

## 1. Aims of the Project

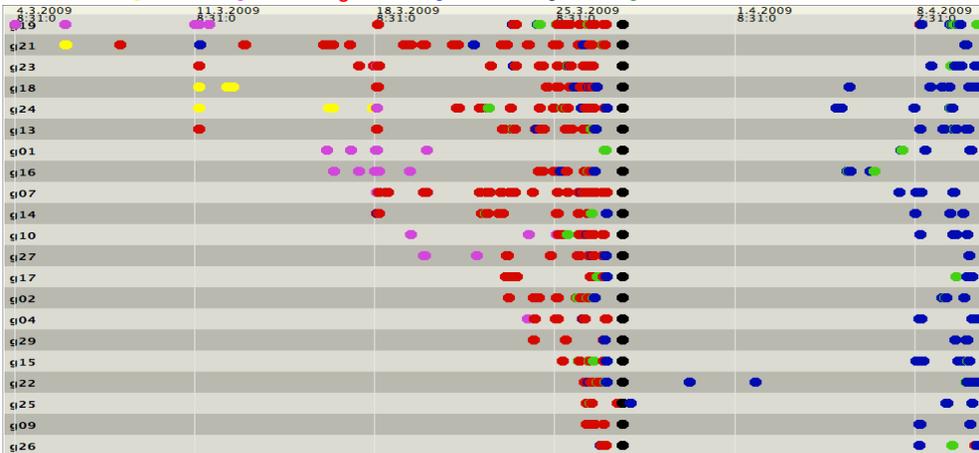
- Develop a new framework for extracting and analysing collaborative writing process.
- Investigate and identify techniques to extract the semantic meaning of text changes during collaborative writing, based on semantic text mining.
- Examining and developing techniques to discover sequence patterns of writing activities that lead to successful outcome and those that may lead to problems, based on semantic process mining.
- Develop a visualisation of feedback on writing process for end-users.

## 2. Introduction

The availability of the Internet has made collaborative writing very easy to implement in schools and at work. This leads to new forms of writing, such as blogging and wiki writing. In addition, the emerging of "cloud computing" tools and Web 2.0 applications such as Google Docs has led to the availability of almost desktop-quality online writing environments. However, despite the widespread of such tools supporting collaborative writing, there is very little support to improve the quality of the process of collaborative writing. Our overall aim is to creating adequate support for these writing activities for improving not only the quality of documents, but also the collaborative writing skills of the authors involved.

There has been abundant research applying data and text mining techniques in order to extract semantic aspects of text and documents. However, this research focuses on the final product of writing (i.e. documents), rather than the process of writing (i.e. writing activities). We are interested in the latter.

Brainstorming    Outlining    Drafting    Revising    Reviewing    Editing

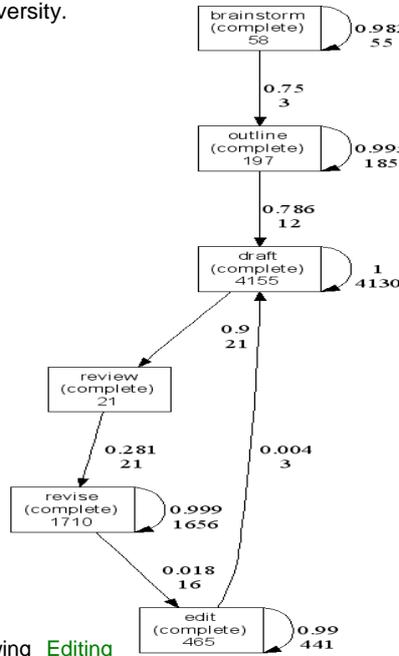


## 3. Extracting Semantic Meaning of Writing Activities

Based on taxonomy of collaborative writing, proposed by Lowry et al., we have six main stages: *brainstorming, outlining, drafting, reviewing, revising and editing*. We created a set of heuristics which map the combination of text change operations, topic changes, and a topic cohesion measure (using Latent Semantic Analysis (LSA)) to stages of writing activities.

## 4. Pilot Study

As a way of evaluating the proposed heuristics and illustrating how data and process mining can be used to analyze writing activities, we conducted a pilot study to investigate writing processes of students in an engineering course taught at our university.



## 4.1 Writing Process Analysis

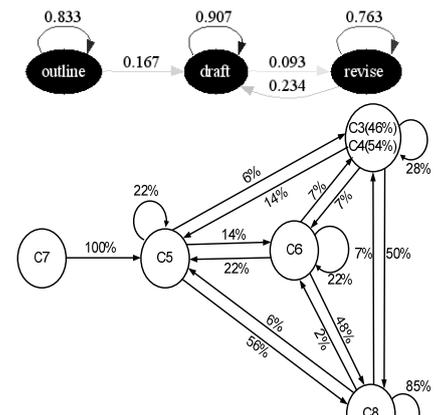
We used ProM to derive, the following:

- Historical snapshot of writing activities.
- Process model of collaborative writing.
- Students' performance sequence and collaboration during writing.

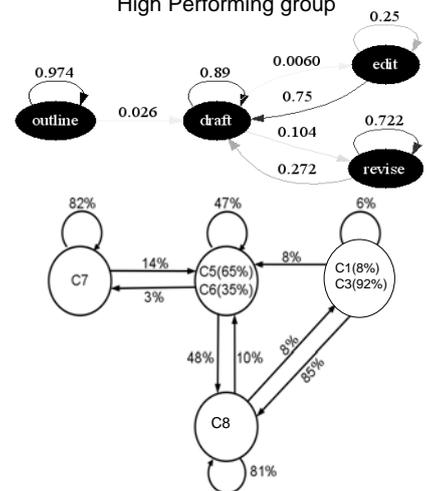
## 4.2 Process Model Using Hidden Markov Models

We also developed two models, a Heuristic Markov Model (Heuristic MM) and a Hidden Markov Model (Hidden MM) for analysing student's writing behavior:

- A Heuristic MM is a Markov model created from a sequence of writing activities, derived from text change operations by applying the semantic heuristics.
- A Hidden MM is built directly from the sequence of text change operations. The Hidden MM can extract writing states, which can not be directly measured, as well as the transitioning probabilities among these states.



High Performing group



Low Performing Group

## 5. Future Work

- Creating a novel modelling representation that takes into account not only events but duration of and between writing activities.
- Exploring the visualization of this new model as a mean of feedback to students and teachers.

