Piloting e-Assessment of Diagrammatic Coursework

CIF 2008/09 Project Completion Report

Ambikesh Jayal and Prof Martin Shepperd
School of IS, Computing & Maths, Brunel University, United Kingdom
Email: {ambikesh.jayal, martin.shepperd}@brunel.ac.uk

Project Name	e-Assessment of Diagrammatic Coursework
Funding Institution	Curriculum For Innovation Funded Project 2008/09, Learning and Teaching Unit (LTDU), Brunel University
Report created by	Ambikesh Jayal, Prof Martin Shepperd, School of IS, Computing & Maths, Brunel University
Project output website	https://sourceforge.net/projects/eassessmentofd/files/
Project output	1. Software for e-assessment of diagram(Open source) 2. Evaluation results 3. Corpus of coursework for further experiments
Description	The project deals with improving the undergraduate assessment by using automated marking for diagrammatic coursework.
Report submission date	14/04/10

Acknowledgements

We would like to thank Professor Ken Darby-Dowman, Prof Rob Macredie (pro Vice Chancellor), Richard Eccles (e-Learning Technologist), Dr Jacques Furter (Acting Deputy Head of School - Learning and Teaching), Dr. Kate Dunton (Education Officer LTDU), Dr Tracy Hall (Reader and module leader CS2077A), Dr Peter Thomas, (Senior Lecturer, Open University) and Selwyn Lloyd (Director, Phosphorix Limited) for their support and constructive feedback throughout the project.

Table of Contents

1. Introduction

This project deals with improving undergraduate assessment by using automated marking for diagrammatic coursework. Automated marking of diagrammatic coursework (e-Assessment), offers many potential benefits including timely feedback, consistency of marking and freeing up of valuable lecturer time. Unfortunately, until recently, e-Assessment of diagrams has proved too challenging because of imprecise semantics, the need to make sense of incomplete or only partially correct solutions and the possibility of a multiplicity of correct solutions. Yet, the use of diagrams is quite commonplace in many disciplines such as Computer Science, Engineering and Design. This research is important because quality and timeliness of feedback is an area about which the students have been consistently critical, so e-Assessment represents a very real opportunity to improve the Brunel student experience.

As part of this project we have developed a software tool for automated marking of diagrammatic coursework and evaluated it on a second year undergraduate module at the Brunel University. This software is customizable, plug and play and extensible. The evaluation results show that the software is effective but further work is required to perfect the accuracy.

The software tool and corpus of coursework are published in sourceforge under the project 'e-Assessment of Diagrammatic Coursework'. The link for this project is as follows.

https://sourceforge.net/projects/eassessmentofd/files/

2. Aims and objectives of the project

There were four objectives and associated work packages to the project:

1. Develop the system as a demonstrator to semi-automate the marking process of diagrammatic coursework and package it as an open-source system

2. Evaluate it on a coursework at Brunel.

3. Package the results as a case on the web with software documentation.

4. Publish research papers and develop a larger research proposal suitable for external funding

We discuss each workpackage outcome in detail in the following sections.

2.1 Developed Open Source Software for e-assessment of diagrams

An open source software tool developed as part of this project and is therefore freely available for use by other researchers and educators. This software is customizable, plug and play and extensible. The software can be configured for various algorithms and parameters using the configuration XML file provided with the software. All the source code of this software has been provided as a compressed zip file. Note that some screen shots of the tool have been provided in the appendix. It consists of five components:

1. DiagramAssessmentTool.jar

   This is the main software tool and uses the other four components mentioned below. Each component is a separate jar file.

2. UMLDiagramXMIAPI.jar

   This component extracts the labels from UML diagrams in XMI format.

3. GenericLabelMatcher.jar

   This component matches the labels.

4. GenericLabelMatcherConcreteClasses.jar

   This component has the concrete implementations of the algorithms used by the GenericLabelMatcher component.

5. GenericLabelMatcherInterface.jar

   This component has the interface for the algorithms used by the GenericLabelMatcher component. The user needs this component containing interfaces in order to compile the concrete implementation of the algorithms.

2.2. Evaluated software on real coursework at Brunel University

The e-assessment software was evaluated on 189 items of 2nd year undergraduate coursework. The marks awarded by the software marks were compared against the marks awarded by the human expert. The maximum mark is 5. Tables 1- 3 show the result of this evaluation.

As shown in Table 1, the mean of difference between machine and human marks is 0.74. Since these marks are out of 5, so on an average the machine has awarded 15% more marks than the human. Table 1 also shows that the standard deviation of machine marks is 1.35 as compared to 0.94 for the human marker, implying that the software has been more inconsistent than human.

Table 2 shows the correlation between the human and machine marks. These results show that the correlation is statistically significant.

Table 3 shows the frequency of various values for the difference between machine and human marks. Figure 1 shows these results graphically. The first row in Table 3 show that for 15.9% of students (30 students) there was no difference between the human and machine marks. Similarly the fifth row in Table 3 show that for 23.8% of students (45 students) the difference between the human and machine marks was 10 % (or 0.5 marks out of 5). Table 3 shows that for 23.8% of students (45 students) the difference between the human and machine marks was 10% (or 0.5 marks out of 5).

Overall the evaluation results show that the software is effective but further work is required to enhance the performance. The main difficulty is in dealing with all the synonyms that students use to label their diagrams, that is synonymous with the labels used by the lecturer in the specimen solution used to judge student solutions.

Statistics

		DIFFERENCE BETWEEN MACHINE AND HUMAN MARKS	MACHINE MARKS	HUMAN MARKS
N	Valid	189	189	189
	Missing	0	0	0
Mean		**.740079**	**3.163360**	**3.021164**
Median		.500000	2.500000	3.000000
Mode		.5000	2.5000	4.0000
Std. Deviation		**.5889757**	**1.3525796**	**.9422566**
Variance		.347	1.829	.888
Skewness		.906	-.151	-.303
Std. Error of Skewness		.177	.177	.177
Range		**2.5000**	**5.0000**	**5.0000**

Table 1: Mean

Correlations

		MACHINE MARKS	HUMAN MARKS
MACHINE MARKS	Pearson Correlation	1	.722(**)
	Sig. (2-tailed)		.000
	N	189	189
HUMAN MARKS	Pearson Correlation	.722(**)	1
	Sig. (2-tailed)	.000	
	N	189	189

** Correlation is significant at the 0.01 level (2-tailed).

Table 2: Correlation

Frequency Table

DIFFERENCE BETWEEN MACHINE AND HUMAN MARKS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.0000	30	15.9	15.9	15.9
	.1250	4	2.1	2.1	18.0
	.2500	10	5.3	5.3	23.3
	.3750	11	5.8	5.8	29.1
	.5000	45	23.8	23.8	52.9
	.6250	5	2.6	2.6	55.6
	.7500	4	2.1	2.1	57.7
	.8750	4	2.1	2.1	59.8
	1.0000	41	21.7	21.7	81.5
	1.2500	3	1.6	1.6	83.1
	1.3750	3	1.6	1.6	84.7
	1.5000	17	9.0	9.0	93.7
	1.7500	1	.5	.5	94.2
	1.8750	2	1.1	1.1	95.2
	2.0000	3	1.6	1.6	96.8
	2.3750	2	1.1	1.1	97.9
	2.5000	4	2.1	2.1	100.0
	Total	189	100.0	100.0	

Table 3: Frequency

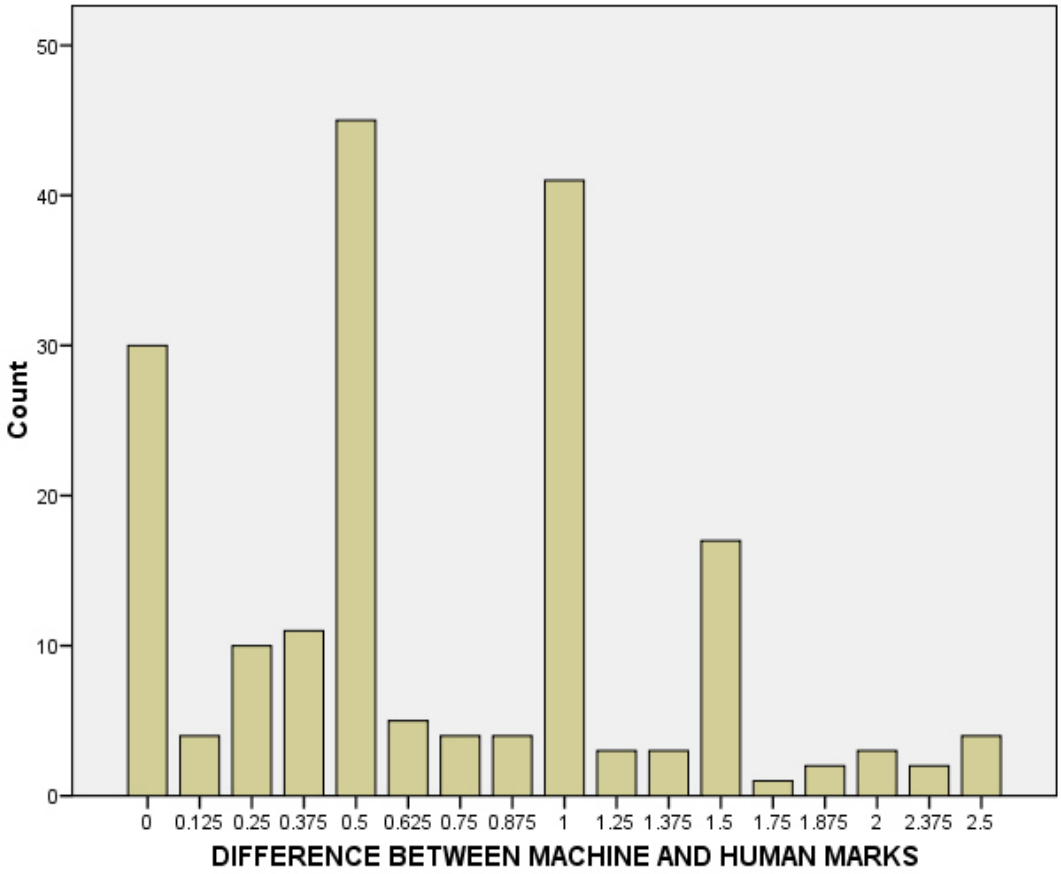


Figure 1: Frequency

2.3. Provided corpus of coursework for experiments

To evaluate the e-assessment software, corpus of coursework was collected from a second year Computer Science undergraduate module at Brunel University. The coursework description consisted of one paragraph of text explaining the process of students eating an evening meal and the students were required to draw an UML activity diagram for this problem. This was an inclass paper based test. A total of 189 student papers were collected and scanned to JPEG image. The labels used by the students in the coursework diagram were manually extracted and used for evaluation. The question text, model solution and a student diagram has been provided in the appendix section of this report.

This corpus of coursework was used to evaluate the e-assessment software developed during this project. The results of this evaluation are presented in the next section.

This corpus of coursework has been made available at sourceforge. It can be used by the research community for further experiments.

2.4. Project publications

The following two publications have resulted from this project.

Jayal, A. and Shepperd, M. (2009). "The Problem of Labels in E-Assessment of Diagrams". *ACM J. Educ. Resour. Comput.*, 8(4), pp1-13, 2009.

This work has already been cited twice by research groups other than ourselves.

Jayal, A. and Shepperd, M. (2009). An improved method for label matching in e-assessment of diagrams. *Electronic journal of the UK Higher Education Academy Subject Centre for Information and Computer Sciences (ICS)*

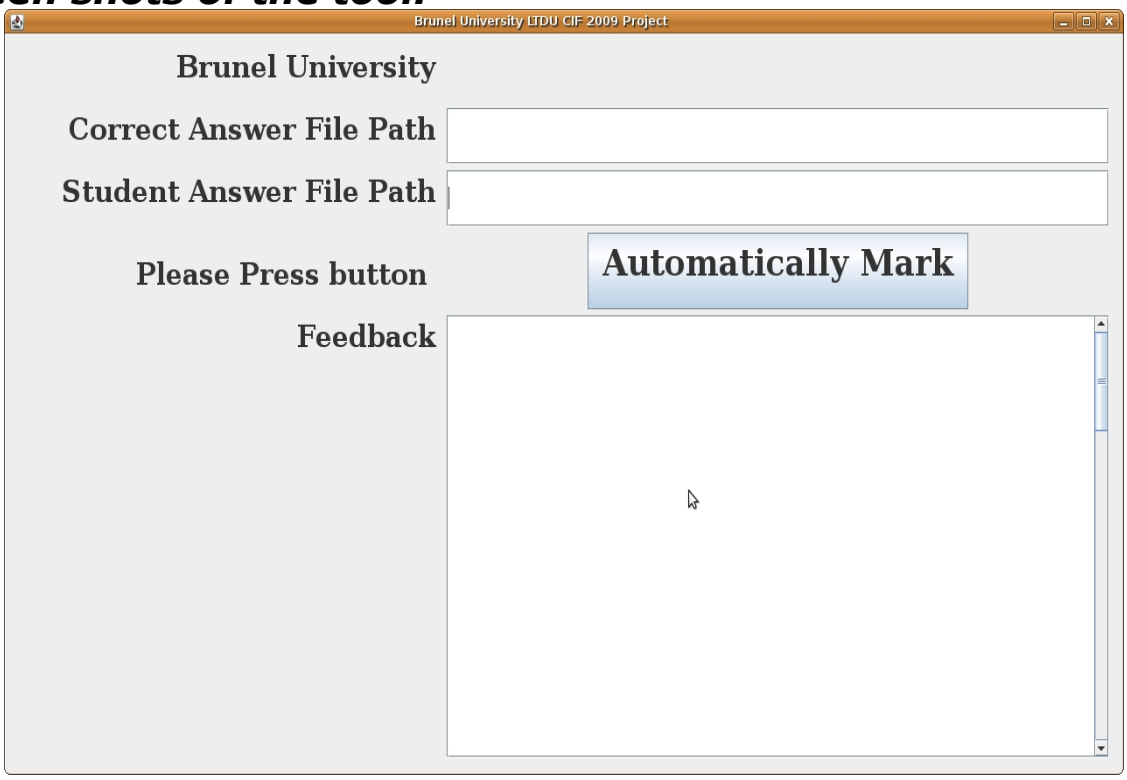
3. Project benefits and Future Work

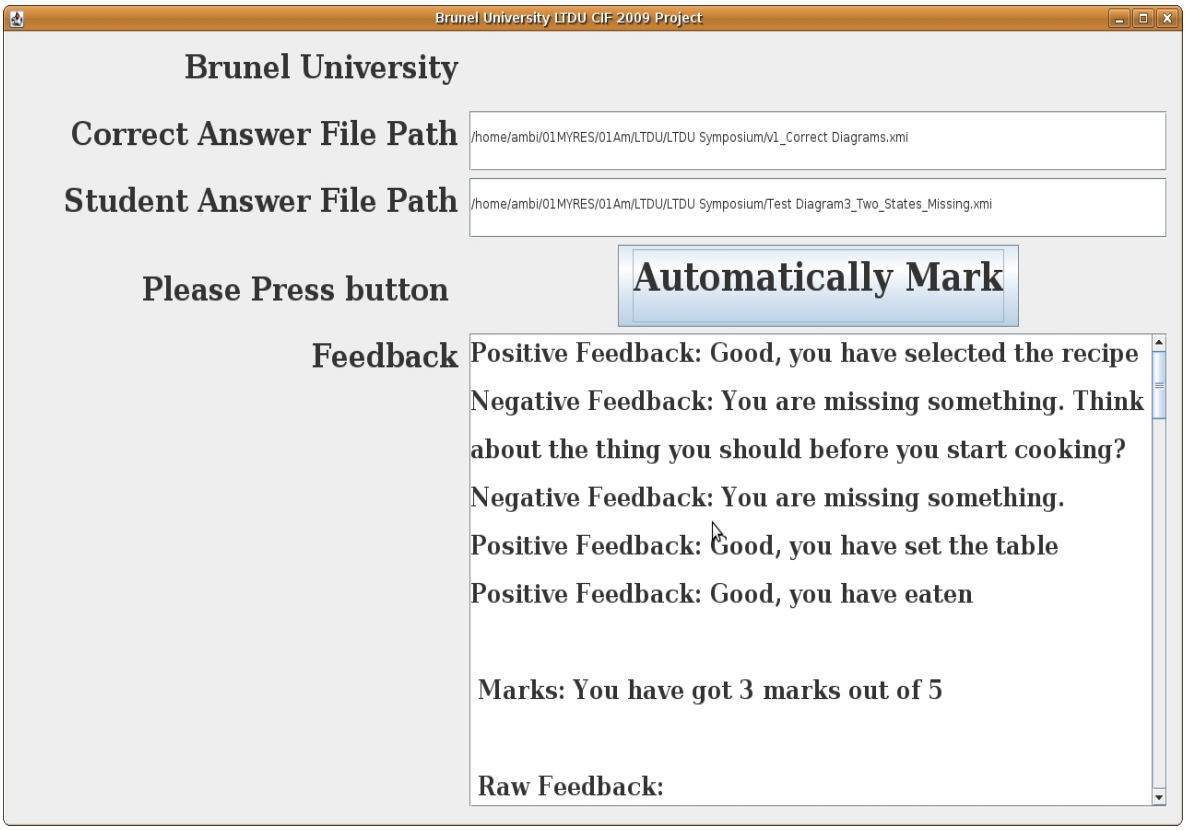
Automated marking is clearly beneficial since it provides students with rapid feedback and relieves academic staff of what can be a very burdensome task. Such systems are widely available including in the Brunel VLE (U-Link). However, these tend to focus on very structured coursework typically with a fixed repertoire of responses such as multiple choice. Marking diagrams automatically is particularly challenging due to the problems of dealing with partially correct and incomplete solutions plus the difficulties of analysing the labels which is a natural language processing problem. Notwithstanding these difficulties we have made substantial progress, produced a general purpose framework, software and a corpus of coursework that might be used by other researchers and educators.

As an extension of this project, two research proposals were submitted for external funding to JISC and HEI but unfortunately neither were successful. However the feedback we have received from JISC is encouraging and we intend to work on some future bids.

Appendix

Screen shots of the tool.



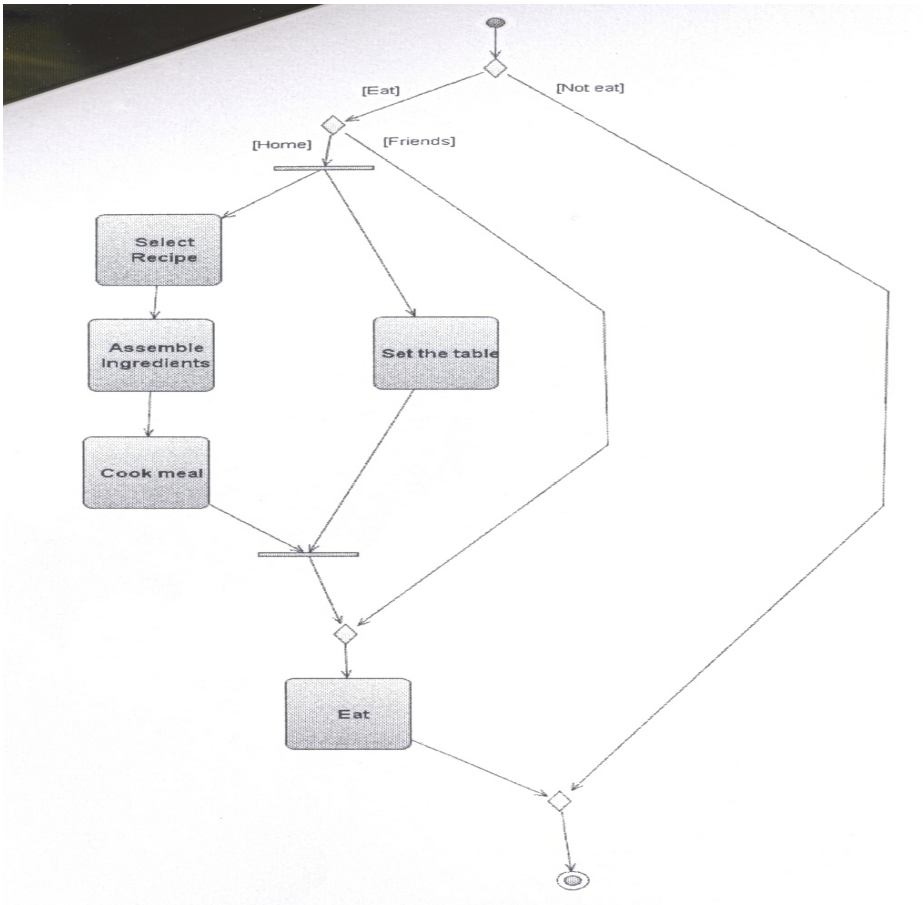

Brunel University LTDU CIF 2009 Project
Brunel University
Correct Answer File Path
/home/ambi/01MYRES/01Am/LTDU/LTDU Symposium/v1_Correct Diagrams.xmi
Student Answer File Path
/home/ambi/01MYRES/01Am/LTDU/LTDU Symposium/Test Diagram3_Two_States_Missing.xmi
Please Press button
Automatically Mark
Feedback
Positive Feedback: Good, you have selected the recipe
Negative Feedback: You are missing something. Think about the thing you should before you start cooking?
Negative Feedback: You are missing something.
Positive Feedback: Good, you have set the table
Positive Feedback: Good, you have eaten
Marks: You have got 3 marks out of 5
Raw Feedback:

Coursework Question Text

Draw an Activity Diagram for the following simplified process of students eating an evening meal:

Students can either: not eat an evening meal, eat at a friend's house or cook some food at home. If a student eats at a friend's house they simply turn up at the arranged time and eat the meal produced by their friend. If a student eats at home they must select a recipe for their meal, assemble the ingredients, cook the meal, set the table then eat the meal.

Model Solution



Sample Student Diagram