

## Authorship Detection

Minutes for 10<sup>th</sup> meeting on 10<sup>th</sup> March 2011

Held in Innova Lunch Area third floor, 4.00PM – 5.00PM

Participants: Derek, Brian, Maryam, Clement, Joel, Jie Dong

Absent: None

### Minutes

#### 1. Project status/progress during last week

- a) Table results were made for each algorithm with the use of SVM.
- b) Have an interface to combine the algorithm.
- c) Derek:
  - a. suggestions on removing punctuation:
    - i. if is . , ; --> remove space
    - ii. if is : ; , - '--> just remove
    - iii. could easily determine by checking if there's a space or not
    - iv. example: 'hello' --> hello, hand-craft --> handcraft
    - v. also check that the algorithm only leaves ASCII keys behind.
    - vi. heading and numbers in the text should be removed as it is not really a word, rather it is just data.
    - vii. there could be other ASCII keys that are use to represent greek letters, need to check the raw data of greek texts.
    - viii. the low probability of the SVM could probably be caused by the short texts.
- d) Brian:
  - a. don't get stray away trying to achieve 100% accuracy, need to start testing on "Letter of Hebrew"
  - b. don't do the data extraction of both training and test data at the same time. this cause incorrect result.
  - c. Keep the data first and compare with the result after implementing the suggestions provided by Brian and Derek.

#### 2. Project goals for upcoming week

- a) Clement:
  - i. Discuss with Joel on using a constant word length for each text.
  - ii. implement interface developed by Joel.
  - iii. modify the interface for delete punctuation.
  - iv. try changing the number of appearing keywords that is chosen. try 5, 10, 15, 20, 25.
  - v. start using "letter of hebrew" to test.
- b) Joel:
  - i. Discuss with Clement on using a constant word length for each text.
  - ii. start using "letter of hebrew" to test.
  - iii. check the result of word function using SVM.

c) Dong Jie

- i. modify trigram algorithm.
- ii. start using "letter of hebrew" to test.
- iii. Apply interface developed by Joel.