Empirical Analysis on Chinese Academic Plagiarism

Yang Shen[1], Huijuan Fu[1], Zitao Liu[2], Pengpeng Liu[1], Qingchuan Fu[2, 3]

(1: School of Information Management, 2: International School of Software, Wuhan University,

3: State Key Laboratory of Software Engineering, Wuhan University, China, 430072)

yshen@whu.edu.cn, fuhuijuan19872006@163.com, hnly228078@qq.com, sdbzlpp@gmail.com,

ifuqingchuan@qq.com

ABSTRACT

This poster, from angels of subjects, authors' social network, authors' combination, and students' plagiarism law, apply self-developed ROST Anti-plagiarism Software to check 3781 papers, do a survey among 450 students, quantitatively analyzed academic plagiarism conditions in China, and draw several conclusions.

Categories and Subject Descriptors

H.3.7 [Digital Libraries] *Systems issues, user issues*; H.3.3 [Information Search and Retrieval] *Selection process, Clustering, Search process.*

General Terms: Measurement

Keywords: ROST Anti-plagiarism Software, Plagiarism Law

INSTRODUCTION

It is a trend to embed anti-plagiarism system to digital library, but there is no widely recognized Chinese anti-plagiarism system. So, the authors believe the empirical analysis of China academic plagiarism makes both academic and social sense.

1. ROST Anti Plagiarism System

ROST anti-plagiarism software automatically cuts texts into small blocks of 200(sizable) Chinese letters, fuzzy matches the blocks with 18.8 billion web pages and 4.9 million literature. Than, we can collect the statistics of the proportion for the similarity degree $\geq 95\%$ (directly plagiarism), and $\geq 80\%$ (modified plagiarism). We consider this proportion as plagiarism (similarity) degree measurable indicator, and conclude paper similarity degree by proportion analysis.

2. RESULT OF ALL RESEARCH

First, through ROST anti-plagiarism software, we sample and check 2250 papers from China Journal Net, 450 papers in each big five academic fields. We find that plagiarism degree, from high to low, are agriculture, social and human sciences, natural science, medicine, and engineering technology.

Second, we gather statistics of the cooperative frequency between authors in journal and communication area and divide it as high, medium, and low three levels. We sample and check 1430 papers in each level. The authors' social relationship is inversely related to plagiarism rate.

Third, we check 101 papers from the 2nd International Conference on Publishing Industry and Publishing Education in the Digital Library, and divide authors into four types, professors, doctors, masters, and undergraduates. It shows cooperation among students' plagiarism rate is obviously higher than between teachers and students, and cooperation among different author types' rate is higher than among the same type.

Forth, we do a survey in Wuhan University with 450 samples, and find that Science and engineering students and lower grades students plagiarize less than liberal arts students and higher grades students. Also, higher grades students do more indirectly modified plagiarism.

3. Acknowledgement