
Socio-Technical Solution to Large-Scale Formal Specification Mining

May1620

Members: Alex Dana & Cody Hanika

Client/Advisor: Professor Hridesh Rajan

Project Website: <http://may1620.github.io/>

What are formal specifications?

- Allow a programmer to describe the behavior of a piece of code in a programmatic and verifiable manner.
 - Consist of a precondition and postcondition pair that specify what must be true about the program state
 - 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 by analyzing pieces of code
-

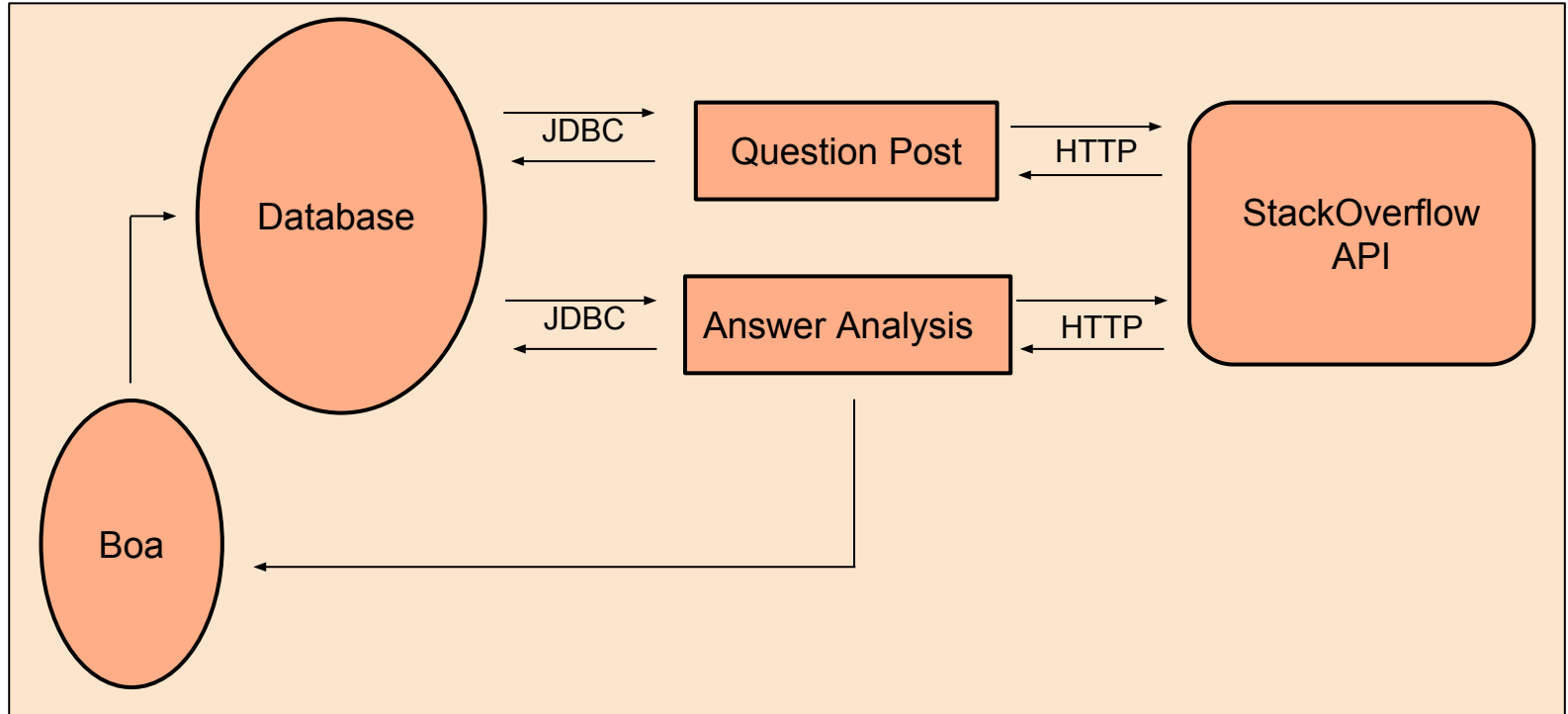
Initial Approach

- Use the success of the popular StackOverflow site to crowdsource the specifications
 - A system will automatically post questions asking for specifications, then automatically pull down the answers
 - Store these specifications in Boa, a software repository mining infrastructure, to make them available to users
 - Eventually link our project with a custom Q&A site based on StackOverflow developed by another group
-

(Big) Risks

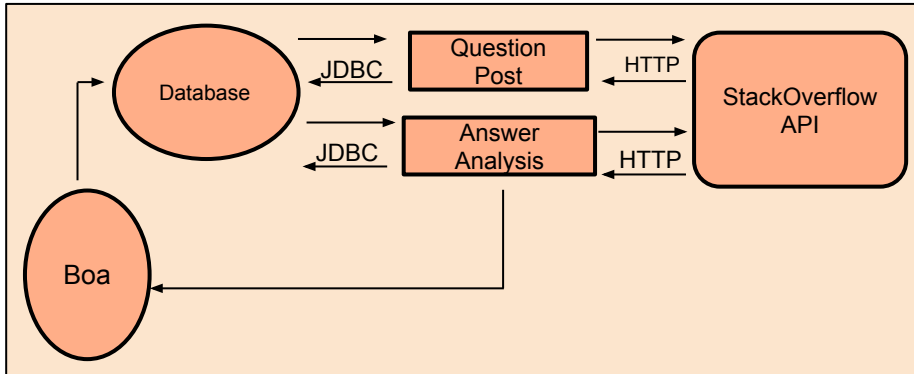
- As formal specifications are not well understood by the general programmer, there may be trouble getting quality responses from users
 - StackOverflow may restrict us to posting very rarely
-

System Overview



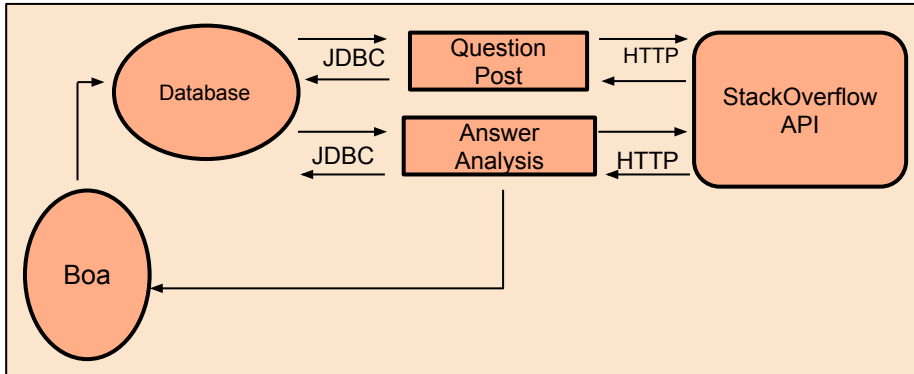
Posting Expanded

- Database grabs a function from Boa that does not have a spec created
- Question is generated based on the function
- Posting requires OAuth 2.0 authentication



Analysis Expanded

- Grabs question ID of functions that currently have questions posted for them (that are not complete)
- We grab those questions via the API and then extract JML from them
 - Use upvotes and accepted answers to determine quality



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 flag

asked Nov 10 '15 at 7:48



May1620 iastate

1 • 1

[add a comment](#)

[start a bounty](#)

Our initial ideal question that would be answered

What assumptions can we safely make about the Java Collections `shuffle(List<?>, Random)` method?



3



So I am looking into the collections `shuffle` method and trying to come up with a list of what is and is not ensured when we run it. There are some obvious cases I've come up which are the following:

1. The list given will contain the same elements after shuffling as before
2. The list may or may not be the same after running the method (you could end up with the same order of elements)
3. The method will run in linear time (I think that this is true but am not 100% positive).

Does this list sum it up or am I missing some possible cases?

java

shuffle

share edit flag

edited Jan 11 at 22:19

asked Jan 11 at 21:36



Alex1620

165 ● 15

asked 3 months ago

viewed 67 times

active 3 months ago

Successful question



11



The [official documentation of `Collections.shuffle`](#) has a lot to say about what will happen. The list will be shuffled using what seems to be the [Fisher-Yates shuffle algorithm](#), which (assuming that random access is available in $O(1)$) runs in time $O(n)$ and space $O(1)$. The implementation will use space $O(n)$ if random access isn't available. Assuming that the underlying random source is totally unbiased, the probability of any particular ordering occurring is equal (that is, you get a uniformly-random distribution over possible permutations).

So, to answer your questions:

1. The list will contain the same elements.
2. They're probably in a different order, but there's a $1 / n!$ chance than they'll be in the same order.
3. The runtime is $O(n)$, and the space usage is either $O(1)$ or $O(n)$ depending on whether your list has random access support.

[share](#) [edit](#) [flag](#)

[edited Jan 11 at 22:33](#)

[answered Jan 11 at 21:46](#)



[templatetypedef](#)

183k



41



461



703

Unstructured answer

What can be safely assumed about the java Collections.sort(List<T>) method after it runs? [on hold]



-8



So if we use the sort method what can we safely expect (assuming it ran properly)? So far using the javadoc I have:

- Our list will be the same size as it was before
- Equal elements won't be rearranged
- Our list will contain the same elements as before

java

sorting

collections

share edit delete flag

asked 1 hour ago



Alex1620

149 • 15

put on hold as unclear what you're asking by [Sean Bright](#), [Cássio Mazzochi Molin](#), [Hovercraft Full Of Eels](#), [trw](#), [Trilarion](#) 52 mins ago

Please clarify your specific problem or add additional details to highlight exactly what you need. As it's currently written, it's hard to tell exactly what you're asking. See the [How to Ask](#) page for help clarifying this question.

If this question can be reworded to fit the rules in the [help center](#), please [edit your question](#).

6 You can assume it will do everything it says in the Javadoc. Apart from just copying and pasting that, what sort of answer are you looking for? – [resueman](#) 1 hour ago

Typical comment/answer

Issues

- The risks we were initially concerned with became real problems
 - The lack of JML knowledge forced us to ask more informal questions which were not typically well received
 - When questions were answered they typically involved telling us to look at the documentation
 - Our StackOverflow account was eventually banned
-

New Approach

- Use currently existing javadoc to gather informal specifications
 - Javadoc solves the structuring problem
 - Generate JML utilizing a natural language processor
 - Attempt to handle the most commonly occurring specifications first in order to maximize coverage
-

Goal for the New Approach

Throws:

`NullPointerException` - if the specified collection is null



```
/*
 * @ ...
 * @ requires c == null;
 * @ signals_only NullPointerException;
 */
public boolean addAll(Collection<? extends E> c)
```

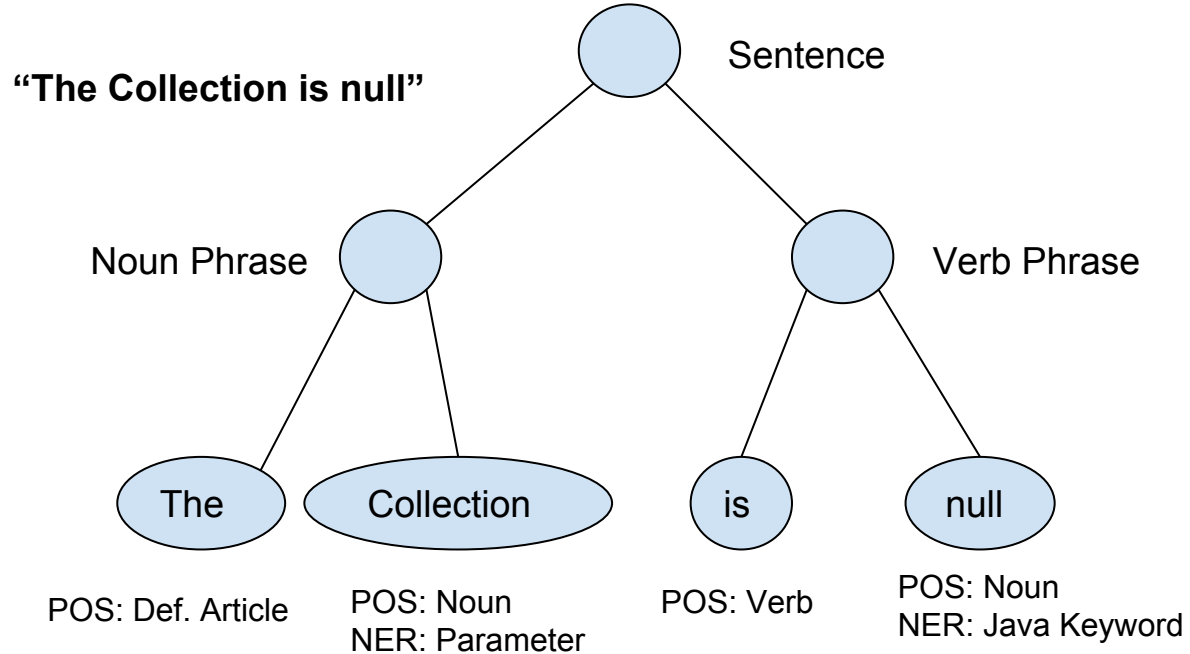
New Risks

- This is a really difficult problem
 - Limited time
 - Our group had no prior knowledge of natural language processors
 - Javadoc is not written in a consistent manner (helpful for parsing)
-

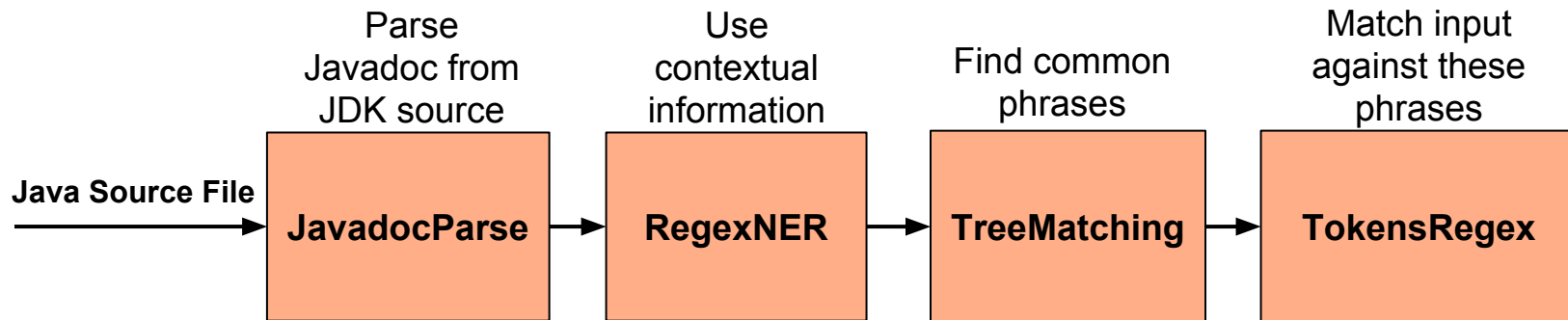
StanfordNLP

- Powerful many-featured natural language processor built with Java
 - Breaks sentences into “tokens” (words) and analyzes based on things such as part of speech
 - The analysis results in a parse tree
-

Example Parse Tree




New System Overview



RegexNER Stage

- We need a way to use contextual information
 - Method name
 - Method parameter types/names
 - Java keywords
- This stage looks at each word and adds an NER annotation that marks words that relate to the context



```
public String substring(int beginIndex, int endIndex)
```

TreeMatching Stage

- As seen earlier SNLP creates parse trees
 - Processing the entire JDK Javadoc results in similar parse trees (or similar subtrees of parse trees)
 - These subtrees represent similar structure of informal specifications
 - We find the most common identical subtrees to determine common informal spec patterns
-

TokensRegex Stage

- TokensRegex takes regular expression “rule” files
- These files dictate how informal specifications will be transformed

```
ENV.defaults["stage"] = 4
{
  ruleType: "tokens",
  pattern: ( [{ner:PARAMETER}] /is/ [{ner:NULL}]),
  result: Format("@requires %s == null", $0[0].value)
}
```

Result: 'Throws NullPointerException if collection is null' -> 'collection is null' ->
'PARAMETER is NULL' -> '@requires c == null; @signals_only NullPointerException;'

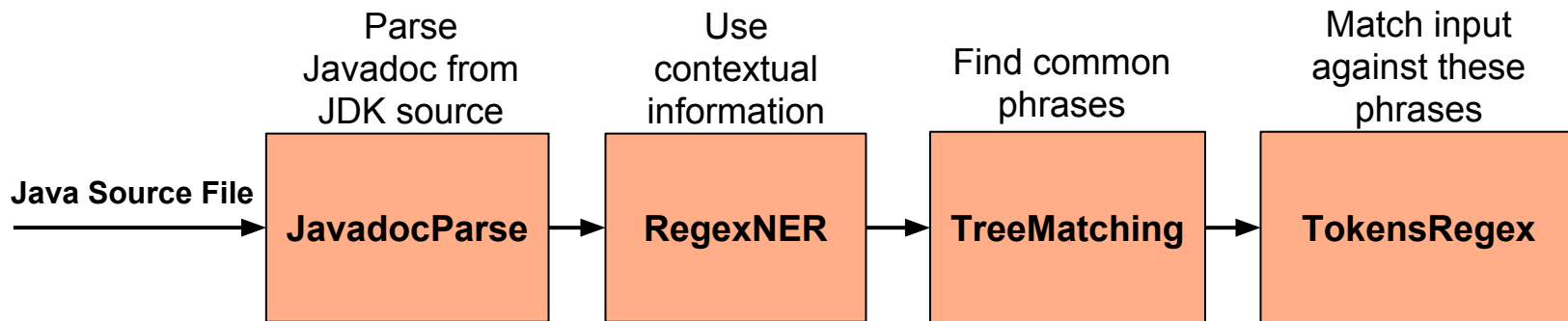
Results

Java package	Match Rate
util	50%
io	11%
lang	25%
security	21%

Closing Remarks

May1620

Project Website: <http://may1620.github.io/>



Analysis run on the Java util package

```
1 public class RecommendationVisitor extends MYVisitor {
2
3     private boolean classified;
4
5     public RecommendationVisitor() {
6         classifier = false;
7     }
8
9     boolean
10     public boolean visit(TypeDeclaration node) {
11         if (node.isPublic()) {
12             classifier = true;
13         }
14     }
15
16     boolean
17     public boolean visit(FieldDeclaration node) {
18         if (node.isPublic()) {
19             classifier = true;
20         }
21     }
22 }
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018

```