PROJECT CASE STUDY

Project Title:

Introducing a 'Forensic Flavour' the teaching databases.

Project Team

Jackie Campbell, Sanela Lazarevski, Maurice Calvert

Abstract:

Databases often receive an uninspired and uninterested response. The curriculum content of a database module generally involves the design of entity-relationship models, SQL programming, application development and advanced database applications such as data warehousing and data mining. These are often taught within the tired and relatively worn case studies of purchase order systems, retail or health care systems. However the current trend for crime scene investigation drama and the frequent stories in the news of personal tragedies involving incorrect data, missing data or data mix-up capture the attention of many. The truth is that crimes require data investigation and expert database witnesses to provide evidence and this requires database knowledge and skill.

This project involved the introduction of a 'forensic flavour' to the teaching of databases as part of an undergraduate Computing Degree to students. The 'forensic flavour' involved introducing investigative and enquiry based learning techniques as well as selecting case studies based around real-life crimes and crime data. The learning objectives remained unchanged for the modules as did the curriculum content. The initial findings are that the students engaged on average 40% better and enjoyed the experience more.

Project aims:

The objectives of the project were:

- To improve students learning by inspiring students with the importance of data, databases;
- To improve students confidence, risk taking, investigation, problem solving skills by introducing self-lead enquiry based learning;
- To improve the general perception of databases.

Initial interviews were held with database experts and the Police in order to identify situations where 1) Data had provided key evidence in a case. 2) The database had provided key evidence in a crime or 3) Data anomalies had led to serious issues. This process raised a number of interesting cases such as the Harold Shipman case and common company frauds (eg fraudulent expenses).

The material for level 5 and level 6 databases was reviewed with a view to creating 'Enquiry based Learning' activities or activities involving problem solving or investigative activities. Module case studies in the field of crime or law were identified. Much of the database material was reworked to incorporate the case studies or use 'investigative' learning activities.

The module was delivered to meet the same learning outcomes but using the newly developed material.

Outputs

A number of new learning activities were created including:

Workshop on 'Horrendous data stories' – students were asked to report on a story were the wrong or missing data had had significant consequences. Stories reported included a girl who died due to being given the wrong blood type in a transfusion, a man who has been

wrongly arrested a number of times, missing SATs results, utility companies who didn't send out bills for 6 months.

Bug fixing – students were given SQL code containing common errors to fix.

Data Analysis – students were given database tables containing data anomalies to discover.

Data Mining – students were given UK crime data to analyse and look for patterns.

Data Dictionary – students were asked to prove suggested anomalies using the database data dictionary.

Assignments were created for both levels based around real UK data.

Two L6 projects were created and have been taken on by students:

1. "Investigation Required"

This project is to create a 'case study' of a system where a fraud has occurred.

Part of the project is to investigate and report on common frauds. Some examples are:

- The rounding error fraud;
- Payments made against 'dead' project codes;
- Payments made to people after they have left the company;
- Purchases against equipment that hasn't been acquired.
- Employees making purchases, expenses claim for their own benefit;
- Identity theft;

The aim of this case study is to use it within teaching (databases, forensics) for students. They will be asked to act as an expert witness and produce a report of evidence to support the charge of fraud.

It is envisaged this project will be developed in Oracle Apex. Other possible considerations are:

- The use of multiple log-ons (to demonstrate how some personnel have access to certain information);
- The importance of the database 'metadata', only available to the DBA to show when data was updated, when users logged on and what they accessed.

2. DBA Consultancy

This project is to create a 'case study' of a system which is performing badly. The aim of this case study is to use it within database teaching for students. They will be asked to act as a Database Administrator (DBA) and produce a report of recommendations to improve the system using appropriate techniques. Examples to include in the system are:

• Concurrency problems (users are locked out, dead lock)

• Slow running queries, which can be evidenced and investigated using QEPs. (such as a cart-prod followed by a SELECT distinct);

- Examples of redundant data, inconsistent data as a result of de-normalisation;
- Examples of slow running queries where de-normalisation would help;
- Examples of out of date indexing plans;
- Examples of insecure systems (security put in client, SQL injection);
- Examples of default setting that lack consideration (logging, audits, indexes).

It is envisaged this project will be developed in Oracle Apex.

Both these projects received much initial interest and have been taken up by students as final year projects. They are to be completed in June 2010 and the intention is to use them in L6 teaching.

The team have two papers under construction on the project. Abstracts have been accepted to the ATINER and TLAD conferences.

Impact

The staff involved in the have enjoyed learning about the Forensic industry. Students appear to have been genuinely motivated and enthused by many of the activities.

Additionally it appears that the project has widened the horizons of databases, some of the developed content has been used with Sports Marketing and Business students – both groups who seemed to appreciate the value and importance of data as a result. At Enterprise events delegates have volunteered their own 'data-mix-up' stories, again suggesting that you don't have to be an expert to appreciate the significant of quality data!

Evaluation

Students on both L5 and L6 modules were surveyed before the modules (L5 database Application Development and L6 Advanced Database Management). They were asked which skills they thought they would be developing and given the options to strongly agree, agree, disagree, strongly disagree or neither agree nor disagree. The skills surveyed were grouped into: Technical skills (SQL, database design), investigative skills (data integrity, data analysis) and Enterprise skills (problem solving, personal time management, researching and scoping a problem, communicating ideas, confidence, risk taking).

Students were surveyed again on completion of the modules and asked to evaluate whether they thought they had developed the skills, again given the same options.

The findings are still being evaluated The initial findings are that the students engaged on average 40% better and enjoyed the experience more.

What Next?

The team intends to continue to consider enquiry based learning and problem based solving. Some more learning material along similar lines is planned for 2010/11.

Some useful contacts have been made in the field of forensics which we will continue to develop.

General Conclusions

Even though, it is difficult to measure learning (as discussed for example by Biggs, 1993), we can report that due to one or more likely, a combination of the factors implemented in the 2009/10 delivery the student experience was improved. Learning is a significant part of the student experience and the improved module feedback and student reflection, although not conclusive, does suggest an improved learning experience.

To conclude, the forensic flavour was appreciated by the students and it was felt by staff and students that there was a general interest in the application of crime based case studies to databases.



Investigation into 'Data Stories'

Jackie Campbell, Leeds Metropolitan University



Introduction

Databases often receive an uninspired and uninterested response. The curriculum content of a database course generally involves entity-relationship modelling. SQL programming, application development and advanced database applications such as data warehousing and data mining. These are often taught within the tired and relatively worn case studies of purchase order systems. retail or health care systems. However the current trend for crime scene investigation drama and the frequent stories in the news of personal tragedies involving incorrect data, missing data or data mix-up capture the attention of many. The truth is that crimes require data investigation and expert database witnesses to provide evidence.

The Database team have introduced a 'Forensic Flavour' to the teaching supported by Enterprise funding. The 'forensic flavour' involves introducing investigative and enguiry based learning techniques as well as selecting case studies based around real-life crimes and crime data. This poster describes one of the activities introduced

Objectives:

·Give students an appreciation of the importance of data and database systems.

·Encourage students to have an interest in their subject background information and something to talk about at interviews. ·Provide the opportunity for students to take responsibility for their own learning by finding their own case study

Activity: Investigate and Report on a Horrendous Data Story

A case study was presented to students via an interactive workshop. Students were told that an 18 year old girl had died during a routine operation and invited to ask guestions.

- The students soon gained the information that the girl had died due to receiving the wrong blood type in her operation.
- Further activities were carried out to analyse the data for the system and how this tragedy could have been prevented using database technology.
- By the end of the session the students were given a hand-out for the case study with the following headings:
 - 1. Data Story description (with references).
 - 2. Data Analysis the event occurred due to missing/incorrect/duplicate data?
 - 3. Data stores in the system.
 - 4. Requirements of the system.
 - 5. Data Model for the system (or sub-system).
 - 6. Data constraints used to prevent data errors.



Practical

Students spent practical time on this assignment activity:

"Investigate and report on your own data quality story. Using the template provided." The students were soon sharing good sites they had found and were generally excited by what they were finding. As a group we discussed and few and moved to the database issues and how the database constraints may have prevented the problems.

Assignment task

The assignment task was a revision exercise so only weighted at 10% of the overall assessment. 40% of the students sent me their examples (unprompted) for formative feedback and the work was a very high guality. Here is an example:

Data Story: Dead girl given truancy warning

"The parents of a girl who died suddenly have received a school letter demanding she improve her attendance"

Megan Gillan, 15, was found dead in the bedroom of her home in Macclesfield. Cheshire two months ago. Her parents say they were "floored" by a Macclesfield High School letter, which threatened to ban Megan from the end of year prom. The school has apologised for the mistake, which they said was down to an error on the computer database. Database error

"Capita software SIMS (School Information Management Systems) maintained Megan's details and clarifies that Megan's name had been taken off the school roll when she died, and removed from the main school database. However, unknown to the school, her details had remained in a different part of the computer system and were called up when the school did a mail merge letter to the parents of all Year 11 students about their prom. The letter called up details of each student's attendance for the whole year to date and because Megan had been on roll in September.

The letter, dated 16th March, said "students must have at least 92% attendance and Megan's is currently 60.4%".

This occurred due to:

This occurred due to duplicate data remained in different parts of the database system, even though Megan's details had been taken off from the main school database. This was caused due to no referential integrity in the main school system with all the other systems.

User Requirements

Insert/select a student who is unable to attend school View student attendance details

List the number of students absent for [X] number of days

How could this story been prevented?

- This could be prevented by adding an integrity check query, which will compare all the students' details in all the available systems and check if there are any data errors remaining in any other systems. Should there be any bad data, it will inform the user and ask them what they want to do with it.
- Other option to prevent Megan's story would be to set up a server link between the two servers to retrieve all the data and populate it at once. This means we would need to set up a referential integrity check using foreign key constraints between both.
- "Referential integrity is a feature provided by relational Database management systems that prevents users or applications from entering inconsistent data" [Michael Blaha, 2005]
- Additionally, they could add an active flag on the students who are not attending the school currently. This means that when the SIM system sends mail merge letter to all the students who had not been attending school, the student who had been removed from the school database
- will be flagged as 'Inactive', which means that their parent will not receive a letter. However having this flag will also mean that statistics can still be built on a schools performance etc but it will prevent inappropriate letters from being sent out.

References Dead girl given truancy warning:

Available from http://news.bbc.co.uk/1/hi/england/manchester/7963081.stm [Accessed 01/10/09]

- Michael Blaha and William Premerlani
- Object-Oriented Modelling and Design for Database Applications. [1998]
- Referential Integrity:
- http://docs.google.com/gview?a=v&g=cache:DvuQACEYBJIJ:www.odbms.org/download/007.02%2520Blaha%2520Referential%2520Integrity%25 20Is%2520Important%2520For%2520Databases%2520November%25202005.PDF+referential+integritv&hl=en&sig=AFQiCNFxVDiTkLGhCCR
- uwo2mGrD36 UBIA



Other examples

Here are some of the other examples presented:

- NASA lost a \$125 million Mars orbiter because one engineering team used metric units while another used English units for a key spacecraft operation.
- · Schools are finding children's "SATS" scripts are being returned unmarked - with the pupils wrongly recorded as having been absent for the tests.
- A British man is being convicted of 8 murders which were carried out in the U.S.A. He claims it is a case of mistaken identity. The British man, Mark Ross, shares the same name, birthday and middle name as the Canadian man wanted for the murders. If the British man cannot prove his innocence then he will be convicted to the death sentence in Florida. The British man, Mark Ross, does not have any living parents and has spent his life travelling across the world. Stephen Jakobi told the BBC website "One is always concerned about American identity cases. We have had some nasty cases of mistaken identity in the past."
- · Many residents of New Orleans surprisingly did not have flood coverage and their homeowners policies covered only wind and not flood losses (Cornrjo, 2006). A number of these policyholders had been told by their agents that they did not need to buy flood coverage, as they were not in a flood zone. Apparently this faulty information was based on maps that had not been updated in decades.

Conclusions and other work

Based on the standard of work, initial feedback and student engagement, the 'Data Story' activity and the introduction of a forensics flavour in database teaching has been successful. We are undertaking more rigorous research on the project and have a couple of papers in progress with a view to publish in June 2010.

References

References :

Archabal, R., Yolini Heart Transplant Des Alter Two Wesler, NY Times, Feb 23, 2003 Arenco, K., Yeshniaal Problems cal Data Poldems in 6AT scores, New York Times, March 8, 2006 CNN, Yand Prices, US says Faality Back to Chinese Embanys Borning', Crinc. Online, VJ, 1098b Chan, WASA's Meire Confusion Causes Mars Obieter Lass', Segtember 20, 1997b Consumer Federation of America, Wilson et Américan Segtember 20, 1997b Scores", 2002 Compio, R., "Searching for a Cause", Bests Review, February, 2006 Dasu, T and Johnson, T., Exploratory Data Mining and Data Cleaning, Wiley 2003 Eckerson, W., "Data Warehousing Special Report: Data Quality and the Bottom Line" Externs, M., 'Data Waterboung Special Report Data Calify and the Bittern Line', Water Vol. Time, Journal V. 1995. Multiplier into Schubine' Report Vol. Time, Journal V. 1995. Multiplier into Schubine' Report Vol. Time, Journal V. 1995. Multiplier into Schubine' Report Vol. Table Calify, 'Data Calify Press, 2001 Report Vol. Table Calify, 'Data Calify Press, 2001 Report Vol. Table Calify, 'Data Calify Press, 2003 Schuber, Data Calify, 'Data Calify, 'Data

[Accessed 12/10/09]