AN EMPIRICAL STUDY OF HOW NOVICE PROGRAMMERS USE THE WEB

Naveen Tula

Dissertation Prepared for the Degree of

MASTER OF SCIENCE

UNIVERSITY OF NORTH TEXAS

May 2016

APPROVED:

Dr. Renee Bryce, Major Professor
Dr. Bill Buckles, Committee Member and
Dr. Philip Sweany, Committee Member
Tula, Naveen. *An Empirical Study of How Novice Programmers Use the Web*. Master of Science (Computer Science), May 2016, 38 pp., 7 tables, 1 figure, references, 32 titles.

Students often use the web as a source of help for problems that they encounter on programming assignments. In this work, we seek to understand how students use the web to search for help on their assignments. We used a mixed methods approach with 344 students who complete a survey and 41 students who participate in a focus group meetings and helped in recording data about their search habits. The survey reveals data about student reported search habits while the focus group uses a web browser plug-in to record actual search patterns. We examine the results collectively and as broken down by class year. Survey results show that at least 2/3 of the students from each class year rely on search engines to locate resources for help with their programming bugs in at least half of their assignments; search habits vary by class year; and the value of different types of resources such as tutorials and forums varies by class year. Focus group results exposes the high frequency web sites used by the students in solving their programming assignments.
Copyright 2016
by
Naveen Tula
ACKNOWLEDGMENTS

I wish to express my gratitude to my thesis advisor, Dr. Renee Bryce, for her generous advice, inspiring guidance and encouragement throughout my research for this work. I would also like to thank the National Science Foundation for support from NSF Grant # 1251067
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ACKNOWLEDGMENTS</th>
<th>iii</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td>CHAPTER 1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 2 RELATED WORK</td>
<td>3</td>
</tr>
<tr>
<td>CHAPTER 3 EXPERIMENTAL SETUP</td>
<td>7</td>
</tr>
<tr>
<td>3.1. Data Collection</td>
<td>7</td>
</tr>
<tr>
<td>3.1.1. Survey</td>
<td>7</td>
</tr>
<tr>
<td>3.1.2. Focus Groups</td>
<td>9</td>
</tr>
<tr>
<td>3.2. Data Analysis</td>
<td>11</td>
</tr>
<tr>
<td>CHAPTER 4 RESULTS AND DISCUSSION</td>
<td>14</td>
</tr>
<tr>
<td>4.1. Survey</td>
<td>14</td>
</tr>
<tr>
<td>4.1.1. Outcome Variables</td>
<td>14</td>
</tr>
<tr>
<td>4.1.2. Resources and Their Helpfulness</td>
<td>15</td>
</tr>
<tr>
<td>4.1.3. Regression Analysis</td>
<td>18</td>
</tr>
<tr>
<td>4.2. Focus Group</td>
<td>20</td>
</tr>
<tr>
<td>CHAPTER 5 CONCLUSION &amp; FUTURE WORK</td>
<td>27</td>
</tr>
<tr>
<td>APPENDIX: SURVEY QUESTIONS</td>
<td>29</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>36</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 4.1. Mean and standard deviation of bivariate associations between class year, resources used, and other predictors and outcome variables. SPT = Solution Parameter Time; SWT = Solution Website Time; TNWP = Total Number of Web Pages

Table 4.2. Significance tests (T-test/F-test) of bivariate associations between class year, resources used, and other predictors and outcome variables.

Table 4.3. Bivariate associations between class year and other variables.

Table 4.4. Bivariate associations between class year and helpfulness of different website categories.

Table 4.5. The most visited websites by the focus group students.

Table 4.6. Website categories used by students in different class years.

Table 4.7. Search terms used by the different class year students for the common programming bugs.
Figure 3.1. A screenshot of the FG search engine: The left side shows fields for students to enter their focus group data and query. The right side shows the query results and buttons for the students to rate the respective pages that they visit.
CHAPTER 1
INTRODUCTION

Computer science is a challenging major for many students because it requires them to learn new concepts while also using a programming language to implement solutions based on those concepts. Programming is not an easy subject to study and requires proper understanding of concepts [17]. Many students struggle to instantiate solutions to problems because they do not fully understand course concepts and further struggle to understand a programming language. Unable to understand the concept or incapable to find a solution is one of the factor the computer science students drop-out from the courses. Understanding the concept relies on the resources used for the learning. Before the age of the Internet, students relied mostly on offline resources such as textbooks, class notes, and classmates to find the information regarding course material. The method of searching for a solution from the offline resources requires a significant amount of time and effort. Today, students use internet resources such as search engines, forums, or online journals to help them with their classes. Internet has capacity to diminish the interaction constraints of traditional classroom learning and can add and retain the knowledge by providing the additional academic resources [2]. However, the process of exploring online resources also requires a noticeable amount of time and effort. Often, students check multiple websites before they can find some relevant information. To interpret the student behavior in using the web for help to solve the programming assignments, we started this study.

In this work, we seek to understand how students search for help when they struggle with course topics and programming bugs. This study involves two parts: survey of 344 students and focus group consisting of 41 students. We conducted the survey on the undergraduate courses which has course work of programming since our research work deals with the students programming assignments and their way of solving through web. Out of 713 students, 344 students responded to the survey and given an extra credit for participation. This survey identifies the different types of resources that students use and the perceived
value of those resources. For instance, students may use some combination of textbooks, course notes, office hours, the web, or other sources of help. We then focus on one particular resource that more than 2/3 of our students from all class years reportedly use on more than half of their assignments – the web. We started the paid focus group study with the 41 students which includes three focus group meetings and each student is asked to use the Focus Group search engine which is designed to record their search terms and websites. Students were selected on the basis of first come first serve. Recorded websites are used to classify the most visited websites by the students. We used the recorded search terms to classify the type of mostly faced bug and then grouped them by the class year. There are several objectives for this study: we can understand relationship between the search effectiveness factors and resources used; We can look at the search terms used by the different class year students for their Programming assignments; Moreover, we can get the students views on the programming and challenging concepts and we can also suggest the students the best websites to use for solving their programming problems. We dig deeper to better understand how students interact with search engines to gather assistance on the web with their assignments. We believe that this knowledge may help us to better assist future students who prefer to use the web to search for help with programming bugs.
CHAPTER 2

RELATED WORK

Most students encounter bugs while working on programming assignments. Fixing bugs in programming assignments has always been a major task and part of the learning process. A significant body of literature exists to classify the types of bugs that students encounter [3, 11, 15] and ways to improve debugging efficiency [6, 19, 10]. For instance, Chmiel et al. [5] report three categories of bugs that programmers encounter: syntax, semantic, and logical errors. It has been shown that students who learn multiple debugging techniques spend less time resolving common programming bugs [5]. While debugging is extremely valuable, we focus on students who choose to use the web to search for help rather than to necessarily use a debugger. Indeed, students may use books, lecture notes, the web, and any other resources that fit their personal preferences and learning style. We hypothesize that many students are drawn to the web to search for help and that some students may be more effective than others in finding valuable help on the web. This knowledge may help us to better understand and assist future students who use the web as a source of help to solve programming bugs on their assignments.

Several approaches exist to keep track of student coding patterns and bugs. Fenwick et al. [8] use their ClockIt Data Logger to track student coding patterns and analyze the behaviors of programming in CS1 and CS2 courses. They observe that students who start early and work on their assignments regularly score higher grades. In our study, we do not focus on the amount of time that students spend on programming assignments, but rather the amount of time that they spend seeking help on the web, online resources that they utilize, and how this corresponds to their success.

There are many causes of programming bugs. Similarly, there are many different approaches and resources that students pursue to solve programming bugs. The most important step is for students to understand the logic behind a problem and solution before solving a bug. On the other hand, a student may understand the logic and instead struggle
with the programming language or APIs. While some students will use a debugger to trace their code to better understand the problem, it is also common for students to seek help from their instructors and peers. These discussions may help students to better understand the material, identify specific problems in their code, and develop solutions. In addition to in-person tutoring from instructors and peers, Metzger [20] reports that the Internet is a major source of assistance. Students use search engines, forums, online journals, and videos to help them with their programming assignments. Students may check multiple websites before they find relevant information. The process may take even longer if students cannot define the problem and important search terms effectively. This process consists of two parts. The first part requires a student to formulate the problem and the second part requires students to understand how to effectively search for help to solve their problem. The second part is dependent on the first part of the process. Nasehi et al. [21] report that formulating a question is the first step toward finding relevant solutions to a question. Moreover, defining the areas of the problem that a student finds difficult helps to narrow the necessary search terms. Not much is known about how students search or the keywords that they use to search the web for help to solve their programming assignments.

Searching the web is a challenging process for many people [27]. This is particularly problematic for introductory programming students who struggle with multiple layers of confusion, including understanding important abstract concepts for problem solving and expertise with a programming language. As frustration accumulates, students may view coding as a difficult process and give up early if they do not find a solution quickly [24]. Giving up on a solution to their programming problem results in low course grades and decreased satisfaction with the course [14]. Consequently, students lose interest in a course and drop-out [25]. On the other hand, many students successfully make it through their courses and we need to better understand resources that they use to support their success.

Previously, there has been study on the high school students web searching behavior. Wallace et al. [31] observed that students do not show much interest in exploring the web results but prefer to find the answers or source for their assignments rather than understanding
it. Schacter et al. [26] analyzed that students do not plan or apply any searching strategies for searching they tend to browse for information seeking. A study trained the students as well as teachers to search the web effectively while they are working on their assignments [9]. In our study, we seek the web searching behavior of undergraduate students and suggested the best websites to solve their programming assignments. Novice programmers face more programming bugs if the programming concepts were difficulty in understanding. In a survey analysis, novice programming students indicated that they face difficulty in understanding and implementing the high-level concepts such as OOP principles and efficient program design [4]. In our focus group meeting, we had a discussion on the programming challenging concepts and students responded that high-level concepts such as multi-dimensional arrays, pointers, recursion and memory management are hard to understand. A study found that the interactions of domain knowledge and web experience that determines the novices and experts search behavior: web experts with domain knowledge were able to select the target document where as domain experts who are novice depended on terminology failed to reach the target document most of the times [13]. A study conducted on 48 students consisting of undergraduates, masters and PhD students to analyze the web searching behavior for academic resources [1]. The study showed that user needs, skills and persistence increases as education level increases. Park et al. investigate how search behavior factors such as search outcome precision, searching time, number of URL nodes, and the number of key words is correlated with the searching experience and gender [23]. In our study, we similarly examine factors related to search outcomes, but instead focus on the strategies, time to find solutions, and types of websites that are helpful to students of different class years. We conduct a large survey with 344 students that report data about their experience with using the web as a source of help for programming bugs and then conduct a focus group that allows a more detailed examination of the search strategies and resources that freshman, sophomore, junior, and senior computer science students use.

In our study, the survey and focus group results states that the students use most commonly forum site stackoverflow.com to learn the programming concepts and solutions. There
were studies demonstrating that Q&A websites such as stackoverflow are providing means for programmers to participate in social learning and also effectively answering the novice programmers questions [21, 28]. In addition to the stackoverflow.com students use the other tutorial sites discussed in the results section.
CHAPTER 3

EXPERIMENTAL SETUP

In our study we use a mixed methods setup with a survey and focus group. Our subjects include undergraduate students enrolled in CS1, CS2, programming languages, and software engineering courses.

3.1. Data Collection

3.1.1. Survey

We conducted an online survey to the 713 undergraduate computer science students in our department and received a total of 344 responses [30]. We emailed the survey link to students and also visited the class sessions to ask them to complete the survey. Some instructors offered extra credit for completing the survey.

The survey collected information to assess students’ efficiency in solving programming bugs (outcome of interest) measured by:

(1) Search Parameter Time (SPT): time taken to define search parameters for the problem.

(2) Solution Website Time (SWT): time taken to find the solution website. Six categories were provided to students to indicate the time taken to find solution website:
   (a) < 2 min
   (b) 2-5 min
   (c) 5-10 min
   (d) 10-15 min
   (e) > 15 min
   (f) Other

(3) Total Number of Webpages (TNWP): the total number of web pages they visited before finding the solution.

Students were also surveyed on
(1) The type of resources used:
   (a) Textbook
   (b) Class Lecture Slides
   (c) Class Notes
   (d) Search Engine
   (e) Specific website
   (f) Other

(2) If students used the focus group website for:
   (a) All assignments
   (b) Most assignment
   (c) A half of assignments
   (d) Less than a half of assignments

(3) How students decided what to search:
   (a) Based on assignment
   (b) Based on error
   (c) Based on other factors

In addition, students also were asked to report on

(1) Usefulness of a website
   (a) Not Helpful
   (b) Somewhat Helpful
   (c) Helpful
   (d) Very Helpful
   (e) Extremely Helpful

(2) Demographics
   (a) Gender
      (i) Male
      (ii) Female
   (b) Class Year
(i) Freshman
(ii) Sophomore
(iii) Junior
(iv) Senior

(3) Expected grade in the course
   (a) A
   (b) B
   (c) C
   (d) D or lower

3.1.2. Focus Groups

We invited all of the students in our department to participate in the focus group and accepted students on a first-come/first-serve basis [16]. We accepted a total of 50 students. A total of 41 students completed all of the requirements and are included in the results section. The distribution of students includes: 11 freshman, 14 sophomore, 10 junior, and 6 senior.

Each student participant attended an orientation and used a web browser plug-in that we named the Focus Group (FG) search engine. The purpose of the FG search engine was to record data about students’ searches and to infer ratings of usefulness for each respective URL. We observed students’ progress remotely each week and met with them in the middle and at the end of the semester to review the progress.

Figure 3.1 shows a sample screenshot of the FG search engine that records student search patterns and website URLs. The engine retrieves information from Google using NodeJS and displays the results in a similar way to the Google search engine. The engine also captures the following information:

(1) Student name,
(2) Assignment goal,
(3) Problem description, and
(4) Search term used.
Figure 3.1. A screenshot of the FG search engine: The left side shows fields for students to enter their focus group data and query. The right side shows the query results and buttons for the students to rate the respective pages that they visit.

Once a student finds information to help solve their problem, they use the “solution button” on the right side of each search result to rate the usefulness of the respective URL.

For a cumulative analysis, student identifiers were stored on a secure machine. The Assignment Goal provides insight into the student’s understanding of their assignment. The Problem Description gives a description of the bug that the student believes exists in the code. They may optionally leave this field blank if they are not sure. Examples include misunderstandings of an API, a null pointer exception error and how to handle it, and questions about syntax. The Search Terms include the actual search terms or phrases that the student enters during their attempt to find helpful information. In addition to this data, we also collected and stored information regarding the links visited by students, how many times a given link was re-visited, the time that it took for the students to describe the problem, and to find a solution.
In addition, we interviewed the 41 students and took their opinions on the course work. During the focus group meeting, we asked the below questions to students and took their opinions:

1. Why did you choose computer science/engineering as a major and was it a difficult choice between that and something else?
2. When you receive a programming assignment, do you begin programming immediately or do you research the assignment first?
3. When you researched a programming assignment this semester, what did you look up?
   a) The entire problem description?
   b) How to implement a specific part of the program?
   c) Any new data structures or functions the program requires?
4. When you came across a bug or a problem in your program, what did you do first?
5. This semester, when you turned to the internet, did you go to a specific website first, or go straight to a search engine? Which ones and why?
6. About how much time do you think you spend solving each bug? Has it gotten easier over the course of the semester?
7. What concept do you find the most challenging?

These questions were repeated in each focus group and analyzed whether students have improved their performance. We used the same questions so we could understand not only students progress but also their perspectives on the problems they were solving.

3.2. Data Analysis

We analyzed the survey data using SPSS statistics, a statistical analysis software [7]. Prior to running the statistical tests, we performed data cleaning on the survey data to ensure accuracy and completeness, to identify missing data patterns, and to regroup levels when extreme values were noted. We have done the data cleaning process in three steps: screening, diagnosing and editing of data abnormality [29]. We searched for the data errors/typos made
by the students and fixed them manually. Univariate analysis is simply defined as analysis of cases that varies with the single variable. Univariate analysis [12] was conducted for all variables to examine frequency distributions, to estimate proportions when variables are discrete, means and standard deviations when variables are continuous. Bi-variate analysis is defined as analysis of two-variables and relationship between them. Bi-variate associations [18] between the three outcome variables, predictor variables and demographics were tested. T-test and Analysis of variance (ANOVA) compare group means. T-test is to compare the means of two groups where as one-way anova can compare more than two groups. T-tests [18] were used to test the mean difference between a two-level categorical variable and a continuous variable. One-way ANOVA [18] is used if the categorical variable is more than two-levels. The left hand side (LHS) variable to be tested should be interval or ratio, whereas the right-hand side (RHS) variable should be binary (categorical) [22]. Bi-variate associations between independent and dependent variables were investigated using cross tabulations with chi-square option [18] (for categorical variables) and significance level set at 0.05. In general, chi-square is used to determine whether a variable has a frequency distribution compared to the one expected. Significance in mean differences were examined for class year and solution website time, search parameter time, number of web pages, forums, example code and tutorials.

To identify significant predictors, linear regression models were run separately for three outcome variables of interest, namely

1. Solution Website Time (SWT)
2. Search Parameter Time (SPT)
3. Total Number of Web Pages (TNWP)

Based on the results from bi-variate analyses, regression models included demographic and predictor variables that were associated with the outcomes. Regression analysis helps one to understand how a typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are fixed. Regression analysis is a statistical technique for studying linear relationships. The general form for the
linear relationship is:

(1) \[ Y = \beta_1 X_1 + \ldots + \beta_k X_k + \epsilon \]

Where \( Y \) is the dependent variable and \( X_1, X_k \) are the independent predictor variables. \( \beta \) represent parameters to be estimated, and \( \epsilon \) is the independent identically distributed normal error [32]. The results were then corrected for the influence of potential confounders, in this case gender and class level. A p-value < 0.05 was considered statistically significant.
The result section categorized into two sections: (4.1) Survey results from 344 students and (4.2) Focus group from 41 students. Our major findings include that the vast majority of our students use the web for help with at least half of their programming assignments. However, the types of programming bugs and value of different types of resources varies as students progress through the program. For instance, students of all class years use online tutorials, but the use of forums increases as students progress through our program. In the remainder of this section, we review these results in more detail.

4.1. Survey

A total of 344 students responded to the survey of which 77% were males. The distribution of the students by class year was as following:

(1) 25% of students were freshman
(2) 36% of students were sophomore
(3) 14.5% of students were junior
(4) 25% of students were senior

The distribution of expected grade was as following:

(1) 38% of students expected A
(2) 48% of students expected B
(3) 12% of students expected C
(4) 2% of students expected D or lower

4.1.1. Outcome Variables

To assess students efficiency in solving debugging problems (outcome of interest) we measured three indicators:

(1) Search Parameter Time (SPT): The mean time taken to define search parameters for the problem was 7.5 minutes (SD±8.3) with a maximum of 30 min.
(2) Solution Website Time (SWT): Time taken to find the solution website. Six response categories were provided to students to indicate time taken to find the solution website. Each category was enumerated from 1 to 5 and the average category to find the solution website was 2.84 which is between categories b and c, i.e., over 5 min but less than 10 minutes.
(a) 14% indicated that they took < 2 min
(b) 33% took 2-5 min
(c) 23% took 5-10 min
(d) 14% took 10-15 min
(e) 16% took more than 15 minutes
(3) Total Number of Webpages (TNWP): the mean number of web pages students visited before finding the solution was 3.6 (SD±2.0) with the maximum number of 20 pages.

4.1.2. Resources and Their Helpfulness

The distribution of resources that students used is as following:
(1) 71% used class lecture slides
(2) 69% used search engines
(3) 55% used class notes
(4) 39% used textbooks
(5) 37% used specific websites
(6) 7% used other resources

The demonstrated results reflect the fact that a resource could be used alone or in combination with other resources.

The distribution of the basis for how students were searching for a solution is
(1) 62% were searching by a specific error
(2) 23% were searching by an assignment topic
(3) 15% were searching by other criteria
The distribution of students who used web resources is as following

1. 36% for the most of their assignments
2. 25% for less than a half of their assignments
3. 21% for a half of their assignments
4. 18% for all of their assignments

Students reported on the level of helpfulness (on a scale from 1 to 5) of the places they visited on the web. The mean of helpfulness was

1. 3.7 (SD±1.04) for forums with the questions similar to their assignment
2. 3.76 (SD±1.06) for websites with examples of the code
3. 3.74 (SD±1.14) for websites with tutorials

We used t-tests and one-way ANOVA to identify significant associations between class level, three outcomes, and other study variables of interest. The results are presented in Tables 4.1, 4.2, 4.3, and 4.4.

Results from the bivariate associations between class year, resources used, and other predictors and outcome variables are reported in Tables 4.1 and 4.2. In these tables, means, standard deviation, F, T and significant difference values are calculated for the variables.

The survey gives the relationship between the search effectiveness factors: (1) Search Parameter Time (SPT), Solution Website Time (SWT), and Total Number of Webpages (TNWP) and (2) resources used and their value for helping to solve programming bugs. The results of t-tests and one-way ANOVA (Tables 4.1 and 4.2) show that textbook use is associated with all three outcome variables. Those who used their textbook on average report taking more time to define search parameters for the problem and to find the solution website; and searched a larger number of web pages. Those who searched the web for help also reported taking more time to define search parameters for the problem and searched a larger number of web pages. The mean number of minutes increased with the number of assignments. Bivariate correlation analyses between three outcome variables and the helpfulness of websites with (1) forums, (2) sites with example code, or (3) tutorials indicated no significant relationship (not shown in the table). Class year was significantly associated
Table 4.1. Mean and standard deviation of bivariate associations between class year, resources used, and other predictors and outcome variables. SPT = Solution Parameter Time; SWT = Solution Website Time; TNWP = Total Number of Web Pages

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Outcome Variables (mean; SD±)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SPT (min.)</td>
</tr>
<tr>
<td>Class Year</td>
<td></td>
</tr>
<tr>
<td>Freshman</td>
<td>7.8 (9.2)</td>
</tr>
<tr>
<td>Sophomore</td>
<td>8.4 (8.8)</td>
</tr>
<tr>
<td>Junior</td>
<td>6.0 (6.4)</td>
</tr>
<tr>
<td>Senior</td>
<td>6.8 (7.6)</td>
</tr>
<tr>
<td>Resources Used</td>
<td></td>
</tr>
<tr>
<td>Textbook</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>9.2 (9.9)</td>
</tr>
<tr>
<td>No</td>
<td>6.4 (6.9)</td>
</tr>
<tr>
<td>Class Notes</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>8.3 (8.9)</td>
</tr>
<tr>
<td>No</td>
<td>6.6 (7.5)</td>
</tr>
<tr>
<td>Search Engine</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>7.2 (8.2)</td>
</tr>
<tr>
<td>No</td>
<td>8.3 (8.7)</td>
</tr>
<tr>
<td>Specific Website</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>7.5 (8.4)</td>
</tr>
<tr>
<td>No</td>
<td>7.5 (8.3)</td>
</tr>
<tr>
<td>Class Lecture Slides</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>7.9 (8.6)</td>
</tr>
<tr>
<td>No</td>
<td>6.6 (7.8)</td>
</tr>
<tr>
<td>Decision to Search by</td>
<td></td>
</tr>
<tr>
<td>Assignment</td>
<td>8.7 (9.7)</td>
</tr>
<tr>
<td>Error</td>
<td>7.0 (7.6)</td>
</tr>
<tr>
<td>Used Websites for a Portion of Assignments</td>
<td></td>
</tr>
<tr>
<td>Less than a Half</td>
<td>5.4 (7.4)</td>
</tr>
<tr>
<td>A Half</td>
<td>7.1 (7.6)</td>
</tr>
<tr>
<td>Most</td>
<td>9.0 (8.8)</td>
</tr>
<tr>
<td>All</td>
<td>7.6 (9.1)</td>
</tr>
<tr>
<td>Used Focus Group Website for a Portion of Assignments</td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>7.3 (8.2)</td>
</tr>
<tr>
<td>Less than a Half</td>
<td>8.3 (8.5)</td>
</tr>
<tr>
<td>A Half</td>
<td>6.6 (7.6)</td>
</tr>
<tr>
<td>All</td>
<td>8.7 (12.2)</td>
</tr>
</tbody>
</table>
Table 4.2. Significance tests (T-test/F-test) of bivariate associations between class year, resources used, and other predictors and outcome variables.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Outcome Variables (significance)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SPT</td>
</tr>
<tr>
<td>Class Year</td>
<td>F=1.35</td>
</tr>
<tr>
<td>Resources Used</td>
<td></td>
</tr>
<tr>
<td>Textbook</td>
<td>T=3.1**</td>
</tr>
<tr>
<td>Class Notes</td>
<td>T=1.9</td>
</tr>
<tr>
<td>Search Engine</td>
<td>T=1.2</td>
</tr>
<tr>
<td>Specific Website</td>
<td>T=0.6</td>
</tr>
<tr>
<td>Class Lecture Slides</td>
<td>T=1.2</td>
</tr>
<tr>
<td>Decision to Search by</td>
<td>T=3.0</td>
</tr>
<tr>
<td>Used Websites for a Portion of Assignments</td>
<td>F=3.0*</td>
</tr>
<tr>
<td>Used Focus Group Website for a Portion of Assignments</td>
<td>F=0.3</td>
</tr>
</tbody>
</table>

* p<0.05
** p<0.01
*** p<0.001

with the use of class notes; search engine and lecture slides and notes; whether students decided to searched by the assignment or by error; and helpfulness of web sites with (1) forums and (2) sites with tutorials.

4.1.3. Regression Analysis

Regression analysis was performed for all three outcomes, i.e. predicting time taken to define search parameters for the problem, time to find the solution website and the number of pages visited before finding the website containing a solution. The regression model intakes the predictors which are significantly associated at the bivariate level with the out-
Table 4.3. Bivariate associations between class year and other variables.

<table>
<thead>
<tr>
<th>Resources Used</th>
<th>Freshm.</th>
<th>Soph.</th>
<th>Junior</th>
<th>Senior</th>
<th>$\chi^2$ (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Textbook</td>
<td>36.5%</td>
<td>41.1%</td>
<td>38.0%</td>
<td>40.0%</td>
<td>0.53 (3)</td>
</tr>
<tr>
<td>Class Notes</td>
<td>65.0%</td>
<td>57.3%</td>
<td>50.0%</td>
<td>45.0%</td>
<td>7.6* (3)</td>
</tr>
<tr>
<td>Search Engine</td>
<td>63.5%</td>
<td>64.5%</td>
<td>60.0%</td>
<td>87.1%</td>
<td>17.3*** (3)</td>
</tr>
<tr>
<td>Specific Website</td>
<td>43.5%</td>
<td>38.0%</td>
<td>28.0%</td>
<td>35.3%</td>
<td>3.4 (3)</td>
</tr>
<tr>
<td>Class Lecture Slides</td>
<td>83.5%</td>
<td>69.4%</td>
<td>76.0%</td>
<td>57.0%</td>
<td>14.6** (3)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Decision to Search by</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignment</td>
<td>25.0%</td>
<td>37.4%</td>
<td>48.0%</td>
<td>22.4%</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>75.0%</td>
<td>62.6%</td>
<td>52.0%</td>
<td>77.6%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Used Websites for a Portion of Assignments</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than a Half</td>
<td>26.0%</td>
<td>24.2%</td>
<td>34.0%</td>
<td>19.0%</td>
<td></td>
</tr>
<tr>
<td>A Half</td>
<td>28.0%</td>
<td>20.2%</td>
<td>18.0%</td>
<td>16.5%</td>
<td></td>
</tr>
<tr>
<td>Most</td>
<td>34.0%</td>
<td>36.3%</td>
<td>32.0%</td>
<td>41.0%</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>12.0%</td>
<td>19.4%</td>
<td>16.0%</td>
<td>23.5%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Used Focus Group Website for a Portion of Assignments</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Never</td>
<td>77.6%</td>
<td>73.4%</td>
<td>76.0%</td>
<td>74.1%</td>
<td></td>
</tr>
<tr>
<td>Less than a Half</td>
<td>15.3%</td>
<td>19.4%</td>
<td>20.0%</td>
<td>14.1%</td>
<td></td>
</tr>
<tr>
<td>A Half</td>
<td>6.0%</td>
<td>2.4%</td>
<td>4.0%</td>
<td>9.4%</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>1.2%</td>
<td>4.8%</td>
<td>0.0%</td>
<td>2.4%</td>
<td></td>
</tr>
</tbody>
</table>

* $p<0.05$
** $p<0.01$
*** $p<0.001$

come variables (SPT, SWT, TNWP). Table 4.2 gives us the predictors that are significantly associated at the bivariate level with the Search Parameter Time (SPT), Solution Website Time (SWT) and Total Number of Webpages (TNWP).
Table 4.4. Bivariate associations between class year and helpfulness of different website categories.

<table>
<thead>
<tr>
<th>Websites</th>
<th>Mean (SD±)</th>
<th>F-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Freshm.</td>
<td>Soph.</td>
</tr>
<tr>
<td>Forums</td>
<td>3.59 (1.1)</td>
<td>3.51 (1.0)</td>
</tr>
<tr>
<td>Sites with Code Examples</td>
<td>3.76 (1.1)</td>
<td>3.61 (0.98)</td>
</tr>
<tr>
<td>Tutorial Sites</td>
<td>3.64 (1.2)</td>
<td>3.71 (1.1)</td>
</tr>
</tbody>
</table>

* p<0.05  
** p<0.01  
*** p<0.001

The regression model predicting Search Parameter Time (SPT) included textbook use, web help, decisions made based on the assignment or programming error, whether they visited the web for help on assignments, and helpfulness of sites with forums. Gender and class year also were included as control variables. Only the textbook use (beta=0.13; p=0.015) and web help (beta=0.14; p=0.011) significantly predicted search parameter time after adjusting for class level. The regression model for time taken to find the solution website included textbook use and decision (made based on the assignment or programming error) as those were significantly associated at bivariate level. Only textbook use significantly predicted (beta=0.21; p<0.001) after adjusting for gender and class year. Regression analysis for number of pages suggested that textbook use (beta=0.14; p=0.01) and whether they had web help for assignments (beta=0.16; p=0.004) significantly predicted after adjusting for gender and class year.

4.2. Focus Group

The Focus group results reveal data about actual student searches and the student perceived value of the webpages that they visited. Table 4.5 contains the most visited websites. Out of the total 450 searches, 35% of students visited the website stackoverflow.com, 27% visited cplusplus.com, 4% visited tutorials.com, 2% visited cprogramming.com, and 2%
Table 4.5. The most visited websites by the focus group students.

<table>
<thead>
<tr>
<th>Websites</th>
<th>Percentage of Visits</th>
</tr>
</thead>
<tbody>
<tr>
<td>stackoverflow.com</td>
<td>35%</td>
</tr>
<tr>
<td>cplusplus.com</td>
<td>27%</td>
</tr>
<tr>
<td>tutorialspoint.com</td>
<td>4%</td>
</tr>
<tr>
<td>cprogramming.com</td>
<td>2%</td>
</tr>
<tr>
<td>youtube.com</td>
<td>2%</td>
</tr>
<tr>
<td>en.cppreference.com</td>
<td>2%</td>
</tr>
<tr>
<td>dreamincode.net</td>
<td>2%</td>
</tr>
</tbody>
</table>

The majority of students find most of their solutions on the stackoverflow.com and cplusplus.com websites. The high frequency of usage encourages us to explore and recommend these sites to future students who seek online resources for help with solving their bugs.

We classified the usage of forum and tutorial websites by the different class year students. Forum sites such as stackoverflow.com, forums.codeguru.com, etc., are discussion sites where students view the conversation of solution to a programming bug. Tutorial sites such as cplusplus.com, tutorialspoint.com, etc., are learning sites where students seek information about concepts with the examples. Table 4.6 shows the percentage of focus group students who used the forums and tutorial websites. Out of all the forum sites visited by the focus group students, 37% of forum sites visited by the freshman, 31% by sophomore, 50% by junior and 65% by seniors. Out of all tutorial sites visited by the focus group students, 59% of tutorial sites visited by freshman, 62% visited by sophomore, 41% by junior and seniors being highest of 27%. There is a clear trend that both forum and tutorial style websites are useful to students. However, tutorials are more frequently visited by freshman and sophomore while forums are more popular with juniors and seniors.

We initially inferred that freshman and sophomores are more interested in learning
Table 4.6. Website categories used by students in different class years.

<table>
<thead>
<tr>
<th>Class Year</th>
<th>Forums(%)</th>
<th>Tutorials(%)</th>
<th>Other(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshman</td>
<td>37</td>
<td>59</td>
<td>4</td>
</tr>
<tr>
<td>Sophomore</td>
<td>31</td>
<td>62</td>
<td>8</td>
</tr>
<tr>
<td>Junior</td>
<td>50</td>
<td>41</td>
<td>9</td>
</tr>
<tr>
<td>Senior</td>
<td>65</td>
<td>27</td>
<td>8</td>
</tr>
</tbody>
</table>

from the tutorial sites whereas the junior and seniors have an increased preference in forums. However, the juniors and seniors still rely quite heavily on tutorials even though this is not typically their top source for finding helpful information to solve bugs. During debriefings at the end of the semester, we learned that many juniors and seniors were seeking the tutorial style sites when they were working in new languages or with new APIs.

Classification of programming bugs is done based on the assignment goal, problem description and search term used in the web browser plug-in. The most commonly faced programming bugs are grouped by the class year level with their search terms used for the corresponding bug. Table 4.7 gives us the search terms used by the different class year students for the commonly faced programming bugs such as syntax, logic, classes & functions, and strings.

We used the search terms in our database to classify the types of programming bugs that students searched for help. Logic bugs are the most common type of bugs faced by all the class year students. Freshman were inexperienced in using the loops and switch statements and face bugs related to these topics. Syntax errors are common in all the class years and mostly faced by the sophomores and juniors. While we were initially surprised that sophomores and juniors had so many difficulties with syntax errors, we learned that many of these students were struggling with new programming languages and were often quite effective in querying for the help that they needed. As the class year progresses, students do not face the bugs related to pointers. We attribute this partly to experience and changes in
Table 4.7. Search terms used by the different class year students for the common programming bugs.

<table>
<thead>
<tr>
<th>Bug Category</th>
<th>Freshman</th>
<th>Sophomore</th>
<th>Junior</th>
<th>Senior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntax</td>
<td>string to int</td>
<td>when to use ampersand in c++ function</td>
<td>branch git commands</td>
<td>cannot convertˈdouble ((\ast)(\text{double}))ˈ to ˈdoubleˈ</td>
</tr>
<tr>
<td>Logic</td>
<td>linker command failed with exit code 1</td>
<td>book checkout c++</td>
<td>std:: runtime error</td>
<td>rails pass variable to partial</td>
</tr>
<tr>
<td>Classes &amp; Functions</td>
<td>array of a class as member variable of class c++</td>
<td>pass array of structures to function in c</td>
<td>c++ put function fstream</td>
<td>responders respond_with delete redirect</td>
</tr>
<tr>
<td>Strings</td>
<td>cppreference size of a string</td>
<td>c++ cast an int as a string</td>
<td>atoi() syntax</td>
<td>c# SQL connection string</td>
</tr>
</tbody>
</table>

the programming languages that students used in their courses. Freshman faced many bugs related to classes and functions, i.e., bugs related to passing of variables to functions, wrongly instantiating an object, etc. Juniors and seniors face the least number of bugs related to classes and functions. Bugs related to arrays & vectors, file I/O and strings are prevalent among all the class year students. Computer environment bugs are mostly faced by the seniors as this bug occurs when working with the different platforms such as connectivity to databases, use of TCP/IP protocols, etc.
Our Focus group meeting discussion gives the students view on programming. Most of the students gave positive responses of the course, and few students reported that few programming concepts were challenging for them. Below are the some of the students perspectives:

(1) Why did you choose computer science/engineering as a major and was it a difficult choice between that and something else?

One of the common reason students interested in taking the computer science is that their curiosity in solving the puzzles and it would be same as the solving the coding problems. The students also indicated that since computer science is a growing field, they felt it would be a good choice for a major.

(2) When you receive a programming assignment, do you begin programming immediately or do you research the assignment first?

27 out of 41 of the students pointed that they researched the assignment before they start. Some of the students stated that it depends on the assignment whether to jump into it or research before.

(3) When you researched a programming assignment this semester, what did you look up?

(a) The entire problem description?
(b) How to implement a specific part of the program?
(c) Any new data structures or functions the program requires?

Nearly 30 students indicated that they look up with the problem description and find the code example similar to their assignment. Very few students said that they look for the main functionality part of the assignments.

(4) When you came across a bug or a problem in your program, what did you do first?

More than half of the students reported that they google it and few students stated that they use the print statements to track the bug and try to fix it if they could or else they would go to stack overflow.

(5) About how much time do you think you spend solving each bug? Has it gotten easier
over the course of the semester?

28 out of 41 students agreed that simple bugs can take five minutes. One student indicated that a bug with many logic errors once took 3-4 hours to figure out. Some reported that it depends on the bug. Everyone stated that it has gotten easier in solving the bugs over the course of the semester.

(6) This semester, when you turned to the internet, did you go to a specific website first, or go straight to a search engine? Which ones and why?

33 students stated that they went straight to a search engine (Google) first because it returns a variety of results. 6 students indicated the ‘cplusplus.com’ for learning the concepts and ‘stackoverflow.com’ for fixing the errors or looking at the syntaxs.

(7) What concept do you find the most challenging?

Majority of the Students indicated that high-level concepts such as multi-dimensional arrays, pointers, recursion and memory management were the most challenging.

We asked below questions same as the survey questions to analyze the quantitative and qualitative results and found out that results were almost similar.

(8) When using the search engine, how did you decide to search for? Do you search by assignment topic or search by error or any other specific type of search do you use?

Less than half of the students reported that they search by error or bug they are facing. Most of the students answered that they search by assignment topic or problem. Two students answered that they use both: at first they search by assignment topic, if they don’t find the solution they would go with the search by error or bug when solving their assignments.

(9) How long will you take to find the solution website?

Nearly 35 out of 41 students answered that they would take less than 5 minutes to find the solution website. Few students reported that they also take less than 5 minutes but if the bug is complex it would take more than 10 minutes for finding the solution website.
(10) How many web pages did you visit before finding the solution?

More than half of students answered that they would visit 3 to 5 websites before finding the solution website and few students reported that they would visit one or two websites and they said that they would select the website from the first three results from the search results.

(11) How long did it take for you to define the search parameters for the problem?

Almost all the students answered that they would take 1 to 2 minutes to define the search parameters for the problem.
CHAPTER 5

CONCLUSION & FUTURE WORK

Our results show that undergraduate students commonly use the web to search for help with programming bugs. In the survey, we found that search engines are the second most popular source of help with programming bugs. We learned that 82% of our students use the web for help with at least half of their programming assignments. Usage of resources and different websites varies by the class year. As the class year progresses, the use of class notes decreases and the use of search engines increases. The mean time taken to define search parameters was 7.5 minutes and the mean number of web pages that they visited for solutions was 3.6. The most helpful websites include those with tutorials, examples of code, and forums that discuss similar questions to that of their assignment.

In the focus group, students mainly use a few common websites to look for solutions to their programming bugs. As class year level progresses, the search by the error is given more preference compared to search by topic. Furthermore, the use of forum sites increases in junior and senior years. From the discussions of focus group meetings, most of the students research the assignment before they start the programming and google to find the example code similar to their programming assignment. By looking at example code students are able to see how certain functions work together and how each part of the code relates, which allows them to code a similar assignment. Students have gotten easier in fixing the bugs over the course of the semester and few students reported that concepts such as multi-dimensional arrays, pointers, recursion and memory management were challenging for them. We also had a discussion on the resources used and behavior of searching. Students searches focused in two main areas: searching for problem specific information and searching for a broad understanding of a topic. For example, reverse string in C using recursion is classified as problem specific, because the student assignment was to reverse the strings from a file using recursion. How to use pointers is an example of broad understanding, because the student was having difficulty assigning pointers to arrays. These searches could be due to many
different reasons; without polling the students for more information, we cannot be conclusive. However, since many of our students do not use class notes when programming. Possibly due to a lack of notes, students may not have fully understood the programming assignment. Many students in our focus group admitted to not taking notes in class. When asked why they did not take notes, participants gave two common responses: You can find almost all that information online really easily and Most the time the teachers are just using the materials supplied with the textbook, they are basically reading to us from the book.

Students also searched by errors (compiler and runtime) and by functions, but these categories were not as prevalent. One student gave an example of searching by an error ”it can be difficult to find help for errors such as a segmentation fault because many errors are specific to each program”. To solve these types of errors often requires looking at the entire program, not just one segment, making it difficult to search the web and they may already realized searching for help is futile.

Our future work will study search patterns of professional programmers and factors that influence their search efficiency. We will use same data collection process and will classify search terms by the number of years of professional experience. Our primary goal is to better understand how novice and professional programmers use the web to effectively search the web for help with programming bugs, including search terms and the types of sites that they gravitate toward to find helpful information. We will identify best practices to help novice and professional programmers to search the web for help with programming bugs.
APPENDIX

SURVEY QUESTIONS
Survey

* Required

Full Name *

Your answer

School Email Address *

Your answer

Current Course *

Your answer

Gender *

- Male
- Female
Class Year *
  ○ Freshman
  ○ Sophomore
  ○ Junior
  ○ Senior

Which resources do you typically use when working on a programming assignment: *
  □ Text Book
  □ Class Notes
  □ Search Engine
  □ A specific Website
  □ Professor's slides and notes
  □ Other: ____________________
How often do you search the web for help solving assignment problems? *
- All Assignments
- Most Assignments
- Half of Assignments
- Less than half of Assignments
- None

How helpful are forums with the questions similar to your assignment? *
For example, stackoverflow.com

<table>
<thead>
<tr>
<th>Not Helpful</th>
<th>Helpful</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

How helpful are websites with examples of the code? *
For example, php.net

<table>
<thead>
<tr>
<th>Not Helpful</th>
<th>Helpful</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
How helpful are websites with tutorials? *
For example, w3schools.com

1  2  3  4  5  
Not Helpful  ☐  ☐  ☐  ☐  ☐  Helpful

How often did you visit focus group website? *
  i.e. mayo.cse.unt.edu/focusgroup

☐ For all Assignments
☐ Half of my assignments
☐ Less than half of my assignments
☐ Never Visit

When using a search engine for solving a problem, how did you decide what to search for? *
What did you search by?

☐ Search by assignment
☐ Search by error
☐ Other :
How long did it take for you to find the solution website? *
Give an approximate time

- less than 2 minutes
- 2 to 5 minutes
- 5 to 10 minutes
- 10 to 15 minutes
- more than 15 minutes
- Other: ___________________________

How many webpages did you visit before finding the solution? *
Give an approximate number

Your answer ___________________________

How long did it take for you to define search parameters for the problem? *
Give an approximate time in minutes to understand the problem and define appropriate search terms for it

Your answer ___________________________
Expected grade of the current course *

- A
- B
- C
- D
- F
- Other: ____________________

SUBMIT
BIBLIOGRAPHY


List of common bugs and programming practices to avoid them, [Accessed: Dec. 04, 2015].

Focus groups: A practical guide for applied research, Sage publications, 2014.


Jmp for basic univariate and multivariate statistics: Methods for researchers and social scientists, SAS Institute, 2013.


What makes a


