# Data Mining for Individualised Hints in e-Learning

*Author: Anna Katrina Dominguez, adom0244@it.usyd.edu.au*
*Supervisor: Dr Kalina Yacef and Dr James Curran*
*School of Information Technologies*

## 1.Introduction

In this project, we aimed to provide hints to students who are learning to program using an existing online e-learning system.

We generated these hints by employing Data Mining techniques both to analyse past data and as part of the live system.

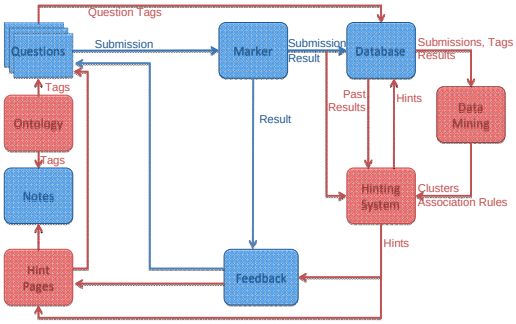
## 2. Problem Statement

The NCSS Challenge is an annual online programming competition designed to teach Python. Students are given notes and then presented with questions for which they submit code.

Often, students who are struggling have little access to help, and find it difficult to apply what they learned from the notes to the questions. We therefore added individualised hints to aid students.

## 3. Data Mining

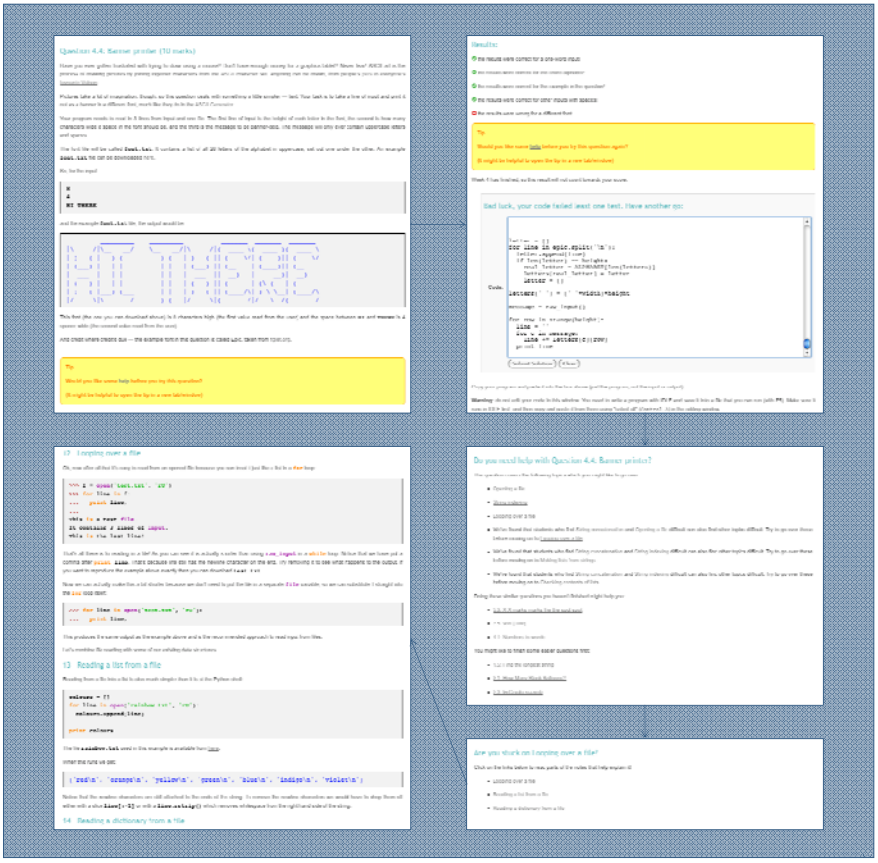
We used the TADA-Ed tool [1] for clustering students and questions and for Association Rule Mining to find common sequences of errors, an idea adapted from Merceron and Yacef [2]. However, our system not only used patterns mined from past data, but mined and integrated data sourced from live usage of the system, a concept which has not been explored by previous research [3].




## 4. Implementation

To add the hints to the existing Challenge system, we created an ontology of basic Python topics and labeled the 2008 and 2009 questions and notes with these topics.

Our system provides hints to students in two forms – as suggestions of other questions to attempt, and as topics which they may like to go over. Since we had labeled the notes with topics, we were able to provide links to relevant sections of the notes for each topic.



## 5. Experiment

We conducted a 5 week experiment with 584 of the 2009 Challenge students, half of whom received hints. Students were clustered into 'Good', 'Medium' and 'Bad' groups weekly, based on their results. We aimed to determine if there was a positive difference between students who were and were not provided with hints.

## 6. Student Feedback

We asked students to provide their opinions of the system, and found many students viewed the hints positively.

*"i found the tips more helpful because when we are using the notes to solve the problem we really don't know where to go and what to do… the hint boxes were a very smart way to access the notes that can help us to solve the problems."*

## 7. Quantitative Results

Students who received the hints had a mean mark of 32% while those without had a mean mark of 40%. This showed to be a significant difference by an Approximate Randomisation test ($p < 0.01$).

## 8. Conclusion

In this project, we implemented a hinting system for students in the NCSS Challenge learning system. We based our hints on patterns mined from both past and current data, and found both numerically and qualitatively that our hints were able to help students who were participating in the Challenge in 2009.

## 9. References

[1] C. Romero and S. Ventura. 2007. Educational data mining: A survey from 1995 to 2005. Expert Systems with Applications, 33(1):135–146.

[2] A. Merceron and K. Yacef. 2005b. Tada-ed for educational data mining. Interactive Multimedia Electronic Journal of Computer-Enhanced Learning, 7(1):267–287.

[3] A. Merceron and K. Yacef. 2003. A web-based tutoring tool with mining facilities to improve learning and teaching. In Proceedings of the 11th International Conference on Artificial Intelligence in Education, pages 201–208. IOS Press.

The University of Sydney