

## ANALYSIS OF WINNOWING ALGORITHMS IN THE TITLE SELECTION SYSTEM OF JOB TRAINING REPORT, FACULTY OF ENGINEERING, UNIVERSITAS MALIKUSSALEH

Rizal\*<sup>1</sup>, Defry Hamdhana<sup>2</sup>

<sup>1,2</sup>Department of Informatics. Faculty of Engineering, Universitas Malikussaleh

\*Corresponding Author : [rizal@unimal.ac.id](mailto:rizal@unimal.ac.id)

### ABSTRACT

The number of practical work reports that have been published by department makes a coordinator of on the job training have difficulty in determining the feasibility of the title of the on the job training report quickly. This happened because each report that was made must be unique in terms of the methods and objects that are examined in a particular institution/company. So that reports that have been published previously did not experience plagiarism. Winnowing algorithm is an algorithm used to detect the similarity of words/sentences in two or more texts that are compared. If there are two texts that are the same, a fingerprint will be formed. So that the practical work coordinator can determine the feasibility of the title of on the job training report quickly and the creation of an online-based on the job training report information system that can be accessed at any time.

**KEY WORDS:** Plagiarism, Winnowing Algorithm, On the job training report

### INTRODUCTION

Job Training is a course that must be taken by all students of the Faculty of Engineering, Universitas Malikussaleh. The main goal of students who will carry out Job Training is to be able to train and apply the knowledge that has been learned in campus before entering the workforce. Students are expected to be able to understand non-academic and non-technical concepts in the world of work. So that it can form a good character in terms of responsibility in work and also form a personality that can interact with people in the company/institute.

The Faculty of Engineering is one of the most favorite faculties at Universitas Malikussaleh by prospective students in choosing their majors/study programs. Thus students who have implemented and collected reports on Job Training have counted more than millions of reports distributed in the University Library, Faculty Library, and Library Programs/Study Programs. The Job Training Report is a mandatory report made by students who take Job Training Courses as a condition so that the grades in these courses can be included in the student's Grade Point Average.

Therefore we need a system that can store all information from the job training report in a computerized manner and can also be accessed online. The data can be used to compare the new titles that you want to submit. So

that the similarity of the report title can be immediately known. And this condition will triggers students to be able to develop their reports more completely and specific. With the existence of a system of practical work report information that is embedded in an intelligent system in recognizing the similarity of titles, the coordinator can quickly provide a solution for the student to be able to update the object or method carried out in the agency / company. The other advantages are the creation of a practical work report data bank that can be searched and retrieved quickly and precisely.

Based on the introduction described above, the problem that the writer formulated is:

1. How do you build and implement a web-based information system of on the job training report?
2. How Winnowing Algorithms can recognize the similarity of the job training report titles?

### LITERATURE REVIEW

#### On The Job Training

Job Training is an activity carried out in the community or in the world or to apply methods that are relevant to what is in the community or through the development path by exploring certain fields of science and their applications. According to the guide working for the Faculty of

Engineering, Universitas Malikussaleh, job training is to provide information to students for:

- a. Get work experience before entering the workforce, and obtain a work certificate (reference) from the agency.
- b. Comparing and applying academic knowledge that has been obtained by contributing knowledge clearly and consistently, with high commitment.
- c. More able to understand non-academic and non-technical concepts in the real world of work, including: "superior-subordinate relations", "peer-to-peer relations", "dead-line", "uncertainty and inconsistency of" specifications " which is sometimes not in accordance with academic theory ", etc.
- d. Job Training does not aim to be a substitute for a final project or student project.
- e. Job Training has a weight of 3 (three) credits for all majors / study programs.

### Plagiarisme

Definition of plagiarism according to the Big Indonesian Dictionary (KBBI) "Plagiarism is plagiarism that violates copyright" (KBBI, 2014). Plagiarism comes from the word plagiarism which means taking the essay (opinions and so on) of others and making it as if the essay (opinion and so on) itself, for example publishing other people's writing on their own behalf.

According to RI Minister of Education Regulation Number 17 of 2010 it is said:

"Plagiarism is an intentional or unintentional act in obtaining or trying to obtain credit or value for a scientific work, by quoting part or all of the work and or scientific work of another party that is recognized as a scientific work, without expressing the source appropriately and adequately".

Plagiarism is a form of misuse of intellectual property rights of others, in which the work is presented and recognized illegally as a result of personal work (Sulianta, 2007). So that it can be concluded that plagiarism is an act of taking other people's ideas or writings without mentioning references and claimed as their own.

According to Soelistyo (2011) there are several types of plagiarism, namely:

1. Word by Word Plagiarism (Word for word Plagiarism). The author uses the words of other authors (exactly) without mentioning the source.
2. Plagiarism of sources. The author uses the ideas of others without giving sufficient recognition (without clearly mentioning the source).
3. Plagiarism of Authorship. The author acknowledges as the author of other people's works of writing.
4. Self Plagiarism. Included in this type is the author publishes one article in more than one editorial

publication. And recycle scientific papers. What is important in self plagiarism is that when taking one's own work, the creation of new works produced must have meaningful changes. This means that the old work is a small part of the new work produced. So that the reader will get new things, which the author really put on the paper that uses old works.

There are several ways that can be done to prevent plagiarism according to Indriyanto (2012), namely:

- a. Honesty in a writer, honesty is the basic thing to uphold and build scientific truth.
- b. Recognition of others, the bibliography is an example of someone's appreciation.
- c. Increasing the role of educators in preventing plagiarism, an educator must accompany his students in writing scientific work.

### Winnowing Algorithms

Winnowing is an algorithm used to carry out the document fingerprinting process (Schleimer, Wilkerson, & Aiken, 2003: 4). The fingerprinting document is a method used to detect the accuracy of copies between documents or only part of the text. The working principle of this document fingerprinting method is to use hashing techniques. The hashing technique is a function that converts each string to a number.

The Winnowing algorithm calculates the hash value of each k-gram, to find the hash value then the hash rolling function is used. Then a window is formed from the hash values. In each window a minimum hash value is selected. If there is more than one hash with the minimum value, the rightmost hash value is selected. Then all selected hash values are stored as fingerprints of a document. The input of the fingerprinting document process is a text file. Then the output will be a set of hash values called fingerprint. This Fingerprint will be used as a basis for comparing the similarities between the texts that have been entered.

The requirements for an algorithm to detect plagiarism according to Schleimer, Wilkerson and Aiken (2003) are:

- a. Whitespace insensitivity, which is matching text files should not be affected by spaces, capital letters, punctuation and so on.
- b. Noise suppression, avoid matching text files with words that are too small or less relevant and are not commonly used words.
- c. Position independence, that is, matching text files should not depend on the position of words so that words in different position sequences can still be recognized if similarities occur.

Following are the steps of the Winnowing Algorithm in detecting text similarities according to Sariyanti, Andharini and Sophan (2014):

1. Removal of irrelevant characters (whitespace insensitivity), including spaces or punctuation.
2. Formation of gram series with size k.
3. Calculation of the hash value of each gram.
4. Divide into certain windows.
5. Selecting multiple hash values becomes document fingerprinting.
6. Determine the similarity percentage between 2 documents with the Jaccard Coefficient equation

## DISCUSSION

### The Proposed Scheme

In this section, I want explain the system scheme about recognition of the similarity of job training report titles.

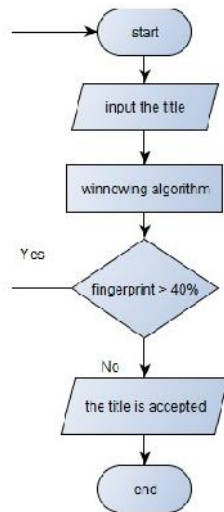


Figure 1: System Scheme

The first step that must be done is that the student must register the data first. The following is a picture of the registration form that must be filled out by students before submitting the title of the practical work report.

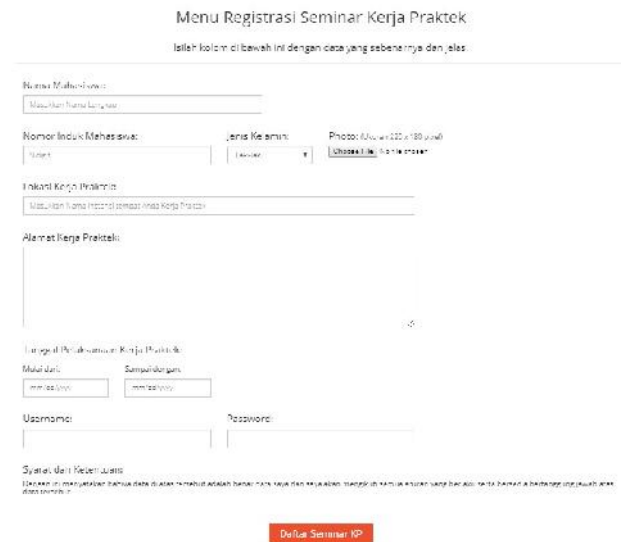


Figure 2: Display of Student Registration Form

After data registration, new students can log into the system to submit a report on the practical work report. Students cannot register if the Student Registration Number (SRN) entered is registered in the database. Students have the opportunity to submit 3 practical work report titles to the system. This is done so that the practical work coordinator can choose the title that falls into the criteria. If the three titles have not been selected by the practical work coordinator, then the title status is still worth Waiting for. And if the practical work coordinator has chosen one of the three titles that the student submits, then the status will change to Accepted.



No	Judul	Dosen Pembimbing	Tahun	Status
1	cobacoba dulu	Muhammad Fikry, S.Kom., M.Kom	2017	Menunggu
2	acara praktek	Dicky Murnidhanu, S.T., M.Kom	2014	Ditunggu
3	lirik puisi ya	Muhammad Fikry, S.Kom., M.Kom	2011	Menunggu

Figure 3: Menu List Submission of Job Training Seminar

Based on the research that has been done on the information system Selection Title Job Training Report, the introduction of the similarity of the title of the practical work report can work well. The system will deactivate the title submission button if the similarity of the title submitted is more than 40% compared to the database reporting the work title practice. Thus the student must

rethink a better title so that the title can be submitted to the system.

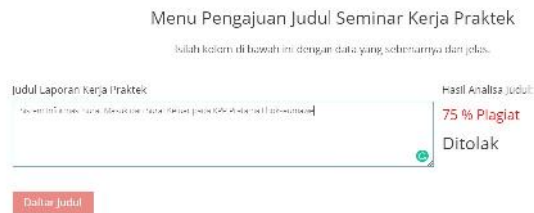


Figure 4: Submission of Report Title Refused Job Training

## CONCLUSION

Winnowing Algorithm is an algorithm that is used to process document fingerprinting to identify plagiarism. The algorithm used to search for hash values in winnowing is rolling hashes. Hash value is a numeric value formed from ASCII calculations for each character.

## ACKNOWLEDGMENTS

Funding: This work is supported by the Faculty of Engineering, Universitas Malikussaleh [Year of Grant 2018]; and research laboratory majoring in informatics, Universitas Malikussaleh.

## REFERENCES

- Connolly, Thomas and Begg, Carolyn. (2010). *Database Systems A Practical Approach to Design, Implementation, and Management Fifth Edition*. Boston: Pearson Education
- Permenpan. (2010). Pencegahan dan Penanggulangan Plagiat di Perguruan Tinggi. *Peraturan Menteri Pendidikan Nasional Nomor 17 Tahun 2010*
- Schleimer, S., Wilkerson, D. and Aiken, A. (2003) Winnowing: Local Algorithms for Document Fingerprinting. *SIGMOD 2003*, San Diego, 9-12 June 2003, 76-85.
- Soelistyo, H. (2011). *Plagiarisme: Pelanggaran Hak Cipta dan Etika*. Yogyakarta: Penerbit Kanisius