem. Analyse what the main problem and what algorithm and data structure needed to solve it. Needed right algorithm and data structure in order to make the approtiate program. Therefore, find and learn about algorithm and data structure that suitable applied to the problem. Search for references about algorithm from journals that ever been made. After looking for it, K-means algorithm and 2-dimensional array decided to be used. Along with it, text mining technique, stopword list, and Nazief-Adriani stemmer are also used. Text mining preprocessing method including tokenization, stopword removal, stemming and term weighting. Stopword list obtained from internet. And learn about Nazief-Adriani stemmer rule.

3. Design

After analysing and learn about K-means algorithm and 2-dimensional data structure, specify the steps how to make the program from start to finish. Illustrate the design into DFD and flowchart. The DFD can be outline of how the program work. And the flowchart will be the process of how the program runs.

4. Implementation

Make the code with implementing the K-means algorithm and 2-dimensional array data structure. Make the program with HTML & PHP as programming language. The code is saved in localhost. Several file created like `index.php` as the main page that contain user interface, `process.php` as the main page for all the background process. Some classes are also created like `cluster.php` that contain K-means process, `cURL.php` that contain the cURL PHP library, `stemmer.php` that contain Nazief-Adriani stemmer, `stopword.php` that contain process to check stopword from article, `termweight.php` that contain term weighting (TF-IDF) process, `tokenisasi.php` that contain process to tokenize the article, and `webscraping.php` that contain the process to get online news articles from `kompas.com`.

5. Testing

After finished all the code, try to test the program with online news articles from `bisniskeuangan.kompas.com` from July 8, 2017. And for the user news articles are taken from the same website but different date, from June 13, 2017. When there is some error found, fix the code.