Authorship Detection

Minutes for 12th meeting on 07th April 2011
Held in Innova Lunch Area third floor, 4.00PM – 5.00PM
Participants: Brian, Maryam, Clement, Joel, Jie Dong
Absent: Derek

1. **Project status/progress during last week**
   - the functions of each algorithm was described again to refresh the memory.
   - English text was verified to be all right and it was the encoding problem of the notepad.
   - Trigram Markov
   - Word Recurrence Interval
     - Results was obtained using the Federalist Text.
     - 15 training data, 12 disputed text which suppose to be written by Madison.
     - Results shown was not favourable. The prediction was very inconsistent, achieving a low accuracy rate of 53% most of the time.
     - it was suggested that the testing could be bias to Madison as only Madison text was taken as testing data.
     - Comparing with the earlier results using the English text, which involves 100 training data and 70 disputed text, the accuracy and consistency was even much lower.
     - WRI might not be suitable for authorship detection.
     - It was suggested to combine the Function Word Frequency developed by Joel to enhance the algorithm.
   - Function word Frequency

2. **Project goals for upcoming week**
   a) Clement:
      - do another testing.
      - implement enhanced version of the WRI.
      - prepare powerpoint slides for the seminar.
      - start to make the video.
   b) Joel:
      - do another testing.
      - implement enhanced version of the WRI.
      - prepare powerpoint slides for the seminar.
      - start to make the video.
   c) Dong Jie
      - do another testing.
      - implement enhanced version of the WRI.
      - prepare powerpoint slides for the seminar.
      - start to make the video.