

# Improved understanding of student interests in Computer Science through Page Tracking Analysis

Gian Calica  
Collaborative Software Development Laboratory  
Information and Computer Sciences  
University of Hawaii

May 13, 2020

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Research Design</b>	<b>2</b>
<b>3</b>	<b>Implementation</b>	<b>3</b>
3.1	PageInterestCollection . . . . .	3
3.2	PageInterestsDailySnapshotCollection . . . . .	4
<b>4</b>	<b>Evaluation</b>	<b>4</b>

# 1 Introduction

The field of computer science has rapidly expanded in such a short amount of time over the last few decades and is continuing to expand at a growing rate. As it continues to expand, so must also the technologies that support it. Every month there are new technologies being released, whether it's a programming language, a framework, or a new library encapsulating domain knowledge or an algorithmic technique for a language, all of which claim to have an advantage over a previous technology.

As the field continues to expand, it is important for the Information and Computer Sciences (ICS) Department in University of Hawaii at Manoa (UHM) to be able to respond and keep up with these fast-changing developments. RadGrad is one way the ICS Department can tell which courses to include in its curriculum. RadGrad can track how many students plan to take certain courses and in which semesters, and this statistic can be used by the students as a way to tell the ICS Department which type of courses they are interested in.

RadGrad also has a way to track how many students are interested in a particular area of Computer Science through interests. However, many of these interests have no course available in the ICS Department to teach students about those interests. This proposal proposes a mechanism called "Page Tracking Analysis" which can discover existing or emergent demand for topic areas by computer science students.

## 2 Research Design

In the context of the RadGrad system, we will define a student's interest in a particular topic with two definitions: 1) if the student has visited the corresponding page for that topic in RadGrad and 2) if the student has added that topic to their favorites.

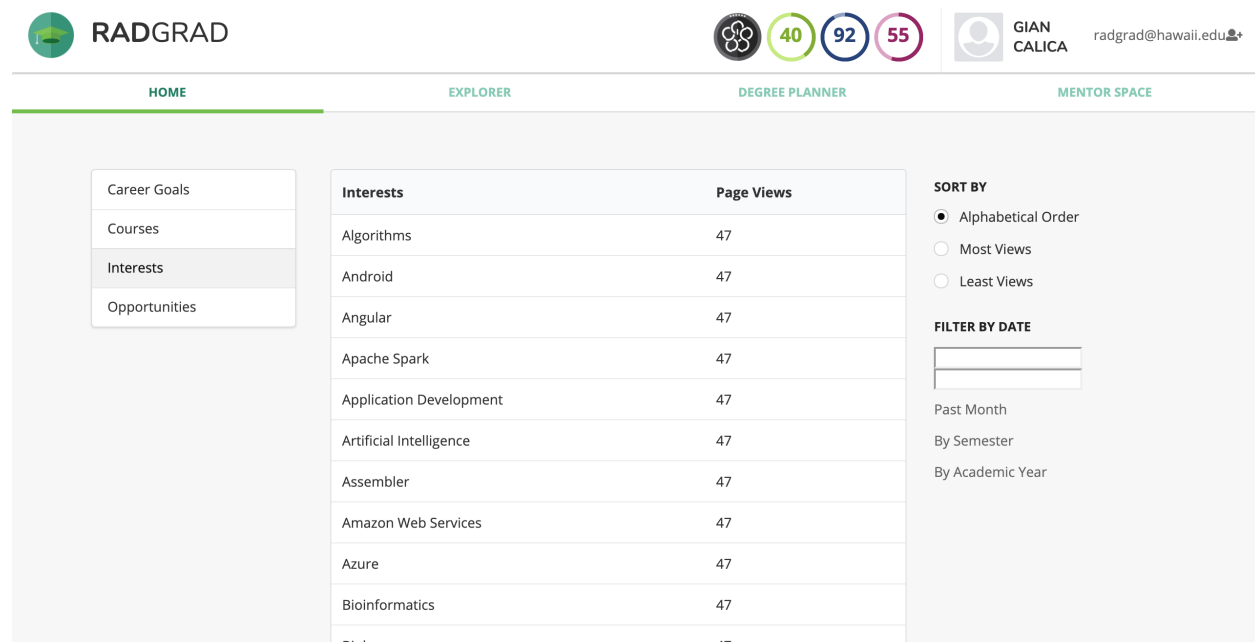


Figure 1: Page Tracking Analysis Scoreboard

With those definitions in mind, this proposal will implement a capability called the Page Tracking Analysis. There will be two dedicated pages under this capability. The first page is implemented like a scoreboard because the RadGrad2 system already has similar pages called scoreboards that function similarly but for

different data. In this dedicated page, there will be a section for each topic category: Career Goals, Courses, Interests, and Opportunities. This is shown in the left menu of Figure 1. Each section will list the names for that area and beside it would list the page visits for each corresponding name. This page has the ability to sort each sections based on alphabetical order, most views, and least views. Users will also be able to filter the data that belongs within a specific date range, the past month, by semester, or by academic year. This is shown in the right side of Figure 1.

The second dedicated page will have a functionality to be able to analyze the trends of page views for each topic category more in-depth. This page will allow a user to select multiple entities from a single topic category and will show graphical charts and plots of the page views only for those selected entities. By default, it will sum all the page views for those entities since the beginning of the database snapshot. Then users will have the ability to filter the data of page views within a specific date range, within the past month, by semester, or by academic year.

With these statistics being tracked, we can now analyze and see if there is a demand for a particular topic that the faculty does not currently know through these page visits. The first page gives us an overview of the page views for all the entities for a particular topic category. The second page allows us to analyze if there is a trend of page views for specific entities within a particular topic category. This gives us a system that can provide insights about student interests that the faculty might not be aware of. One question that could be explored would be, “Why did the interest Data Science or career goal Data Scientist suddenly spike up in page visits in this certain date range?” Then when we investigate this cause, we might find that a course related to Data Science was actually offered in that semester of that date range. From this research investigation, we might conclude that our RadGrad team should promote and look for opportunities related to Data Science in that particular date range.

Just like the purpose of collecting “Students Participating” data for a Course or Opportunity, the ICS faculty and department can also use this dataset to see which areas they should emphasize and focus on teaching the students.

As students also explore the field of computer science beyond what their classes teach, analyzing which fields in computer science they are most interested in allows us to also have a glimpse of what the field of computer science is heading towards. This can better prepare the ICS Department to adjust the courses and curriculum to meet with the standards of the constantly expanding computer science industry.

### **3 Implementation**

Although we define a student visiting the page of a particular topic as the student being interested in that topic, we do not want to count each page visit counting towards their interest. There are certain circumstances such as accidentally clicking the page even though they were not interested in that topic that might skew the data. Some factors to consider are time between a student visiting the page for a topic versus a student leaving the page or time between a student visiting the page for a topic versus a student adding that topic to their favorites. See the implementation details of `PageInterestCollection` for a detailed definition of a student interest view.

To implement the two dedicated pages of Page Tracking Analysis, we will be implementing two collections: `PageInterestCollection` and `PageInterestsDailySnapshotCollection`.

#### **3.1 PageInterestCollection**

A document object of this collection represents a student’s one interest view for a specific topic category page. The different topic categories that a page could be classified under are Career Goal, Course, Interest, and Opportunity.

A student interest view is defined as follows:

1. When they Favorite that page and don't unfavorite it before leaving the page
2. They play the Teaser video at that page or open it as an external link
3. They click an external link in the description of that page
4. They spend a minimum of 5 seconds viewing that page

The schema for a document object of this collection is defined as follows:

- username (string) - the username that this document is associated with
- category (string) - the category that this document represents (career-goals, courses, interests, and opportunities)
- name (string) - the name that this document represents in the form of a slug (data-scientist, google-summer-of-code, etc...)
- timestamp (date) - the timestamp of when this document was created

### 3.2 PageInterestsDailySnapshotCollection

This collection is automatically populated every day using a cron job. A document object of this collection represents a snapshot of the page views for all topic categories for that day.

The schema for a document object of this collection is defined as follows:

- careerGoals ([object])
- courses ([object])
- interests ([object])
- opportunities ([object])
- timestamp (date) - the timestamp of when this document was created

In the defined schema, the careerGoals, courses, interests, and opportunities are all arrays of objects. Each object represents a specific area of that topic category. The schema of the objects is defined as follows:

- name (string) - the name that this object represents in the form of a slug (data-scientist, google-summer-of-code, etc...)
- views (number) - the number of aggregated student interest views for this object

## 4 Evaluation

It would be more appropriate if this new feature is evaluated after it has been deployed to RadGrad2 for at least a two semesters since it will be implemented in that version of RadGrad. Nonetheless, this feature can still be evaluated to some degree if we manually feed the data of page views from RadGrad1 to this implementation in RadGrad2. However, this evaluation might not accurately reveal if there are unknown threads when it comes to student interests. As discussed in the Implementation section, our implementation of page views will differ in RadGrad2 from RadGrad1. We won't necessarily be counting every single page

view as a student interest due to the fact that misclicks can happen or a student may visit a single page multiple times within a session.

When we do evaluate this feature with a semester or two of RadGrad2 data, we can evaluate trends in two ways. The first way corresponds to the two dedicated pages of the Page Tracking Analysis. The first page can be used to find general trend of page views for all areas in each topic category and the second page can be used to target specific areas of each topic category to view their trend of page views.

The second way corresponds to evaluating the trend of page views for students as they go through Freshman year through their Senior year. One problem that lies with this is classifying who is Freshman and who is Sophomore. This is hard to classify because everybody goes through their degree program differently. There are some who have more balanced semesters with ICS specific courses along with non-ICS courses and there are others who end up taking three or more ICS specific courses in a single semester. Therefore, I have classified these school years into two parts: the first part corresponds to students who are less than halfway through their program and the second part corresponds to students who are halfway through and nearing the end of their program. This is because the first part would just be students taking the required ICS courses for their degree program. The trend of student interests at this time period may not necessarily change from what they were already interested in at the time of their RadGrad onboarding. We compare this first part to the second part as this is when students are most likely finished with all, if not most, of their required courses and are finally taking ICS electives that they are free to choose. Comparing their page views at this period of time to their page views of the first part can give us an overview of how their student interests changed over time. Did the students' change of interest match the type of courses they have taken? Did it even change at all from the first part? Or did it change, but not necessarily matching with the type of courses they have taken? This second way seems to be another feature to implement that specifically looks at the period of time that a student is currently in (whether they are the first or second part), queries their page views from those parts, and finally gives an output to display these queries.