Improving Stack Overflow Tag Prediction Using Eye Tracking

Jenna Wise (Computer Science and Mathematics) jlwise@student.ysu.edu – student

Ali Morris (Computer Science) almorris01@student.ysu.edu – student

Alyssa Pawluk (Computer Science) alpawluk@student.ysu.edu – student

Alina Lazar (Computer Science) alazar@ysu.edu – faculty advisor

Bonita Sharif (Computer Science) bsharif@ysu.edu – faculty advisor

Youngstown State University

Project Description

The Stack Overflow website is an essential and a growing resource among the community of coders all over the world. Software developers use it to post questions and answers related to programming and computer science problems they need to solve. Questions such as seeking input on some efficient and time-saving methods of coding a particular program, getting help on solving various bottlenecks in coding are commonly seen. Over the years the website has slowly evolved into a large free repository of knowledge. Currently, Stack Overflow includes 12 million questions, 19 million answers and 47 million comments all available to download in a data dump of size 70GB. The data is made publically under the Creative Commons cc-by-sa 3.0 license. Given the availability and size of the dataset many researchers from fields such as information retrieval, text mining and machine learning have been working with this dataset.

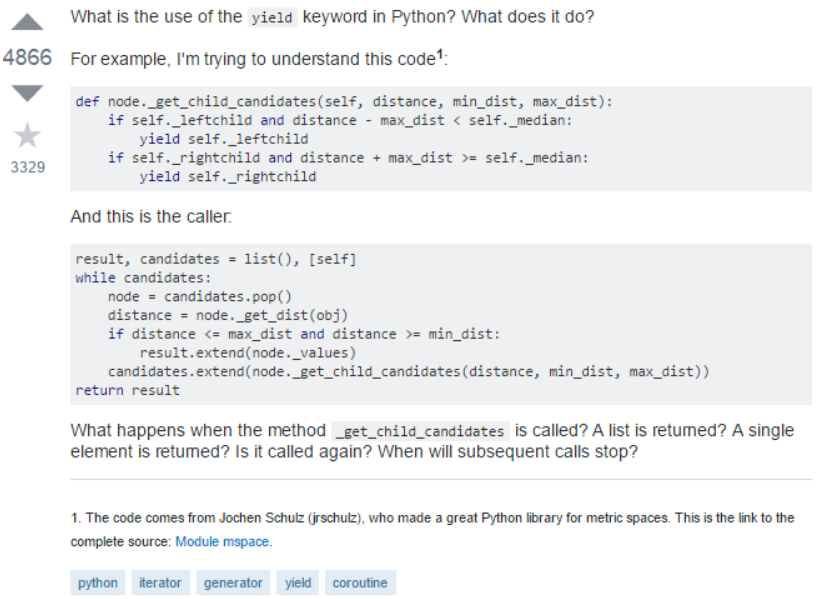

Figure 1: A Stack Overflow question showing five tags at the bottom.

When users submit questions on Stack Overflow they need to submit at least one and up to five tags in addition to their question (see Figure 1). These tags attached to each question broadly identify the programming language talked about, the problem type in discussion and maybe some other fine grained categories the question belongs to. The tags associated to each question help with information retrieval or user queries. For example, it may be very useful when users try to identify duplicate questions or related questions to a particular problem. Several approaches [1]–

[4] for automatically generating tags from short questions or text have been recently developed. The Stack Overflow dataset is perfect for this problem as it provides the ground truth (as the author of the question adds these tags). The problem with this approach is that this multi-label, multi-class classification solution does not provide very good accuracy. The best accuracy reported is 68.47% [1]. Our goal for this project is to use eye-tracking studies to improve the prediction accuracy for Stack Overflow question tags.

In the last couple of years there have been many studies [5]–[7] designed to understand where humans focus while looking at a static image to predict gaze allocation. Several eye tracking datasets were constructed by the above studies and are freely available for research and replication. Given any type of text, image or video the eye-tracker will show where the users look and for how long. The eye-tracking device can collect information such as eye fixations and saccades. Scanpaths are a set of fixations in order. See Figure 2 for an example of fixations, saccades, and scanpaths. The circles indicate fixations, the radius of the circle represents the duration of the fixation, and the lines between fixations are saccades. The numbers shown in the circles are the fixation number. A sequence of fixations and saccades is a scanpath.

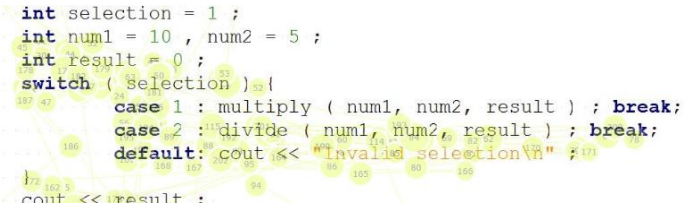

Figure 2: Scan path of an expert reading a C++ program.

Eye tracking has only recently been used in the field of software engineering to study how developers work [8], [9], [10]. Although it is new to software engineering, eye tracking has been used to study cognitive processes involved in the comprehension of visual material for over thirty years [11]. The main premise is that mental processes are triggered when people focus on certain things [12], [13]. This focus is referred to as visual attention.

In our project, the tasks are related to software engineering. In particular, the task will be to read Stack Overflow questions including code to identify up to five tags or keywords. Eye trackers generate a huge amount of data per session per user. An eye tracker running at 60 Hz will generate 60 samples per second. This can easily add up to a lot of data during a session that typically lasts anywhere between 20 minutes to 60 minutes. This data is very rich in thought processes, strategies, techniques that a user utilizes while they are working on a task. Fixations and saccades are the two most widely used eye movement measures.

Our main goal this year is to use eye fixations to improve the accuracy of automatic tag prediction. The main benefits of this prediction are 1) to automatically identify the keywords in short text and source code snippets and 2) to determine what people focus on when they read Stack Overflow questions.

Research Questions and Hypotheses

For this project we will focus on the following research questions.

- To what degree do programmers focus on the keywords that tag extraction techniques extract?
- To what degree do the top-n keywords from our approach and the standard approach match the keywords chosen by the author of the Stack Overflow question?

- What are the best machine learning algorithms (informed by eye gaze) that can be successfully used to make such predictions?

Methods

The problem of tag prediction can be considered a classification problem. A machine learning algorithm will be trained using a high dimensional feature vector extracted from the question's text, the included code snippet and the eye-tracking data to predict a tag value. We will use the eye tracking data to determine weights for keywords that are read within the question. The samples in the training set will be questions from an existing Stack Overflow dataset [14] that was generated from the Stack Overflow data dump. The set of features will be generated using natural language processing methods such as *tf-idf* but also from the code snippets. Eye movement and gaze fixations will be used to generate weights for the main features [15]. We will also include other characteristics from the eye tracking data such as regressions, text and source code element visited, and sequence of visits in time.

From source code, two types of features can be extracted [16]. First, there are the low-level lexical features such as "word length", "contains digit", "word unpredictability", "contains uppercase", "is all uppercase". Second, syntactic features include abstract syntax trees, variables, pointers, instructions, basic blocks, and procedures. Especially important are the abstract syntax trees (AST) that can provided additional features [17] such as maximum depth of an AST node, term frequency of possible AST node types, term frequency inverse document frequency of possible AST node types, average depth of possible AST node types, and term frequency of the programming language keywords.

To train and test the model, the dataset will be divided into a training set and a smaller testing set. From each question we will randomly choose five positive tag labels and five negative tag labels. To get good results, the features need to be normalized to have zero mean and variance equal to 1. We will evaluate the following classification algorithms: Linear Support Vector Machines (SVM), Naïve Bayes, and Random Forests.

We will gather eye tracking data for question readability tasks. The study will be conducted in our eye tracking lab at the Computer Science and Information Systems department. The eye tracker generates additional data such as pupil dilation, saccade durations, dwell time, fixation duration, to name a few. We will obtain IRB approval from YSU's IRB before we conduct the study. The students will also do the online training on how to deal with human subjects.

Performance Metrics

Our models will be evaluated using classification accuracy, precision, recall, F1-measure and receiver operating characteristic (ROC) curves. The ROC curves will be obtained computing the true positive rates versus the false positive rate. We plan to test several models using different sets of features. All these measures will be used in conjunction to obtain a more comprehensive picture of the classifier's capabilities.

Impact on the Goal of CREU

The goal of the CREU project is to encourage females and minorities to pursue graduate work and study in the field of Computer Science. This project will provide realistic research experience for our female undergraduates, by active involvement in the planning, execution and interpretation of scientific research. Well-developed research projects can significantly enrich the educational experience for undergraduate students. Working on this research project, the students will be able to enhance their computer and programming skills, apply those skills to investigate

scientific problems, learn how to formulate questions and problems and to participate in the discovery of new knowledge. A good research experience can foster an enthusiasm for lifelong learning and a desire to continue education beyond the bachelor's degree. The students will be exposed to both sides of the scientific investigation: hypothesis testing and development of theoretical explanations of observations. No science education is complete without research related activities, technical writing and oral presentations.

Jenna Wise is a senior pursuing a double major in Mathematics and Computer Science with a 4.0 GPA. She is an honors student and is very interested in pursuing a research project in Computer Science. Her mathematical background will be ideal for this type of research project that involves many mathematical models for prediction. This is a project that can potentially be a foundation for her senior thesis. The experiential nature of this project will add to the science of teaching and learning in software development.

Ali Morris is a senior student pursuing a degree in Computer Science. As an undergraduate female with intentions of pursuing graduate work in Computer Science, this project will continue to give her a practice with practical applications for her further research. She is very interested in experimental design and processes to conduct studies and the human aspect of learning about how developers read code. She is also the president of the ACM-W student chapter at YSU.

Alyssa Pawluk is a junior student pursuing a degree in Computer Science. She started as a journalism major but changed to Computer Science after her sophomore year. She is very excited to join our team to learn about how a research project is conducted. As an undergraduate female with intentions of pursuing graduate work in Computer Science, this project will continue to give her confidence to keep pursuing a degree in Computer Science.

The girls we have chosen for this year's CREU have skills that work synergistically to achieve the goals of this project. They are all part of the ACM-W student chapter at YSU and strive to build awareness by being change leaders in STEM fields. The students intend to present this project at next year's QUEST at Youngstown State University (a program highlighting student research). The students will likely also submit findings at regional and/or national scientific meetings (Mining Software Repositories and International Conference on Program Comprehension) within their fields. They will also participate in the YSU STEM Showcase that is held every year and open to all nearby schools. This will help get more students interested in Computer Science projects. Results generated by this project will be included in one or more future manuscripts, with participating students afforded full opportunity to share the responsibilities and rewards of authorship.

Student Activities and Responsibilities

Specific tasks for the students will include: literature search and review, review write-ups, reading, presenting, and discussing research articles, designing and implementing the experimental studies, data processing, data analysis and interpretation, running data mining algorithms, summarizing and preparing results for presentations and publications, YSU QUEST 2017 participation and writing the final report. The students will also be mentored to prepare a conference paper that will be sent to the International Conference on Program Comprehension (ICPC) or the International Conference on Mining Software Repositories (MSR).

The primary responsibility of the students is to participate in all phases of the project: proposal, development, experiments, and dissemination. The students will be required to do weekly independent work and to schedule team meetings with the faculty advisors. The faculty advisors will meet with the students every week. Email and a central repository will be used for

questions, announcements, and document interchange. A blog will also be setup during the first week with clear shared responsibility for updates.

Faculty Activities and Responsibilities

Dr. Alina Lazar will work to actively mentor the student and continuously supervise their progress during the one year period. She together with her colleague (Dr. Sharif) will meet with the students on regular basis to guide their activities and answer their questions related to the project. Dr. Lazar has extensive experience in data mining, machine learning and artificial intelligence and she has written several papers related to that field. Dr. Lazar will help the students with the data analysis, and with the final report and also with the preparation of a conference paper. She will lead and oversee the data mining algorithm aspect of the project.

Dr. Sharif has conducted over 10 empirical studies in the field of software engineering using various means of data collection such as questionnaires and mostly eye tracking hardware and software. She will be responsible for developing concrete research hypotheses for each of the research questions mentioned above. She will mentor the students to design the experiments that address the research questions. This will involve experimental design including selection of variables, selection of tasks, and appropriate methods before the study is implemented. The students will be part of the design and variable selection. After the data is collected, she will also be responsible for guiding the student to choose appropriate methods of analyzing the data and help with publication write-up.

The overall guidance and mentoring will not refer only to this project but also provide insights about how to apply and how to succeed in graduate school, about being a female scientist and what the options are after graduate school.

Project Timeline

Please refer to Table 1 for the project timeline.

Table 1. Project Timeline

Task Name	2016				2017							
	S	O	N	D	J	F	M	A	M	J	J	A
Literature review on tag prediction for Stack Overflow documents.	■	■										
Develop and conduct an experimental study on Stack Overflow questions using the eye tracking lab. Get IRB approval.	■	■										
Format the data for prediction. Generate and rank classification features.			■	■								
Develop a new weighting scheme for the existing features using eye tracking data. Run data mining algorithms on the generated features in the data set.				■	■	■	■	■				
Statistically analyze the classification results					■	■	■	■				
Dissemination of results through papers and communications at specific conferences.								■	■	■	■	■

Budget

For the proposed project we are requesting $3000 for each student. The additional $1500 will be used to support the student travel to TAPIA. While working on the project the student will

be encouraged to apply for the Undergraduate Student Research Grant Award sponsored by the Youngstown State University and other scholarships. We plan on applying for the summer extension next February. A summary of the budget is given below.

Items	
Academic year for Jenna	$3000
Academic year for Alyssa	$3000
Academic year for Ali	$3000
Travel allowance	$1500
Total requested:	**$10500**

Role of the CREU project within the larger scope of this research

The Department of Computer Science and Information Systems at Youngstown State University has always valued undergraduate research and provided resources for projects. The Chair, the Dean, and the Provost all encourage research activity and provide faculty with necessary time to conduct such research with students. The faculty advisors created a new usability lab in 2012 as part of an internal grant that has access to state-of-the-art machines and eye trackers that students can use for research projects. Another lab that our students will use is the Software Engineering Research and Empirical Studies Lab that was established by Dr. Sharif. The students will use both labs for conducting the eye tracking studies, data analysis and running the experiments.

The results of this project will directly impact other projects that are currently underway at the Software Engineering and Empirical Studies Lab. The girls will occasionally collaborate with other lab members to make sure the work is extensible to other related projects. This will help advance not just this project but also other projects that are conducted in the lab.

Prior results of CREU projects

Bonita Sharif and Alina Lazar worked with students Jenna Wise and Jessica Whitely during the CREU 2015-2016 project titled: "*Predicting Areas of Interest in Code Reading*". Jenna presented the results of this project at YSU's QUEST in April 2016. We also got a poster accepted at ACM TAPIA Conference 2016. Jenna will be presenting this work in September 2016. We are working on a submission to the SWAN 2016 workshop on Software Analytics that is due on July 1st 2016.

Bonita Sharif and Alina Lazar also worked with students Jenna Wise and Jessica Whitely during the CREU 2014-2015 project titled: "*Mining Eye-tracking Data to Determine Developer Expertise and Task Difficulty in Software Development*". Jenna and Jessica presented the project "*An Eye-tracking Experiment Studying Problem Solving Behavior*" at the Ohio Celebration of Women in Computing (OCWIC 2015) conference and a talk at YSU's QUEST 2015 - a forum for student scholarship. Jessica also presented her senior project *"Towards Understanding Student Problem Solving Behavior in Algorithms via Eye Tracking"* in the department.

Alina Lazar and Bonita Sharif advised Sarah Ritchey during the CREU 2013-2014 project titled: "*Classification Algorithms for Detecting Duplicate Bug Reports in Large Open Repositories*". We got two papers published at the International Conference on Mining Software Repositories in May 2014. Sarah also presented this work at YSU's QUEST 2014, a forum for student scholarship, Ohio MAA Spring Sectional Mathematics Meeting, and the Pi Mu Epsilon meeting. We also got a poster accepted at the Grace Hopper Celebration 2014. This was presented

by Dr. Sharif as Sarah had other engagements during that time. Sarah is now doing her Ph.D. at Duke University.

Bonita Sharif and Alina Lazar guided student Rachel Turner during the CREU 2012 -2013 project titled: "*An Empirical Investigation on Code Debugging and Understanding: An Eye-tracking Perspective*". Rachel presented the poster "*C ++ vs. Python: An Eye-tracking Assessment*" at the Ohio Celebration of Women in Computing (OCWIC 2013) conference and a talk at the YSU's QUEST 2013 A Forum for Student Scholarship. We also published a paper at the Eye Tracking Research and Applications Conference (ETRA) in March of 2014.

Alina Lazar advised two prior CREU projects in 2004-2005 and 2006-2007, and one MROW project in 2007-2008. Darcy Davis the participant student in the 2004-2005 project just finished her PhD in Computer Science and Engineering at Notre Dame University. Louise Popio who participated as a student during 2006-2007, received her Master's in Information Sciences and Technology from Pennsylvania State University in 2010. Irena Lanc participated in the 2007-2008 MROW project and received her PhD in Computer Science and Engineering from The University of Notre Dame in 2014. Erin Pfeil the other student that did the 2007-2008 MROW project, is working on her PhD in Ecology at the University of Pittsburgh.

Student Transcripts

The student transcripts follow the references.

References

[1] A. K. Saha, R. K. Saha, and K. A. Schneider, "A discriminative model approach for suggesting tags automatically for stack overflow questions," in *Proceedings of the 10th Working Conference on Mining Software Repositories*, 2013, pp. 73–76.

[2] M. Lipczak and E. Milios, "Learning in efficient tag recommendation," in *Proceedings of the fourth ACM conference on Recommender systems*, 2010, pp. 167–174.

[3] C. Stanley and M. D. Byrne, "Predicting tags for stackoverflow posts," in *Proceedings of ICCM*, 2013, vol. 2013.

[4] S. Beyer and M. Pinzger, "Synonym suggestion for tags on stack overflow," in *Proceedings of the 2015 IEEE 23rd International Conference on Program Comprehension*, 2015, pp. 94–103.

[5] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *Computer Vision, 2009 IEEE 12th international conference on*, 2009, pp. 2106–2113.

[6] T. Saikh, S. Bangalore, M. Carl, and S. Bandyopadhyay, "Predicting Source Gaze Fixation Duration: A Machine Learning Approach."

[7] Q. Zhao and C. Koch, "Learning saliency-based visual attention: a review," *Signal Process.*, vol. 93, no. 6, pp. 1401–1407, 2013.

[8] B. Walters, T. Shaffer, B. Sharif, and H. Kagdi, "Capturing software traceability links from developers' eye gazes," in *Proceedings of the 22nd International Conference on Program Comprehension*, 2014, pp. 201–204.

[9] B. Walters, M. Falcone, A. Shibble, and B. Sharif, "Towards an eye-tracking enabled IDE for software traceability tasks," in *Proceedings of the 2013 International Workshop on Traceability in Emerging Forms of Software Engineering*, 2013, pp. 51–54.

[10] R. Turner, M. Falcone, B. Sharif, and A. Lazar, “An eye-tracking study assessing the comprehension of c++ and Python source code,” in *Proceedings of the Symposium on Eye Tracking Research and Applications*, Safety Harbor, Florida, 2014, pp. 231–234.
[11] K. Rayner, K. H. Chace, T. J. Slattery, and J. Ashby, “Eye Movements as Reflections of Comprehension Processes in Reading,” *Sci. Stud. Read.*, vol. 10, no. 3, pp. 241–255, Jul. 2006.
[12] A. T. Duchowski, E. Medlin, A. Gramopadhye, B. Melloy, and S. Nair, “Binocular Eye Tracking in VR for Visual Inspection Training,” New York, NY, USA, 2001, pp. 1–8.
[13] M. A. Just and P. A. Carpenter, “A theory of reading: From eye fixations to comprehension.,” *Psychol. Rev.*, vol. 87, no. 4, pp. 329–354, 1980.
[14] “Facebook Recruiting III - Keyword Extraction,” *Facebook Recruiting III - Keyword Extraction*. [Online]. Available: https://www.kaggle.com/c/facebook-recruiting-iii-keyword-extraction.
[15] P. Rodeghero, C. McMillan, P. W. McBurney, N. Bosch, and S. D’Mello, “Improving automated source code summarization via an eye-tracking study of programmers,” in *Proceedings of the 36th International Conference on Software Engineering*, 2014, pp. 390–401.
[16] P. Martínez-Gómez, T. Hara, C. Chen, K. Tomita, Y. Kano, and A. Aizawa, “Synthesizing image representations of linguistic and topological features for predicting areas of attention,” in *Proceedings of the 12th Pacific Rim international conference on Trends in Artificial Intelligence*, 2012, pp. 312–323.
[17] A. Caliskan-Islam, R. Harang, A. Liu, A. Narayanan, C. Voss, F. Yamaguchi, and R. Greenstadt, “De-anonymizing Programmers via Code Stylometry.”