# Knowing when to target students with timely academic learning support: Not a minefield with data mining

**Elizabeth McCarthy**
University of Southern
Queensland

The strategic scheduling of timely engagement opportunities with academic learning support, targeting specific student cohorts requires intentional, informed and coordinated planning. Currently these timing decisions appear to be made with a limited student focus, which considers individual course units only as opposed to having an awareness of the schedule constraints imposed by the students' full course workload. Hence, in order to respect the full student academic workload, and maximise the quantity and quality of opportunities for students to engage with learning advisors, a means to capture and work with the composition and distribution of student full workload is needed. A data mining approach is proposed in this concise paper, where public domain information accessed from the back end html language of course unit information webpages is collected and consolidated in graphical form. The resulting visualisation of the students' academic learning activities provides a quick and convenient means for academics to make informed scheduling decisions. The case study presented describes the implementation of the data mining in the context of discipline specific academic learning advisors at the University of Southern Queensland servicing three campuses under the 'One-University' model.

## Introduction

Despite the intention to plan and schedule learning activities for the student, the logistical arrangements do not tend to consider a holistic view of the student's total commitments under the full study load. Instead academic schedules are generally designed based on the micro academic resource level. Gill (2015) flagged this as an issue where a review of academic practices indicated that units/courses are managed in an independent, and largely isolated, modularised manner. The consequence of this lack of communication about a student's total academic commitments is the high susceptibility for clashes to occur or the incidence of lengthy, highly concentrated blocks of learning activities, both resulting in less than optimal student engagement and performance. This closed nature of scheduling means that students are forced to prioritise the application of their attention, which has varying levels of success depending upon the time management skills of the student (Gill, 2015; Kyndt, Berghmans, Dochy, & Bulckens, 2014). By investigating methods to consolidate and communicate the distribution of academic workload of students, a greater awareness of student behaviour may be achieved. Currently, there are no tools available to do so, hence the arduous task is completed manually.

In this concise paper, we explore the implementation of educational data mining to curate and distil scheduling data, and present the collective students' academic commitments with data visualisation, ready for use in human decision making. Despite having a number of potential uses in the higher education sector, in particular, this concise paper will consider the scheduling of on-campus classes during the week, and the scheduling of assessment throughout the semester, and the value of this information for the university's academic learning advisors.

Typically, the scheduling of learning activities has a limited consideration of total students' academic commitments due to the following restrictions:

- The strategic consolidation of information will require considering a variety of combinations of courses/units to cater for the diverse composition of students' academic commitments within the targeted cohort.
- The opportunities afforded by the awareness of students' total workload and scheduling of activities will only eventuate if academics are provided convenient access to information.

- The collection of the relevant open access data is problematic due to its sparse storage on a plethora of individual course/unit related webpages.

Overcoming these restrictions would enable more convenient access to information which would be used to inform the management of student academic workload.

Exploration of novel methods of engaging with the total student academic commitments' is considered essential in incorporating a holistic student centred approach into strategic pedagogical practice. A technological strategy which adopts data mining and data visualisation practices would permit academics to access reliable public domain information and create knowledge in a consolidated and easily interpreted form (Mohamad & Tasir, 2013; Cristobal Romero & Ventura, 2007). While data intelligent methods are used in literature to optimise the class timetables scheduling (Deris, Omatu, Ohta, & Samat, 1997; Lai, Wu, Hsueh, Huang, & Hwang, 2008; MirHassani & Habibi, 2013; Pillay, 2014; Rudová, Müller, & Murray, 2011), they generally consider a single course/unit as opposed to the holistic student course workload composition, or are teacher or resource centric (Thompson, 2005). Hence they do not capture the true nature of the distribution of student workload throughout the semester.

However, as noted by Kitto, Lupton, Davis, and Waters (2016), there is an interdisciplinary gap between educators and computer scientists, which compounds the problems mentioned before. Academics interested in scheduling events which consider the students' total academic commitments do not possess the technical computing skills required for developing their own novel technologies.

Another significant issue is the widespread lack of awareness and appreciation of data science and visualisation which would be invaluable for engaging in evidence based practice (Peña-Ayala, 2014; Cristóbal Romero & Ventura, 2010). The lack of engagement with evidence based practice means that decisions are likely to be made based on anecdotal and redundant historical information, rather than catering for the current cohorts of students. Academic services offered from the University of Southern Queensland (USQ) library, such as timely offering of workshops and other scheduled learning activities would benefit from access and usage of complete student academic workload information. This prompts us to introduce the case study that will be referenced to for the remainder of the paper.

## Case study: logistics of academic learning advisors

The University of Southern Queensland's (USQ) Academic Learning Advisors are disciplined based specialists who service the university's three campuses at Toowoomba, Ipswich and Springfield, in addition to the significant online cohort which represents up to 70% of USQ's student body (*The University of Southern Queensland Annual Report*, 2016). As members of USQ's Library academic staff, the learning advisors provide students and staff with academic learning advice in mathematics and discipline aligned academic language. With time being a precious commodity, the learning advisors require complex time management strategies to maximise productivity and maximise the engagement opportunities with their clients across the three campuses throughout the semester.

To achieve higher productivity, learning advisors would prefer to travel to service campuses at times and on days when their designated cohorts will have a strong presence to improve the likelihood of student engagement. Learning advisors would also need to be aware of major assessment deadlines for each cohort, to allow necessary planning for surges in demand. Hence the provision of student academic scheduling information is useful at weekly, daily and hourly temporal scales.

# Method

The use of a particular data mining technique called web scraping or web mining was implemented using the programming language Python 2.7, aided specifically by functions from the BeautifulSoup module, which specialises in web scraping capabilities. Web scraping involves inspecting html code for specific fields such as those shown in Figure 1, and retrieving the corresponding arguments.

```
<event id="7400" timesort="09001100" colour="D4FFD4" date="26/02/2017 ... 4/06/2017">
<day>0</day>
<prettytimes>9:00 AM-11:00 AM Workshop</prettytimes>
<starttime>9:00 AM</starttime>
<endtime>11:00 AM</endtime>
<category>Workshop</category>
<prettyweeks>1-6, 9-15 (Sem 1 Teaching 2017)</prettyweeks>
<rawweeks>YYYYYNNYYYYYYNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN</rawweeks>
<resources>
<module title="Course">
<item>CMS1000</item></module>
<staff title="Staff">
<item>Warnest, Bev</item></staff>
<room title="Room">
<item>A5LR2 (Springfield)</item></room></resources></event>
```

*Figure 1: Sample of html code behind the timetable for Springfield Semester 1, 2017 offering for CMS1000* ([https://www.usq.edu.au/timetables/Sem12017Sfield/m7399.xml](https://www.usq.edu.au/timetables/Sem12017Sfield/m7399.xml))

In alignment with the various USQ learning advisor designated disciplines, the courses/units have been segregated into their corresponding themes: mathematics (MATH), research and higher degrees (HDR), education (EDU), engineering, built environment and information technology (ENGBENVIT), management, commerce and law (MGTCOMLAW), creative arts, society and culture

(ARTCOMMN, also includes enabling programs such as tertiary preparation programs), sciences and health (HEAWBLSCI). These have been identified and colour coded in Figures 2 and 3.

## Program and course classification

Simple text files (*.txt) lists are generated from iterating through the course specifications pages at [https://www.usq.edu.au/course/specification/.](https://www.usq.edu.au/course/specification/) Courses may also be classified according to faculty/section, school or department, Australian Standard Classification of Education (ASCED) code or year level. More customised automated clustering also possible with finer definitions of rules.

### Class timing

USQ's class timetable web pages are openly accessible from launch page at [https://www.usq.edu.au/current-students/organise-enrolment/timetables/class](https://www.usq.edu.au/current-students/organise-enrolment/timetables/class) , which provides individual class timetables for Toowoomba and a combined Springfield/Ipswich into the three semesters over the academic year. Incremental counters are used to each acknowledge each instance in the hourly clusters. Clusters may be used to filter by discipline, class type, semester number, year level, campus.

### Assessment timing

Workload information may be interrogated on a course by course basis from the freely available course specifications web pages, separated into year and then course offering at [https://www.usq.edu.au/course/specification/](https://www.usq.edu.au/course/specification/). Specific assessment information may be extracted using filters, including assignment due date, type (for example quiz, essay, report, assignment or presentation), total marks and weighting. Fully scaleable, the data can be used to present assessment spread for nominated courses, up to a comprehensive distribution of all courses' assessment deadlines across the semester.

## Results and discussion

The web scraping and compilation of information from the 805 course specification webpages detailing the courses/units offered by USQ was achieved in the order of minutes, which may vary depending upon the processing power of the computer used. The course specification page data mining activity yielded lists of all courses offered at USQ have been separated according to their corresponding themes, and are used in the process of grouping the courses for the graphical representations of class and assessment timings.

### Class timing

From the daily bar charts (Figure 2) depicting the distribution of classes throughout each day, the discipline based learning advisors are able to decide best days to service different campuses based on the courses scheduled. Learning advisors may strategically choose days where their cohort are highly represented on campus, and schedule academic learning advisor engagement events during the gaps noticed in the distribution of the class times. While the data extracted from the website alone does not indicate class sizes, further data sets incorporating student study mode status are possible to further inform users.

### Assessment timing

The quick inspection of the semester wide, and university wide, assessment deadlines (Figure 3) presents the academic learning advisor with information that is useful as a guide for predicting peaks in demand. At a micro level, this may be customised for specific cohorts, and used to help students become aware of the academic deadlines throughout the semester, and provide feedback to academics regarding competing academic demands.

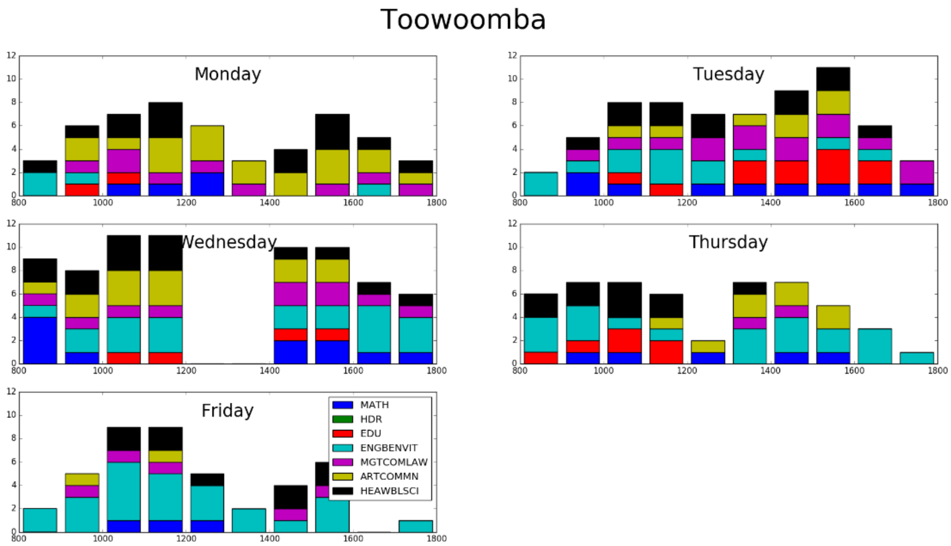


*Figure 2: Distribution of number of first year on campus classes (lectures only) scheduled each day of the week at Toowoomba during Semester 1 2017, classified by course themes*

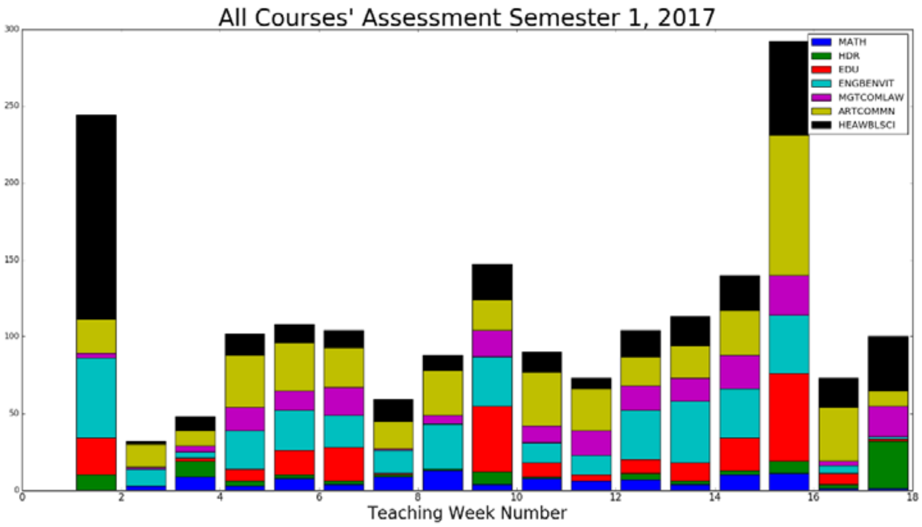


*Figure 3: Distribution of all courses' assessments (excluding end of semester examinations) across Semester 1, 2017, classified by course themes*

## Conclusion and future scope

The adoption of web scraping, as a form of data mining, has enabled the academic learning advisor, and potentially other members of the university community, to have access to information regarding student's academic commitments, in bulk. In addition to providing this information to learning advisors, who may assess the best application of their time (both day, and periods throughout the semester) at different campuses, this also has the potential to be inform other student and staff centred operations. To date, this consolidation of information has been requested by members of learning advising, student services and student experience teams at USQ.

Further applications of this work could involve using the filters in these algorithms to present more comprehensively the nature of students' academic workload, and may provide insight into the theoretical workloads of the academics servicing these courses. With the integration of other data science methods, such as using machine learning clustering techniques, means there is potential to smooth the distribution of assessment throughout the semester, based on assessment weighting (proportional to expected student effort in hours) and due dates, to reduce the incidence of high concentrations of deadlines. Capturing the true distribution of workloads provides the opportunity to use this measurable evidence in the negotiation of course workloads and more effectively understand student and academic workload stress.

## References

Deris, S., Omatu, S., Ohta, H., & Samat, P. (1997). University timetabling by constraint-based reasoning: A case study. *Journal of the Operational Research Society*, 48(12), 1178-1190.

Gill, B. (2015). Talking about the elephant in the room: Improving fundamental assessment practices. *Student Success*, 6(2), 53-63.

Kitto, K., Lupton, M., Davis, K., & Waters, Z. (2016). Incorporating student-facing learning analytics into pedagogical practice.

Kyndt, E., Berghmans, I., Dochy, F., & Bulckens, L. (2014). 'Time is not enough.' Workload in higher education: A student perspective. *Higher Education Research & Development*, 33(4), 684-698.

Lai, L. F., Wu, C.-C., Hsueh, N.-L., Huang, L.-T., & Hwang, S.-F. (2008). An artificial intelligence approach to course timetabling. *International Journal on Artificial Intelligence Tools*, 17(01), 223-240.

MirHassani, S., & Habibi, F. (2013). Solution approaches to the course timetabling problem. *Artificial Intelligence Review*, 1-17.

Mohamad, S. K., & Tasir, Z. (2013). Educational data mining: A review. *Procedia-Social and Behavioral Sciences*, 97, 320-324.

Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications*, 41(4), 1432-1462.

Pillay, N. (2014). A survey of school timetabling research. *Annals of Operations Research*, 218(1), 261-293.

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert systems with applications*, 33(1), 135-146.

Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics*, Part C (Applications and Reviews), 40(6), 601-618.

Rudová, H., Müller, T., & Murray, K. (2011). Complex university course timetabling. *Journal of Scheduling*, 14(2), 187-207.

Thompson, G. M. (2005). Using information on unconstrained student demand to improve university course schedules. *Journal of Operations Management*, 23(2), 197-208.

The University of Southern Queensland Annual Report. (2016). Retrieved from https://www.usq.edu.au/about-usq/about-us/plans-reports/annual-report

**Contact author:** Elizabeth McCarthy, Elizabeth.Mc.Carthy@usq.edu.au
**Please cite as:** McCarthy, E. (2017). Knowing when to target students with timely academic learning support: Not a minefield with data mining. In H. Partridge, K. Davis, & J. Thomas. (Eds.), *Me, Us, IT! Proceedings ASCILITE2017: 34th International Conference on Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education* (pp. 128-132).

Note: All published papers are refereed, having undergone a double-blind peer-review process.