Author Identification Using Natural Language Processing

自然言語処理を用いた筆者特定

S1260227 Kento Miura Supervisor John Blake

1. Outline of the research

(Motivation/background and Goal)

・My research was motivated by my curiosity

to know how much anonymity can be

guaranteed on the Internet.

・The purpose of this research is to develop

useful tools for Digital Forensics and to

improve their functions and UI.

・My task is to improve the accuracy of

author matching and to evaluate the usability

of the website.

・Frequency of word usage

- We will examine the frequency of word

usage in two prepared documents.

By examining the trends,

we will try to identify the authors.

・Readability

- We measure the readability (decomposition

of words by lexical level) of two prepared

documents.

To achieve this function, we compare the

vocabulary levels of the prepared samples

and The similarity rate is calculated

according to the distribution of

the vocabulary of the documents

to be compared.

1. Approach/Methodology

・Tools used: some web frameworks

(maybe Vue or Svelte) [1], Regex [2], Bash [3],

Python [4].

・First, some web frameworks was used to create

e a basic prototype of the author specific site

・Create multimedia resources and regular

expressions.

- I create multimedia resources and regular expressions to add.

・While I make the website easy to use

I will make the website easy to use, while doing

usability studies with some students.

3.Current Results and Status

・I made a prototype of a readability calculator and a word frequency calculator.

-The prototypes were created using Bash and python.

- The frequency calculator needs to be further improved.

- I would like to try using this tool in practice.

- Make the design easy to read

Work to make the UI easier to understand.

- I understand the basic regular expressions, but

I need to improve it a little more.

4.Remaining Tasks and Tentative Schedule

Things to do

- Detailed implementation of word frequency functions

- Implementation of web design

- Conducting usability studies

- Improving from usability studies

- Writing a Graduation Thesis

• Gantt chartグラフ

自動的に生成された説明

References

[1]Man page of Bash

<https://linuxjm.osdn.jp/html/GNU_bash/man1/bash.1.html>

[2] Python for NLP: Developing an Automatic Text Filler using N-Grams

<https://stackabuse.com/python-for-nlp-developing-an-automatic-text-filler-using-n-grams/>

[3]List of Regular Expressions

<http://gimite.net/help/devas-ja/all_regex.html>