

Ways to improve Course Selection and Recommendation at North Carolina State University

Fuxing Luan
Master of Computer Science
North Carolina State University
Raleigh, North Carolina

Fuzail Misarwala
Master of Computer Science
North Carolina State University
Raleigh, North Carolina

Zubin Samuel Thampi
Master of Computer Science
North Carolina State University
Raleigh, North Carolina

Abstract—Course selection is one of the most important aspects for a graduate student given the short amount of time they have, to learn as much as they can, while maintaining a good academic standing. The process of course selection is not as easy as it seems. It is a long and arduous decision making process that relies on information available to students based on their needs. In this paper, we have provided new methods for students to gain access to more information, and made it easier for them to decide what courses they should enroll for. We have also conducted some user evaluations and surveys to find out which one of our solutions works better and more effectively based on user feedback.

Keywords—Course Selection, Course Recommendation, First Semester, Computer Science, Masters, North Carolina State University.

I. INTRODUCTION

The tools for new students about to begin a program at the university to make decisions on enrollment are few in number. They have to rely on the small amount of information they get through their peers (if any) and the incomplete and uncomprehensive information they receive through course descriptions or the university website. Through our initial surveys we found that this information is not nearly enough, as the criteria that students actually consider are lot more including feedback from previously enrolled students, grade distribution, relevant information about course work such as inclusion of projects, assignments, distribution of exams, and overall workload. Most of these criteria mentioned above, either do not exist or students are unaware about the ways to gain access to them.

We aim to provide the students with all of the information they need when they enroll for courses. Based on our past surveys, we concluded that the three most efficient ways to counter the existing problem are:

1. Provide feedback about courses from senior students who have already enrolled in them. This helps people get answers to questions which cannot be answered through a simple course description. It deals with providing more detailed, and subjective information such as details about workload, assignments, professor reviews, or any sort of information that is unavailable anywhere else. This deals with the aspects of a course which one can know more about only through experience. We created a common forum on the popular networking website Piazza. It is a platform where anyone can ask a question by creating a thread with responses from other

students or even create polls. The reason for choosing Piazza as a platform were that it is very easy to use for a large scale user base like this. Another big advantage is the search feature where anyone can lookup keywords, and even modify their search to look through only a specific course as well. Piazza also retains all its past posts and all their comments so it is easy for someone to look up a question which may have already been asked before instead of creating a new post, thus avoiding repetition.

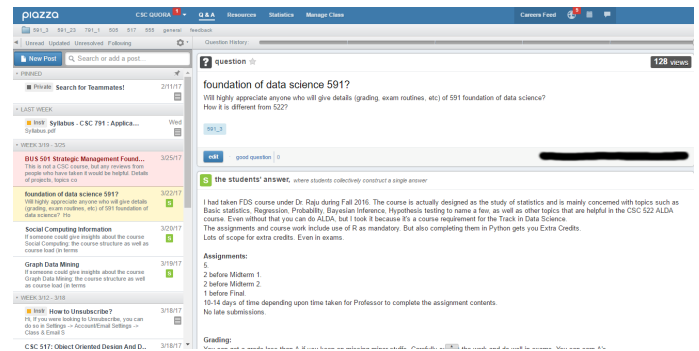


Fig. 1. A screenshot of the Piazza forum

2. Gather information about courses in as much detail as possible and provide it in a convenient location. A lot of the new students do not know where to get more information about course descriptions, syllabus, or grade distribution information. We built a website using Ruby on Rails which serves as a convenient spot for consolidated information. The data we provide includes aggregate grade distribution information, specific grade distributions, detailed course descriptions, and syllabus information. The website is easy to navigate. It includes individual pages for every course which provides an overview about the course, a syllabus, and grade distribution information about that specific course. The aggregate grade distribution information is presented on a separate webpage in the form of a table. The table can also be sorted in the order of Number of students enrolled suggesting the popularity of the course, or %As that is the percentage of students that scored an A out of the total number of students that enrolled.

3. Our third solution was targeted at people that know what criteria they want to base their decisions on, but still cannot understand which course is a close match to what they need. This method makes use of text mining techniques to make direct suggestions to the users on what courses are a good fit for them. Based on the data we collected through our earlier

Course Helper				SERVICES	PORTFOLIO	ABOUT	TEAM	CONTACT
<p>An introduction to software life cycle models; size estimation; cost and schedule estimation; project management; risk management; formal technical reviews; analysis, design, coding and testing methods; configuration management and change control; and software reliability estimation. Emphasis on large development projects. An individual project required following good software engineering practices throughout the semester.</p>				<p>Object-Oriented Design and Development</p> <p>Object-Oriented Languages and Systems</p> <p>DevOps</p> <p>Artificial Intelligence I</p> <p>Automated Learning and Data Analysis</p> <p>Computational Methods for Molecular Biology</p> <p>Database Management concepts and Systems</p> <p>Advanced Data Structures</p> <p>Management Decision and Control Systems</p> <p>Cloud Computing Technology</p>				
Semester	Professor	A%	TotalA%					
F12	Xia	46	54	85%				
F13	Murphy-Hill	63	72	87%				
F14	Parrin	103	108	95%				
F15	Murphy-Hill	50	59	84%				
F16	Parrin	85	97	87%				
S14	Dagino	38	38	100%				
S14	Dagino	18	22	81%				
S15	Hendrix	32	32	100%				
S15	Dagino	13	20	65%				
S16	Hendrix	26	57	45%				

Course Helper

SERVICES

PORTFOLIO

ABOUT

TEAM

CONTACT

GRADE DISTRIBUTION

Course ID	A%	B%	C%	Total Enrollment
501	67%	26%	3%	725
503	64%	27%	6%	303
505	51%	38%	8%	1437
506	80%	17%	1%	78
510	84%	13%	0%	559
512	85%	10%	2%	108
515	78%	21%	0%	265
517	80%	15%	0%	466
520	56%	36%	3%	325
522	59%	36%	3%	592
540	71%	24%	2%	799
541	68%	22%	6%	500
547	60%	39%	0%	69

Fig. 2. Wesbite pages showing a course page, and the grade distribution page

surveys, we put down a set of the most fundamental criteria a student uses. We provide sliders for a student to decide what they care about more, and the user clicks submit. A text mining algorithm runs through all the reviews and feedback posted on Piazza through every course and calculates a rating for each criteria. These ratings are matched against what the user selected, and a list of courses is given to a user in ascending order of the error, that is the closest match to a users input appears on top. Based on this the user can have his decision made a lot easier and they can save a lot of hours of looking through information by working through this mechanism in a matter of seconds.

Course Helper				SERVICES	PORTFOLIO	ABOUT	TEAM	CONTACT
TEXT MINING								
Please select your preferences on what other people thought about courses on the following categories.								
Professor Rating:	<input type="range"/>							
Grades:	<input type="range"/>							
Content:	<input type="range"/>							
Job Prospectives:	<input type="range"/>							
Work Load:	<input type="range"/>							
<input type="button" value="Submit"/>	222							

Fig. 3. Sliders for the text mining solution

II. DIVISION OF WORK

We took up an agile method of development for our project. Our project was managed using GitHub, and we created multiple Milestones and Issues to track the progress. Even though working on Agile methodology, we took up a couple of concepts from Spiral design model, specifically commit partitions. Our first step was to identify the parts our Project would not work without. We prioritized these, and came up with solutions before laying the roadmap for the whole project.

For the project implementation, we decided to take ownership of one implementation each, and work exclusively on them. This was mainly because, within the team, we had expertise in different domains. Since we had just about a month to come up with working demos, we thought it would be unwise to spend two weeks coming up with a common platform, and learning that. Instead, we allowed flexibility for each of us to implement our solutions in a language/medium of choice, and finally integrate them together. We also planned to meet together frequently to update each other, and to work on integrating our solutions.

The technologies we worked on together include Ruby, Rails, Python, Java, JavaScript and NodeJS.

III. EXPERIMENTS

Our first commit partition was to identify a forum which could enroll multiple users (we approximated in hundreds), and which the students in North Carolina State University are familiar with. We got down to two options - Piazza and Slack. We chose Piazza because of the following reasons

1. There is a Node JS API available for Piazza online, which we found really easy to use.
2. Piazza provides support for emails, polls, notes, file attachments, and collective answers which would together address the needs of multiple students.
3. We found that more students in North Carolina State University are familiar with Piazza as compared to Slack.
4. Piazza computes statistics of the group, like number of active students, average response time etc, which we could potentially use to evaluate our solutions.

Our second commit partition was to get the information about Courses (and respective Grade Distribution) offered in the past 3-4 years. The info was available on NCSU websites, but we could not find a straightforward method to get them. It would make our lives a lot easier if we could get read-only access to overall grade distribution data at the University. We shot the Computer Science department an email, to try our luck. As expected, we got a long email stating federal laws that prevent the Department from giving us access. The final resort was to parse HTML content, and store the available data in some sort of data structure. We used a Python library called BeautifulSoup to parse html content, and store them in Python lists and Dictionaries. Later, we imported these into a database and used it in our website.

The grade distribution was even more difficult to get, since the data could only be exported from the website in a .csv format. We had to manually download the data for the relevant semesters. We then wrote some java code which read the data, and imported them to the database.

On Piazza API part, we first found a Python API online, and did a pilot on this. Unfortunately, we could not get the API working, and there was minimal support on the API from the owners. Later, we found a NodeJS API for Piazza, and our pilot on this was successful.

On Piazza, we decided to create a folder each for Graduate Courses offered in the Computer Science Department. The folder names are the Course ID, which is usually a 3 digit

number. It should make it easier for students to search on the forum. (As opposed to typing the complete name of the subject). We added Instructions and Guidelines to use the forum, on the resources page. We inputted some random course review data on the forum, to test the pilot project, and check feasibility of the text mining solution.

The Piazza forum was the groundwork for our Text Mining solution. Our first challenge was to convert the course reviews and data (available in text format) from the Piazza forum into vectors (numbers), so that we could run some algorithm over it. We researched on the Python Natural Language Processing toolkit, and found that Doc2Vec Library, which we found was perfect for our application. We preferred Doc2Vec over other methods of converting text to vectors (like Word2Vec, BagofWords etc) because Doc2Vec actually takes the context of the data into consideration. The data from Piazza forum, might be responses to an initial question, or follow up answers. We wanted an algorithm which will actually consider these aspects too.

The second challenge was to relate student preferences, and the data. For this, we identified 5 factors - Professor Rating, Grade Distribution, Course Content, Job Prospects and Workload, which we found were relevant to students from our Survey done in January. For each factor, we elicit a requirement from the student on a scale of 0 to 1 (Bad vs Good). Using the data from the forum, we calculate the score for each course based of these factors on a scale of 0 to 1. (This basically indicates the opinion of other students on these factors). To get this score, we identified related words for each factor from our trained model, and compared them to a list of good and bad words we got from the internet. Out of n related words, if g are good words,

$$Score = \frac{g}{n}$$

which would be on a 0-1 scale.

For each course, we found the Squared Error between the requirement vector of the student and the actual score vector. The lesser the error, the more the course matches with the requirements of the student.

$$Error = \sum_{factors} (Requirement - Score)^2$$

On the website, we designed 3 main pages on Ruby on Rails - Course Details, Grade Distribution, and a page for Text Mining feature. We also designed a database model to represent the information in the website precisely. The features for each page are as detailed in the Introduction above.

IV. EVALUATION OF EXPERIMENTS

In order to have a better understanding of the effectiveness of the program and its potential problems, we invited 15 students in the Software Engineering class to use and test the application. We first gave these users a basic introductions to our application, then let them use the three different approaches one by one and track the time they spent on each one. Finally, we asked whether the program was helpful and quick enough for them to narrow down the courses they needed. To gather feedback information, we designed 12 questions on Google Form categorized by three different approaches.

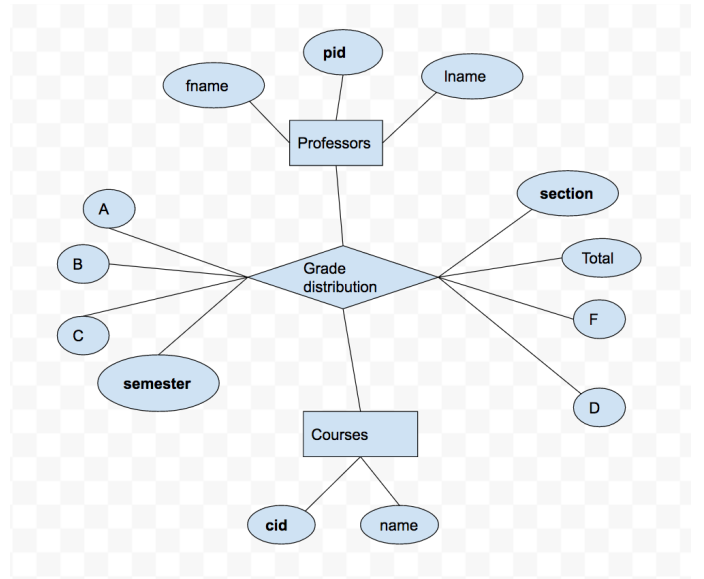


Fig. 4. ER Diagram for our database model

Rate how effective the website was in helping you get more information about your future courses. (5 = Very Effective, 1 = Not Effective)
(15 responses)

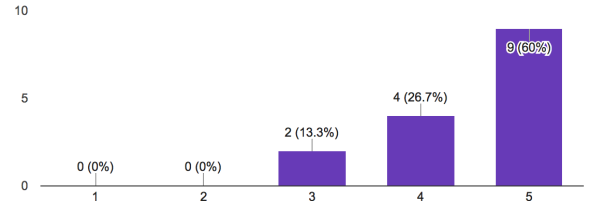


Fig. 5. Question 1

1. The Website

Based on the result, most users were satisfied with the usefulness of information presented on the website.

Do you think the website was extensive enough to cover all the criteria that you consider when trying to decide a course ? (5 = Covers almost everything I need to know, 1 = The information was of no use at all).
(15 responses)

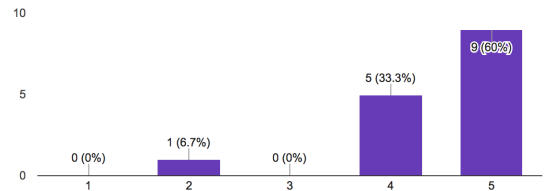


Fig. 6. Question 2

Also, over 93% users believed the information on the website were comprehensive. However, one user did make a negative feedback on this question - "Would like to see the sources and legitimacy of data, when the data is updated, things like that displayed clearly so that I can trust it ". We think this is a very good point that we need to improve in the future. A

website that provides statistical information does need to be more authoritative, accurate and professional.

Any comments about potential improvements to that solution ?

Your answer

Fig. 7. Question 3

Several users came up with their ideas:
 “Ratings by user based on a scale ”
 “I think if you could find a way to automate that data collection it would be very beneficial. Look into a tool called Tracster which provides that information. ”
 “Probably focus more on making the forum better in ways to make it easier for people to choose courses. ”
 “Focus on technical aspect instead of more graphic design. ”
 “More description on the result.
 “It was a great attempt just add some data for Text Mining ”

2. Piazza Form

Rate how helpful the Piazza forums were in terms of getting feedback and reviews. (5 = Very Helpful, 1 = Not Helpful).
 (15 responses)

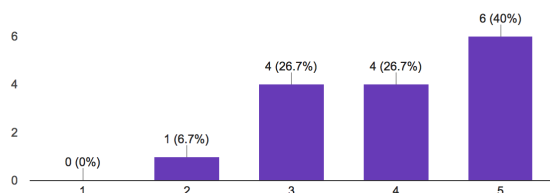


Fig. 8. Question 4

From the evaluation, we found the Piazza forum approach is not as popular as the website. Up to one third of the users are not satisfied with the forums. However, the results did not contain too much surprise for us. This is because the forums are merely created for a month, which means the number of users and posts is still limited. There is no doubt that the users in our evaluation section might have a difficult time finding useful information. But we believe this situation will have a turnaround as long as more and more new coming students ask questions and seniors make corresponding responses.

Do you have any additional comments on the forum ? (11 responses)

Fig. 9. Question 5

Some users provided very sincere opinions to us.

“It might be difficult to get enough people to contribute to the forum to get enough information. Probably a similar problem to all forums. You get a lot of lurkers and not many contributors. People may also get annoyed when being enrolled into something like this so it would have to be a voluntary effort, further limiting participants. ”
 “Would help if courses had internal segregation as well based on professors, because my experience has been that

each professor takes a course in his own way, differs a lot from professor to professor. Also can have some ratings like projects or class interaction etc ”

We think every student should know that choosing correct courses is crucial to his success, so most incoming students are likely to spend their time on knowing courses’ information. As long as we find a way to let them know the forum, those students must be willing to participating the discussions. Besides, we believe they will keep active on the forum since they need to find information for at least two more semesters. At that time, when new students come in, they can become the seniors who answer questions. ”

Indeed, different professors have different styles in teaching. In the future, we will invite more instrutors to join the forums. We believe it is mutual beneficial since instrutors hope to see the students in the class are the people who are really interested in the course. ”

Additionally, Piazza itself provided some statistics on the forum, which we could potentially use for evaluation. Presently, there isn’t enough data on the forum to draw any conclusions, but we assume it would be different one the forum picks up pace.

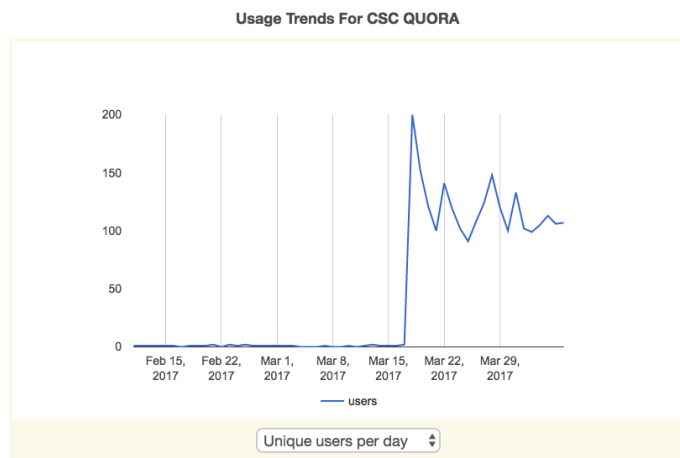


Fig. 10. Unique users per day on Piazza

Currently, we have about 30 students who have made some contribution on the forum, and around 100 students who have read more than 10 posts on the forum. So, its safe to say there is a community that believes the idea might have potential.

(3) Text Mining

Similar to the website approach, text mining is another very popular method. Even though there is not sufficient topics or replies on our Piazza forums, which is used for text mining, most people still gave us positive feedback on the results it returned.

As we introduce above, we set five criteria for the text mining – Professor Rating, Grades, Content, Job Prospectives, Work Load. The survey result suggests that more than 86 % people agree with this kind of settings.

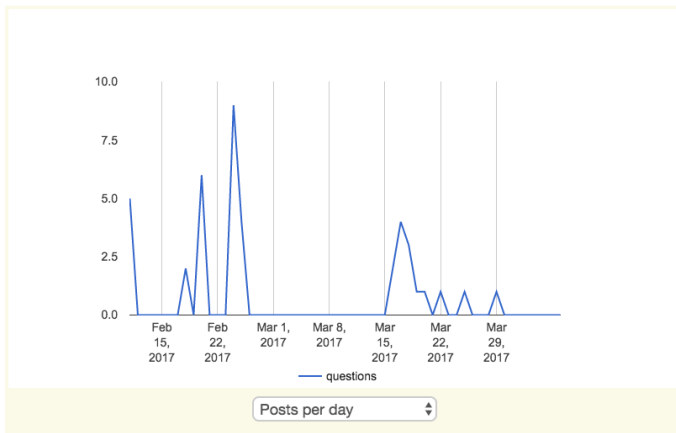


Fig. 11. Number of posts per day on Piazza

How effective do you think the text mining solution was ? (15 responses)

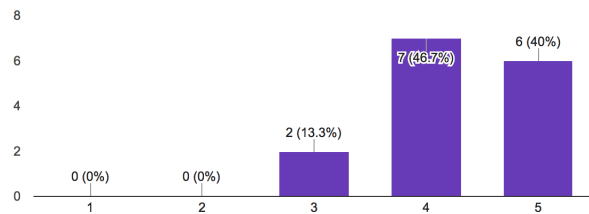


Fig. 12. Question 6

Was the set of criteria in the text mining solution enough? (15 responses)

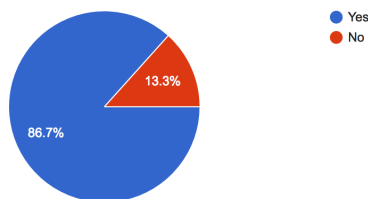


Fig. 13. Question 7

Did you think there was anything missing in the text mining solution, or any criteria we could add ? (12 responses)

Fig. 14. Question 8

Among the two users who held different ideas, one believe that "Based on professor, theory or practical workload are factors I look for", and the other said "Can add skills that needed in the course."

We do agree with these two suggestions. If possible, we plan to put skill tags on each course categorized by professors in the future, which will make the search results more comprehensive and accurate.

To summarize, website and text mining are the two most popular methods. Other than the implementation itself, we believe it has something to do with people's habit. Because

Select one of the three solutions that you think was the best/your favorite. (15 responses)

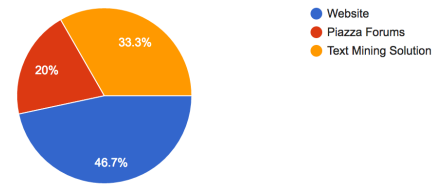


Fig. 15. Question 9

either website or text mining can provide you an answer very quickly. On the other side, using forum, although you might get more subjective and detailed answers, it also takes a lot of time to post and wait. In today's fast-pace life, there might not be as many people as before to keep patient.

V. CONCLUSION

Our main aim of this endeavour was to curb the problems new students have with course selection, at a fundamental level. The primary issues with course selection have always been lack of detailed descriptions, and lack of subjective feedback, which are incidentally, some of the most sought-after pieces of information. Through our three approaches, we have helped counter these issues. We created a website that provides as much information as we can gather, in an all-in-one convenient location. The data ranges from syllabus, to grade distribution information giving a student more hints about the course. Besides that, the most important feature of this solution is the time one would save looking up all the different aspects of course selection. Along with text mining, the website turned out to be quite popular in terms of utility and efficiency. The piazza forum is a great way for students to get in touch with their peers and get a more subjective outlook on the courses they want to enroll for. Feedback was one of the most important criteria for students and this forum has helped generate that for students. Thus, a combination of these three resources covers all the important criteria that every student needs to think about before enrollment. Overall, we hope we can make improvements to our systems and provide a better, and more convenient experience for the users.

VI. CHITS

fkkhueaa
fhjmauua
fmjviuoi
chjsioii
cddruueu
cjkqooaa
cklfuuee
bjltooeu
fjbvaiou
fckmouaa
cfgmuuui
djknueoo
dcmhiioo
dblbuaau

fjmbuoao