Advanced Science Letters
Vol. XXXXXXXXX

Internship Logbook Mining (Illustration in Politeknik Negeri Batam)

Mir'atul Khusna Mufida[1], Muhammad Nashrullah[2], Monalisa[3]
[1,2]Informatics Engineering, Politeknik Negeri Batam, Batam 29461, Indonesia
[3]Multimedia and Networking, Politeknik Negeri Batam, Batam 29461, Indonesia

Internship is a part of the job training for final-year student as one of graduation requirement. During internship, student has to make weekly report called logbook to monitor their activity in the office. Logbook contains basic internship information such as company details and students activity in the workday validated by the supervisor. This documents are very crucial to check whether the internship student get the appropriate workplace in term of company profile and daily tasks. Moreover logbook is able to determine what kind of job description offered by the Internship Company or institution. Student internship logbook monitored by the lecturer as their supervisor and coordinator to categorize student daily task. The supervisor and internship coordinator examine the logbook manually. The manual manner of logbook content identification is time consuming and inefficient. This research aims to tackle this problem by implementing text mining. Logbook contents is categorized by rule-based classifier. The solution is collaborating data mining technics and web-technology to classify logbook content automatically without too much effort.

Keywords: Internship, Logbook, Data Mining, Rule-based.

1. INTRODUCTION

Internship is a part of the job training that is usually taken by the students as one of the requirements for graduation. During the internship, students are required to prepare an internship report containing a general description of the company and a description of the internship. The internship description is in the section called apprentice logbook. One of the logbook functions is as the material used by the supervisor and internship coordinator to categorize job type parameters based on the content on the logbook.

The problems that arise regarding the analysis of the content of apprentices' logbooks conducted by supervisors and internship coordinators should be done manually. The process of reading a logbook written descriptively of course will take a long time and the process of concluding the suitability between student activities with the parameters of the type of work is still done manually.

In accordance with the above problems, it is necessary to have a system to facilitate the supervisor and internship coordinator in categorizing the intern book logbook content. Using the rule based method, the categorization will be done by matching the content of the new apprentice logbook with the rules that have been made based on the job type parameters on the previous internship logbook content.

2. STATE OF THE ART

A. Internship

Internships are part of the job training that is usually taken by the students as one of the requirements for graduation. During the internship, students are required to prepare an internship report containing a general description of the company and a description of the internship. The internship report is prepared as an evaluation method for lecturer and coordinator to assess

the appropriateness between the description of the internship activity to demonstrate their skillset and match it to the campus graduate profile competence.

B. Text mining and rule-based classifier

Text mining is a variation of data mining trying to find interesting patterns from a large collection of textual data. Text mining is a data mining that is performed by a computer to get something new, something unknown or recover implicitly implicit information, derived from information that is extracted automatically from different text data sources (Feldman & Sanger, 2007).

In the book Introduction to Data Mining, Tan, Steinbach and Kumar explained that the rule-based method works by classifying data using a set of rules "if ⋯ then ⋯" or can be written (condition) → y, where condition is the conjunction of attributes and y is the class label. Rule based classifier will form the rules (R) to group data that have the same behavior using some particular parameters.

3. ANALYSIS AND DESIGN

A. General Description

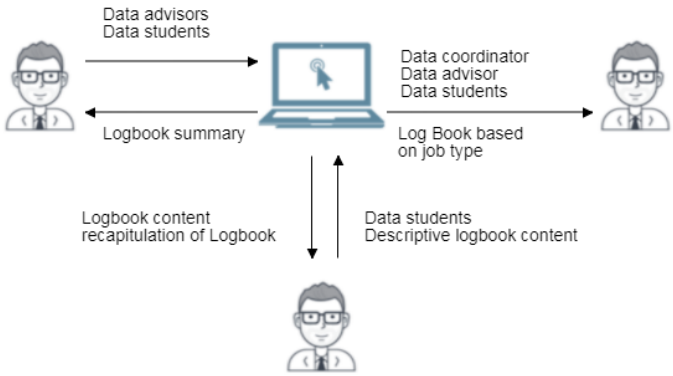

Fig.1. General Description

Figure 1 describes about the system architecture. In this system containing three different users namely student, supervisor and coordinator as listed in Table 1.

Table 1. User lists and access

No	User Type	Access
1	Student	- login - add logbook content to the system - update logbook content - search logbook content - view logbook content - select logbook content - logout
2	Supervisor	- login - register student - preprocessing - logbook content grouping - view grouping result - logout
3	Coordinator	- register - login - Add the static result parameter of job description - define and classify the result - view the result and job description - logout

Students login using their username and password as their credentials. Then, students create weekly logbook by explaining their activity in the internship workplace. The duration of internship is 4 months so they need to submit at least 16 weeks of logbook. So each student supposed to have 16 logbooks to note their activity for internship program.

Problem comes from manual logbook submission. And supervisor need a lot of effort to resolve logbook content carefully to determine any possible problems. In this paper we demonstrate automatic solution to proceed logbook contents. A system that be able to add logbook content and ease supervisor and coordinator make a decision.

In addition, the system also displays the recapitulation of the logbook content of each student. The supervisor performs login to access logbook content and analyze internship document.

Supervisor register students to give them access to the system. Supervisor can categorize the student internship logbook content per week so that each student has 16 logbook summaries per internship periods.

Raw data will preprocess to make it ready to continue the mining process. The preprocessing consists of tokenizing, filtering and stemming. Tokenizing eliminates delimiters such as periods (.), Commas (,), spaces and numeric characters present in the word. Tokenizing can also perform the tolowercase function, which converts all the characters into lowercase letters. Filtering is the process of eliminating words that do not contain meaning or stopword. Stemming is the process of discarding both the prefix and the suffix until it becomes the basic word. The data obtained from preprocessing will be categorized using the rule based method.

Furthermore, the application views the categorization result of their logbook contents. coordinator can categorize the logbook of a student using the logbook conclusion at the categorization stage on the mentor's side. Then the system displays the desired categorization result based on the given job descriptions.

B. Flowchart System Design

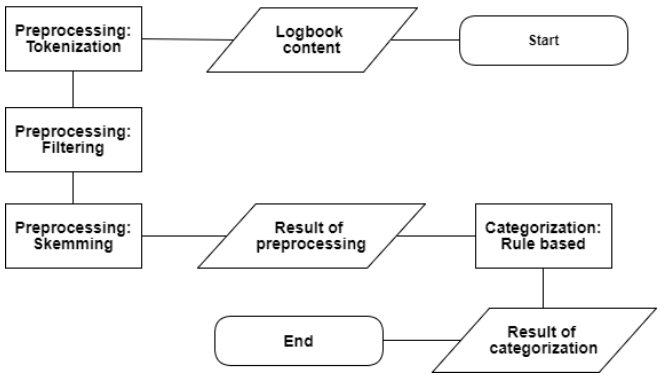

Fig.2. Flowchart Design System

Figure 2 represents the categorization algorithms. First of all the logbook content is entered into the system. Then the system performs preprocessing i.e., tokenizing, filtering and stemming. The results of the preprocessing process are categorized using the rule base method to obtain the results of the categorization. System designs by UML as Figure 3.

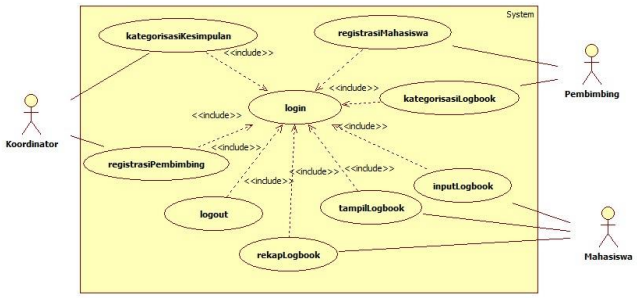

Fig. 3 Use case diagram of system

C. Categorization Method

In this research the categorization is done by the method of rule-based classifier, so it is necessary to create a rule to be able to produce a category. The rule are as follows:
If the logbook input is the same as the keyword in the database, then the number of keywords will be calculated. The most keyword is the category.

Keyword will save in the database as designed below Figure 4.

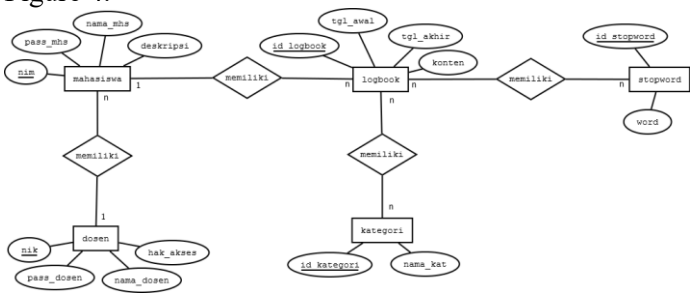

Fig. 4 Design ER Diagram

Here is the application implementation shown by Figure 5

4. IMPLEMENTATION AND EXPERIMENTAL RESULT

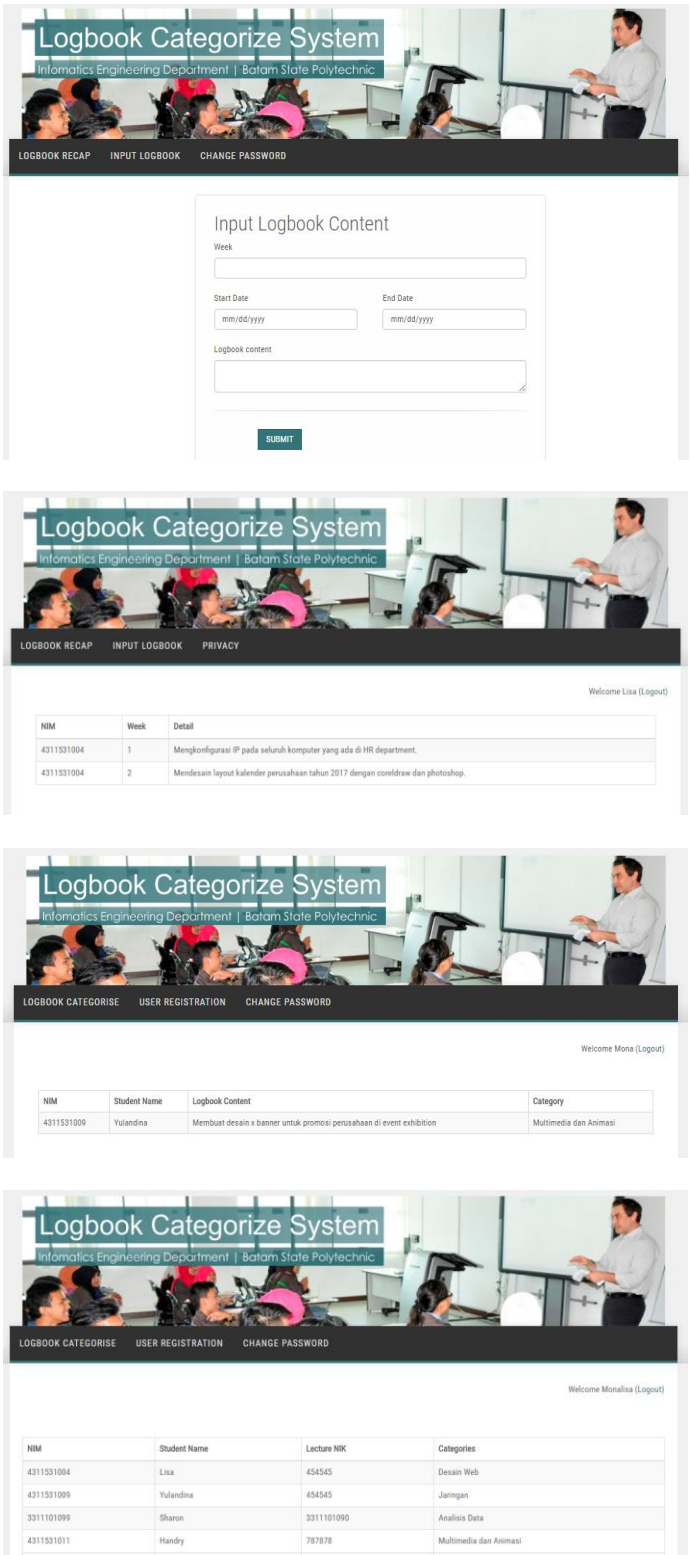

Fig. 5 Implementation Result

From the test results concluded that the application has functioned according to the design. The application can handle the following features:

1. Login user

2. User Registration
3. Input logbook
4. Displays a logbook recapitulation
5. Perform preprocessing and categorization and display it.
6. Logout user
7. Change user password

5. CONCLUSIONS

Based on Analysis, Design and Implementation on Internship Logbook Mining (Illustration in Politeknik Negeri Batam) we conclude that:

1. Applications can categorize logbook content automatically by applying rule-based classifier.
2. The rule based method can be implemented to categorize internship logbook content efficiently.
3. Designed and developed system contains some main feature to support supervisor, and coordinator to summarize student internship logbook content.
4. Student also use this system to submit and create tidy documentation of their work.
5. More keyword to mine the logbook will increase the system performance significantly.

REFERENCES

[1] Feldman, R & Sanger, J. 2007. The *Text mining* Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press: New York.

[2] Hariyanto, B. 2004. Rekayasa Sistem Berorientasi Objek. Bandung: Informatika Bandung

[3] Henderi, Syafei. 2007. Diktat Kuliah UML STMK AMIKOM. STMIK AMIKOM: Bandung.

[4] Juju, Dominikus. 2006. Tip dan Trik Desain Web untuk Pemula. Jakarta: PT. Alex Media Komputindo

[5] Kamus Besar Bahasa Indonesia

[6] Kurniawan, Bambang, Effendi, Syahril, dan Salim Sitompul, Opim. 2012. Klasifikasi Konten Berita Dengan Metode *Text mining*. Universitas Sumatera Utara.

[7] Loudon, Kenneth C, Laudon, Jane P. 2007. Management Information System, 10th ed. Pearson Prentic Hall: Pearson Education, Inc.

[8] Mufida, Mir'atul H. 2015. Garis-Garis Besar Program Perkuliahan Magang Industri Teknik Informatika. Batam: Politeknik Negeri Batam.

[9] Nugroho, B. 2008. Membuat Sistem Informasi Berbasis Web dengan PHP dan MySQL. Yogyakarta: Gava Media.

[10] Tan, Pang-Ning, Steinbach, Micheal dan Kumar, Vipin. 2006. Introduction to Data Mining. Boston: Pearson Education, Inc.

[11] Tim Wahana Komputer. 2010. Mendesain Website Dinamis dan Menarik dengan Adobe Dreamweaver CS4. Yogyakarta: Penerbit Andi.

[12] Tim Wahana Komputer. 2010. Panduan Praktis Menguasai Pemrograman Web dengan JavaScript. Yogyakarta: Penerbit Andi.