MAY 1620

Cody Hanika (Team Leader)
Deeksha Juneja (Communication Leader)
Alex Dana (Key Concept Holder, Web Master)

A Sociotechnical Solution to Large-Scale Formal Specification Mining

Formal Specifications

❏   Formal specifications allow a programmer to describe the behavior of a piece of code in a programmatic and verifiable manner.
❏   These consist of a precondition and postcondition pair that specify what must be true about the program state before and after the code executes
❏   Allow programmers to more easily write correct code
❏   Research has been conducted to create formal specification languages (JML) that can be applied to many programming languages

Informal vs. Formal Specifications

```
/*
 * This function returns a divided by b. An exception will be thrown if
 * b is 0.
 */
public static int divide1(int a, int b) { return a / b; }

/*@    public normal_behavior
  @       requires b != 0;
  @       ensures \result == a / b;
  @ also
  @    public exceptional_behavior
  @       requires b == 0;
  @       signals_only ArithmeticException
  */
public static int divide2(int a, int b) { return a / b; }
```

The Problem

❏ Formal specifications are not widely available to programmers

❏ Many programmers do not know how to use formal specifications

❏ It can be difficult to write new formal specifications without a lot of experience using them

❏ It is difficult to automatically generate formal specifications from informal specifications (code comments) or by analyzing pieces of code

Our Approach / Functional Requirements

- ❏ A socio-technical solution that will leverage the success of community collaboration platforms such as StackOverflow
- ❏ Create specification for trivial methods automatically
- ❏ Post questions about specifications on StackOverflow to receive answers
- ❏ Parse and analyze these answers and store the solutions for others to use

Potential Risk

❏ Because there are so few programmers with experience in formal specifications, it is difficult to get (correct) responses to questions

❏ Using an automated method to post to StackOverflow is a gray area in the terms and conditions for the site

❏ Our application is planned to be merged with May1639, which will provide a platform like StackOverflow

Technology Platforms Used

- ❏ OpenJPA was used for persistence of database objects
- ❏ Stanford NLP is being used for answer analysis on StackOverflow posts
- ❏ OpenJML for JML analysis
- ❏ Protocol Buffers for scaleable storage
- ❏ The Boa platform is being used to make our specifications widely available

System Overview
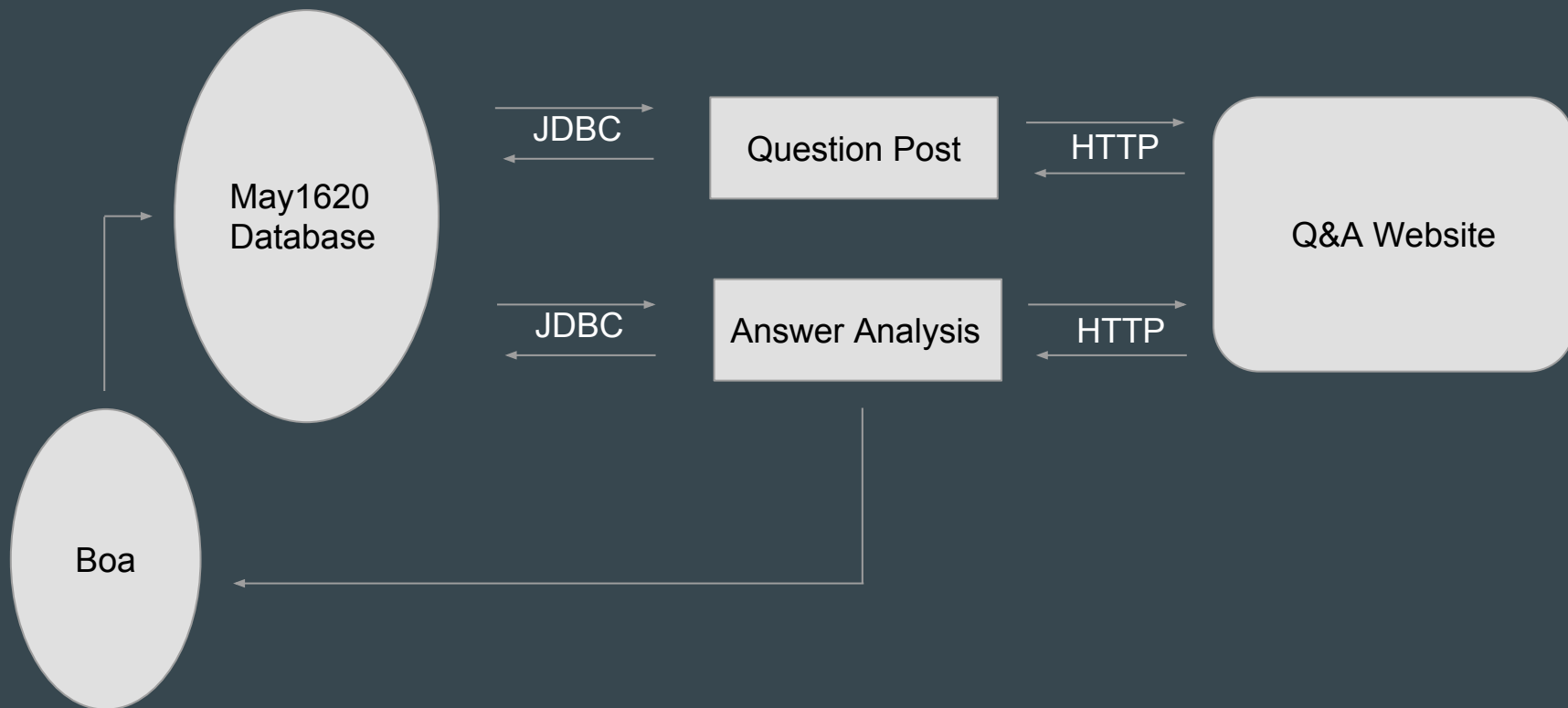
Posting

❏ The input to our system will be from a Boa datafile
❏ We will automatically generate the specification for trivial methods
❏ We will use the Boa data to automatically post a question on underspecified methods
  ❏ We will ask for an answer that provides the JML specification
❏ The posting is done via http requests to StackOverflow which are authorized using OAuth2.0
❏ Once a question is posted we store the question id in our database to return and analyze later
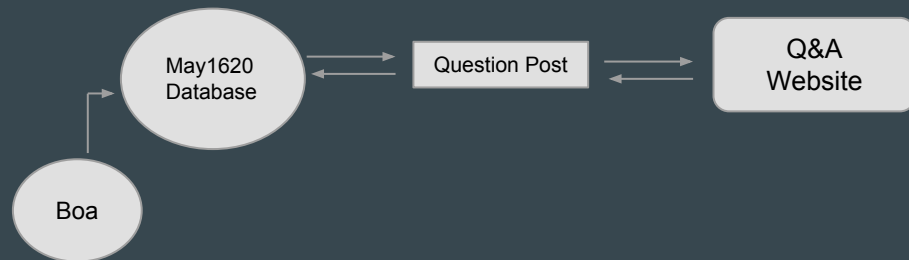
Analysis

❑ We will use our database to obtain a previously posted question for analysis

❑ The question will be retrieved using http requests

❑ The Stanford NLP library will then be used to analyze the answers and comments to the questions in order to extract important pieces of information

❑ The JML provided in the StackOverflow answers will be parsed and stored back into the Boa system for later use

May 1620 Database → Answer Analysis → Q&A Website

Answer Analysis → Boa
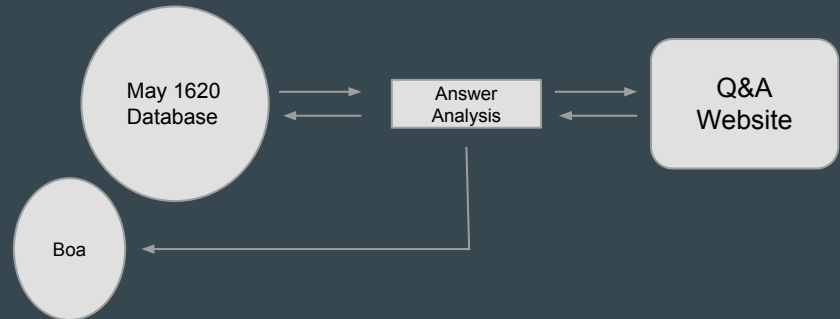
Current Project Status

- ❏ Currently the Post area of the application is complete
- ❏ Not much progress has been made with the Analysis but this is being looked into now
- ❏ Database issues have caused some setbacks but we are working on getting a dedicated VM for this purpose
- ❏ Need to still develop a method for parsing the JML and storing the data back into Boa

Social Challenges

❏ Our initial questions posted on StackOverflow have not received much response
❏ We believe that this is due to a lack of understanding of JML specifications within the community
❏ We are looking into different methods of phrasing these questions in order to maximize useful responses

   ❏ Asking straightforward questions without examples seems to be a quick way to get no response from StackOverflow users

   ❏ We have created different users to try and see if account name, profile picture, reputation, etc. influences the rate of answering

   ❏ We are creating new questions with more detailed examples of what we want from an answer and it has shown improved results

Example Questions - No Responses

What is the proper JML for the java Scanner.hasNextInt(int) method?

▲

0

▼

I am using the java.util.Scanner class. I would like to have the JML for the nextInt(int) function. I am mainly concerned with the requires, ensures, assignable, and the signals_only clauses (although more is always better). So what is the JML for this method?

java   java.util.scanner   specifications   jml

★

share edit

asked Nov 10 at 7:48

May1620 iastate
1 ● 1

Example Questions - More Details

What is the postcondition for the java PriorityQueue.remove(Object) method?

I am looking into creating JML specifications for the java.util.PriorityQueue.remove(Object object) method. So far I have thought of the following precondition:

```
//@ requires object != null;
//@ requires this.size > 0;
```

I am now trying to figure out the postcondition. So what is the ensures field for this method? I feel that is should involve the size() method and making sure that the data is no longer in the queue but I am not sure how to write this.

java   priority-queue   specifications   formal-verification   jml

share edit delete flag

asked just now
Alex1620
1 • 1

May1620 - http://may1620.github.io/
Alex Dana, Cody Hanika, Deeksha Juneja

❏ Formal specifications allow programmers to more easily write correct code
   ❏ They are not currently widely available and are often misunderstood
❏ Our solution will combine previous efforts of a purely technical approach with the social aspect of StackOverflow