# Authorship Detection: Who Wrote the Letter to the Hebrews?

Project by Jie Dong, Leng Yang Tan, Tien-en Joel Phua

Supervised by Prof. Derek Abbott  Co-supervised by Dr. Brian Ng

SCHOOL OF ELECTRICAL AND ELECTRONIC ENGINEERING

THE UNIVERSITY of ADELAIDE

## 1. Controversy

George Miligan (1860 ~ 1934)

"It was unlikely Paul was the author of Hebrews because the anonymity of the letter was not consistent with Paul's pattern"

Clement of Alexandria (c.150 ~ c.215)

"Paul was likely the author of Hebrews originally and not the Greek Version. The latter is the work of Luke who translated Paul's letter"

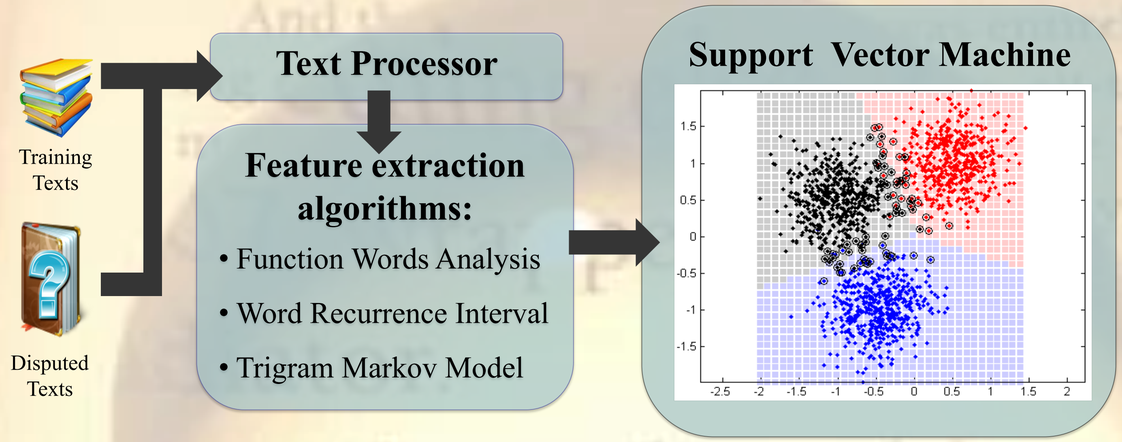
## 2. Introduction

The New Testament of the Bible contains a number of text that have disputed or unknown authorship. One of the texts that has been widely debated would be the Letter to the Hebrews. Three feature extraction algorithms with Support Vector Machine were implemented to develop a classification model. The results obtained would help in identifying the rightful author of Letter to the Hebrews.

## 3. Project Objective

The objective of this project is to identify the author of the Letter to the Hebrews. This project will provide a non-biased approach in authorship attribution. It will also implement the use of the Support Vector Machine to develop a classification model and analyse the use of different kernel functions in authorship attribution.

## 4. Approach




### Function Word Analysis

It is the analysis of the percentage of a set of function words used in a particular text.

### Word Recurrence Interval

It is the number of words in between successive amount of a set of chosen keywords.

### Trigram Markov Model

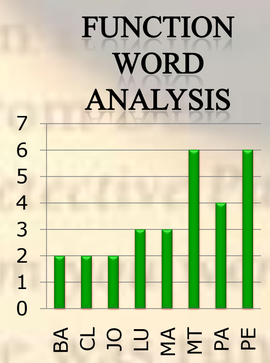
It is the probability of the next word appearing after the previous two words.

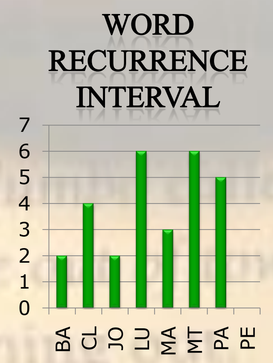
### Support Vector Machine

It is a classifier that uses hyperplanes to divide a set of data items into groups according to the feature vectors.

## 5. Result

| Authors | Abbrs. |
|---|---|
| Barnabas | BA |
| Clement | CL |
| John | JO |
| Luke | LU |
| Mark | MA |
| Matthew | MT |
| Paul | PA |
| Peter | PE |





## 6. Conclusion

The feature extraction algorithms together with the Support Vector Machine were able to accurately attribute a disputed text to its rightful author. However, the algorithms require a sufficiently large number of training data in order to make the classification accurate. In the case of the New Testament texts, only a small set of training data was available. This limits the accuracy of the classification model. In this case, the algorithms were only able to accurately eliminate unlikely authors from the list.

## 7. Future Applications

Plagiarism Analysis


Search Engine


Biological Analysis

## References

[1] Putnins, T. J., Signoriello, D. J., Jain, S. Berryman, M. J., & Abbott, D., Who wrote the Letter to the Hebrews? Data mining for detection of text authorship, University of Adelaide, 2005

[2] Putnins, T. J., Signoriello, D. J., Jain, S. Berryman, M. J., & Abbott, D., "Advanced text authorship detection methods and their application to biblical texts", Proc. SPIE: Complex Systems 6039 ed. Axel Bender, Brisbane, Qld., Australia, December 11-14, 2005

[3] Baayen, H., Halteren, H. V., Neijt, A. & Tweedie, F., An experiment in authorship attribution, 6th JADT, 2002