

Data Mining to Investigate University Expectations of Work

Guoxin Tang, University of Louisville, Louisville, Kentucky

ABSTRACT

Objective: To investigate the expectations the faculty has of students, and how varied those expectations are.

Methods: Course syllabi from representative departments were collected to investigate variability in course requirements between faculty within a department, and across departments (107 total). In addition, information will be collected to examine department expectations of faculty work. For this study, we used SAS Text Miner to investigate the syllabi. In addition, the syllabi were coded for the number of tests, quizzes, and homework assignments, the presence of participation requirements or attendance requirements, and weighting of the final to compare manual coding to SAS Text Miner results. The syllabi were coded as to whether they included a statement indicating the expected number of hours each student should spend studying course material.

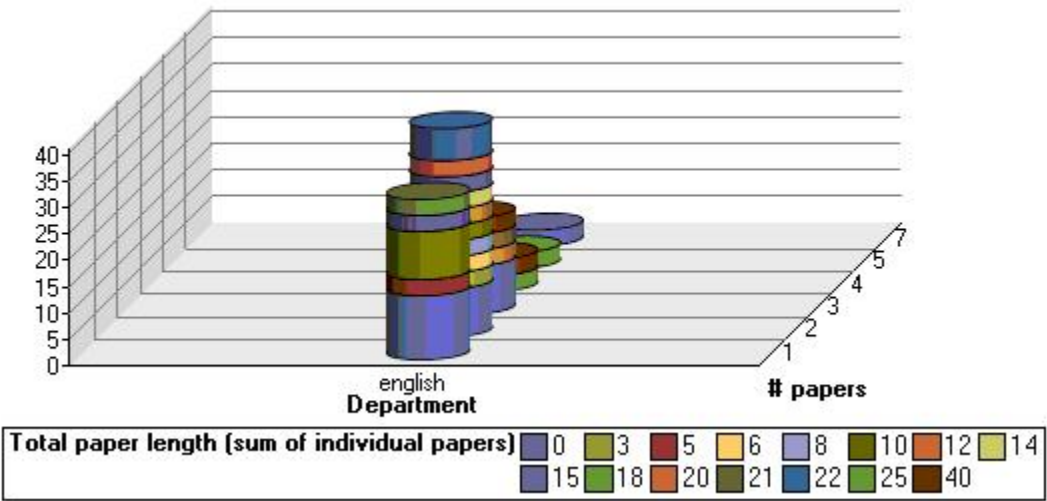
Result: Text Miner divided the 107 syllabi into 10 clusters. The examination reveals that most syllabi, including those from the mathematics, physics and English departments, focused on requirements and grading. Approximately 58% of syllabi were concerned about requirements; the interest in course requirements was maximal for the mathematics department.

INTRODUCTION

Course syllabi contain information concerning expectations of student work. The primary purpose of a syllabus is to communicate to one's students what the course is about, why the course is taught, and what will be required of the students for them to complete the course with a passing grade. It is the purpose of this paper to demonstrate the variability in expectations of different departments.

There are many different expectations of the faculty work in different departments, even in the same department. Consider Figure 1, for example, which shows the proportion of syllabi that include paper assignments. Such assignments are clearly very important for the English department, compared to Mathematics and Physics Departments. There are no requirements for papers in those two departments. However, there is considerable variability within the English Department.

Figure1. Total paper length required by the English Department. There is considerable variability in the required total length required of the semester



In this study, data analysis and text mining are used to examine the relationship between expectations of faculty and student work for the purpose of teaching skill improvement. It is also of interest to compare the results based on two different analysis methods: text mining versus manual coding

METHODS

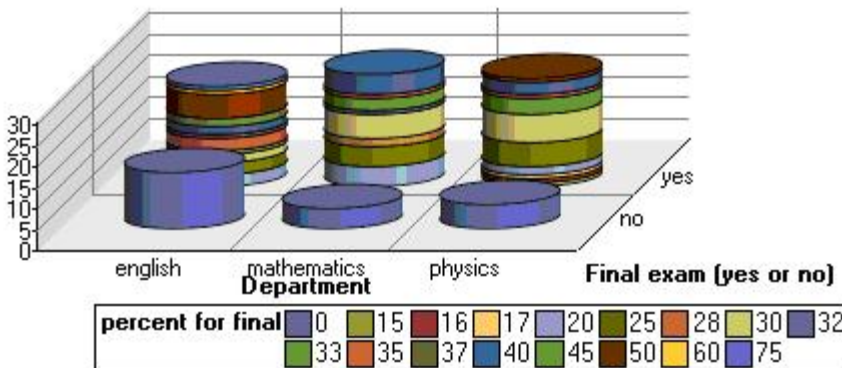
A total of 107 copies of syllabi from English, Mathematics and Physics departments were collected to examine department expectations of faculty work. The syllabi were scanned individually into a computer directory, and then the macro, %tmfilter, was used to create a SAS dataset. Information in the syllabi was also manually coded into a SAS dataset. The variables include the number of papers, total paper length (sum of individual papers), number of tests, final exam, quiz, midterm exam, percent for discussion (or participation), percent for attendance, percent for regular homework, percent for final exam, percent for quiz, and percent for the midterm exam. We used SAS Enterprise Guide to examine the data. It made the information of syllabi more visible. The other method used was SAS text mining to examine all materials in text version. Here, Text Miner provides an alternative method to analyze data and collect useful information.

RESULT

Results from Manual Coding

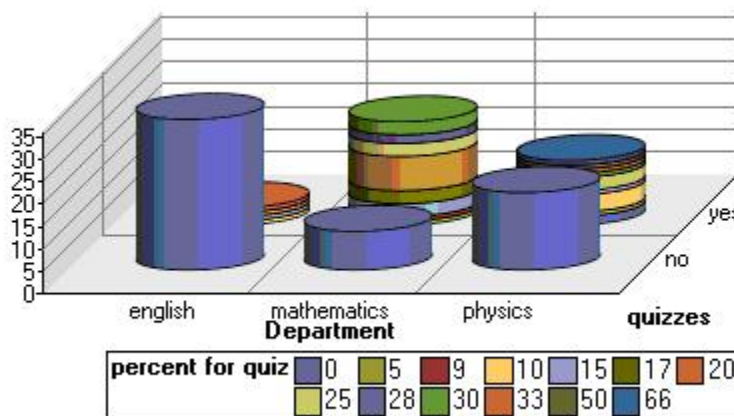
Firstly, manual coding was used to examine expectations of work. Figures 2 to 6 show the information derived from syllabi according to the final exam, quiz and midterm exam. Note that the English Department requires papers as a big part of the final grade. Tests are used for reference.

Figure2. The importance of final exam in the three departments



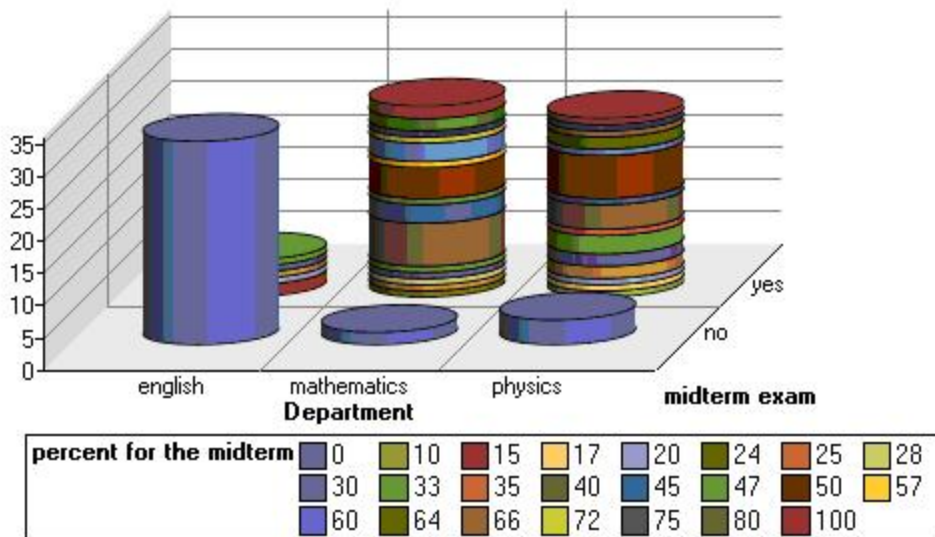
Note that the final exam is required for most courses in the departments of mathematics and physics. The percentage of the final grade for the final exam, however, varies from 20% to 40%.

Figure3. The importance of quizzes for the three departments



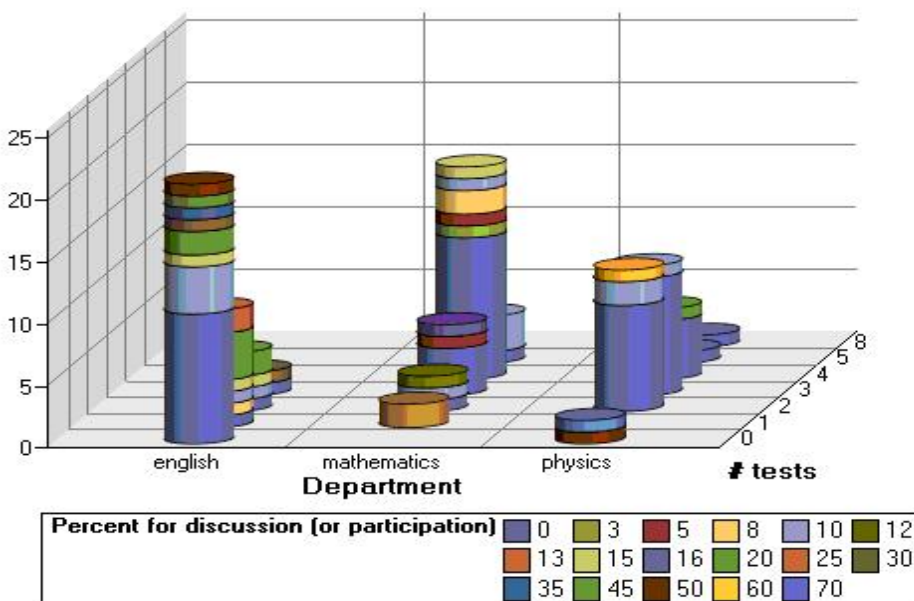
It shows that quizzes play an important role in the final grade of Mathematics and Physics compared with the English department. More than 75% of the courses in the Mathematics Department and about 50% of the courses in the Physics Department require quizzes, compared with 15% of the courses in the English Department.

Figure4. The importance of the midterm exam for the different departments



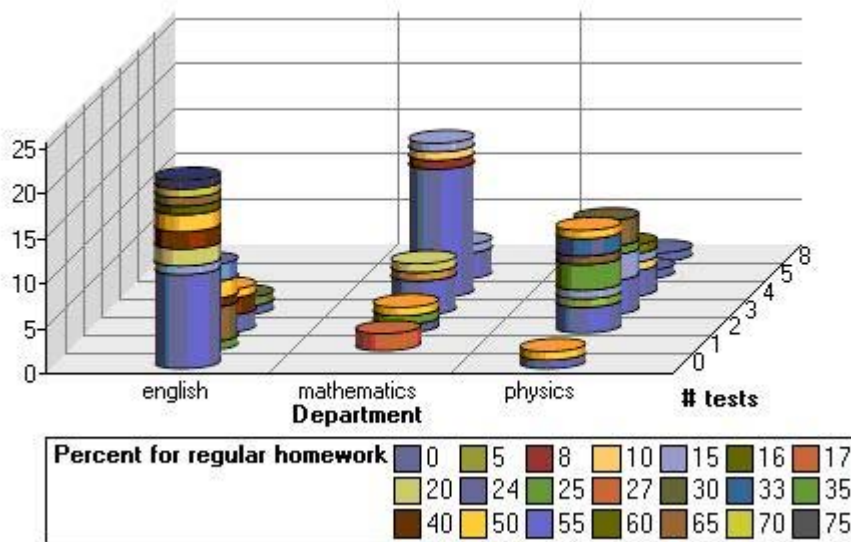
Note that about 95% and 90% courses in the Mathematics and Physics Departments respectively require a midterm exam, and so do about 15% of the courses in English.

Figure5. The percentage of discussion according to the number of tests for the different departments



It is clear that for most courses of the Mathematics and Physics departments, the percentage of the grade from discussion decreases as the number of tests increases.

Figure6. The percentage of homework according to the number of tests and different departments



Note that the percentage of homework decreases with increasing of number of tests. For mathematics, when the number of tests is more than 2, the weight of homework in the final grade becomes very small.

Text Miner Results

Another step was to investigate all the syllabi by using text mining. All syllabi were scanned into a Windows directory as text. Text mining can cluster syllabi into similar groups to identify the weights of expectation. The %tmfilter macro can read documents of different formats that are stored on the file system, and create a SAS data set that can be used as input for the text miner node. The following code shows how to generate a SAS data set::

```
%tmfilter(dataset=work.txtinput, dir=c:\testdir, numchars=32000);
```

In this analysis, Text Miner divided the 107 syllabi into 10 clusters. Table 1 gives the clusters. From the description, cluster 1 with a frequency of 15 shows the discussion and writing assignments in the English Department. Cluster 5, with frequency 24, shows exams and requirements in Mathematics.

Table1. Table of Clusters of Syllabi

# ▲	Descriptive Terms	Freq	Percentage	RMS Std.
1	+ teach, + practice, + presentation, + discussion, + late, + project, + include, + write, + discuss, + day	15	0.135135135...	0.1462852...
5	+ calculator, math, + exam, + disability, + section, + requirement, + test, + cover, + require, + problem	24	0.216216216...	0.0837102...
6	+ homework, + lecture, + date, + work, + email, + test, + cover, + grade, + exam, + make	10	0.090090090...	0.1043340...
7	+ test, + material, + make, can, + expect, + cover, + grade, as, + do, + week	14	0.126126126...	0.1311981...
9	+ grade, physics, + score, + exam, + physic, + homework, + encourage, office, + follow, no	9	0.081081081...	0.0963927...

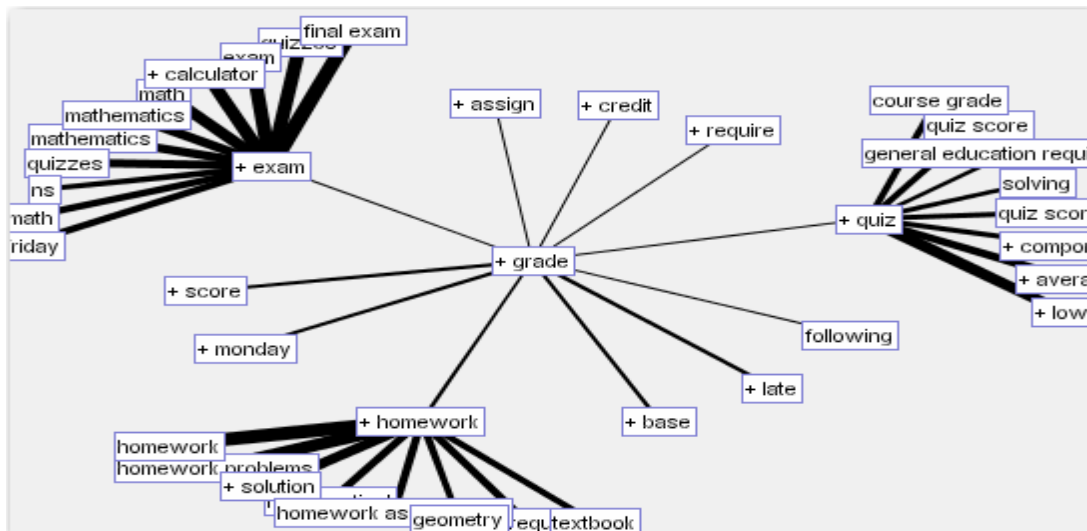
Cluster labels for Table 1 are suggested below:

1. Discussion & writing

2. Math exam
3. Homework
4. Tests
5. Physics tests

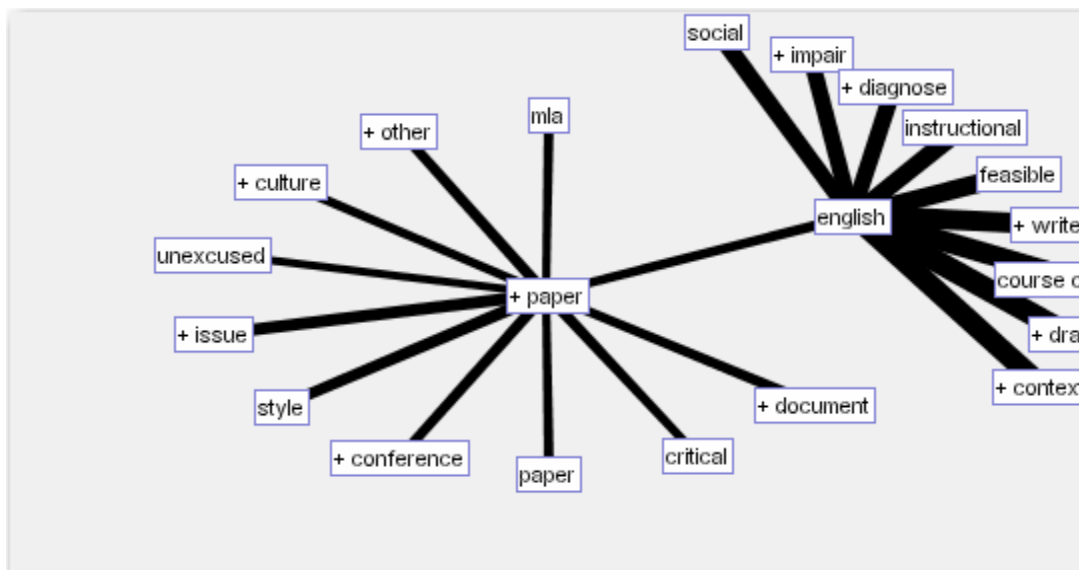
Text Miner can also give concept links, showing how different terms are related in the documents. Figures 7 to 9 show the results of analysis by using text mining.

Figure7. Concept links for the term “grade”



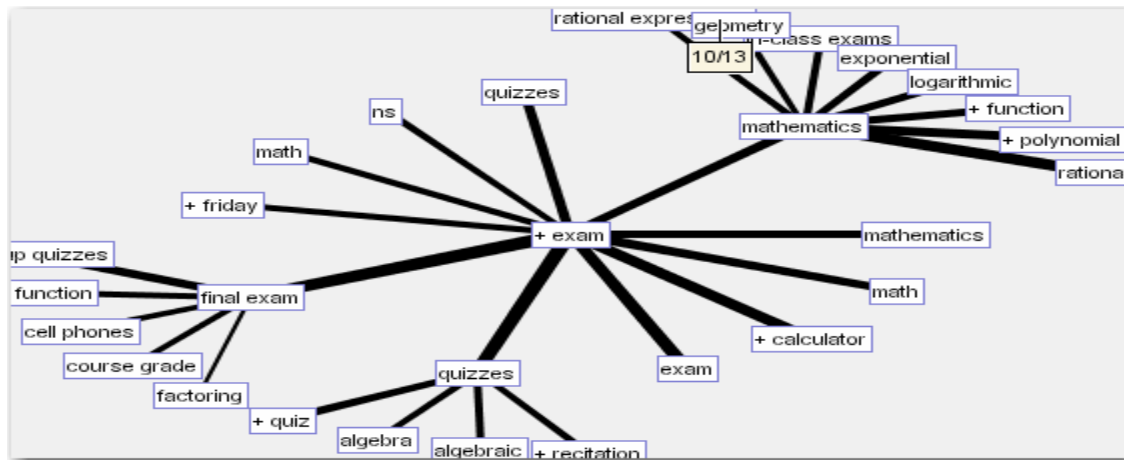
Here, 'grade' is related to 'exam', 'quiz', and homework, and so on. It is clear that the exams, quizzes and homework are the most important parts for the final grade. Note that the emphasis here is on mathematics.

Figure8. Concept links for the term “paper”



The most meaningful term related to “paper” is ‘english’. This shows clearly that the English Department requires papers as the way to grade.

Figure9. Concept links for the term “exam”



Note that the linked terms are ‘final exam’, ‘quizzes’ and ‘mathematics’, etc. Exam is the primary way to evaluate a student in mathematics department.

CONCLUSION

The analysis tells that most syllabi, including those from the mathematics, physics and English departments, focused on grading and requirements about papers and exams. The writing plays an important role in English and exam does in mathematics and physics departments. The results of this research reveal the relationship between the expectations of faculties and student work, and help the faculty to develop the teaching skill to improve the teaching purpose.

Manual coding and text mining present the similar results. Therefore, the use of SAS Text Miner and text analysis can be substituted for the time-consuming process of manual coding and analysis.

ACKNOWLEDGMENTS

1. My Graduate Advisor; Dr. Patricia Cerrito
2. SESUG 2007 for giving me the opportunity to present my paper

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Guoxin Tang
 Ph.D. Student in Applied Mathematics
 328 Natural Sciences Building
 Department of Mathematics
 University of Louisville
 Louisville, KY 40292
 Work phone: 502-852-7012
 E-mail: g0tang01@louisville.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.