Informed Consent

Understandable Test Generation Through LLM-generated summaries

Background

Unit test cases can be generated by several tools. One can achieve a high coverage, but it hard to understand because of the names of the variables and lack of documentation, for example. There exists for this several tools which serve as a solution to improve understand

This survey

In this survey, we will evaluate the impact on understandability by looking at summaries cruby different tools. This survey does not aim to evaluate the automatically written test cases

The first part is about your experience with Java and test generation frameworks. The secc is split into four rounds, with the first two rounds having 4 different test cases, and the last t rounds having 3 different test cases. It starts with the original test case without any summand/or comments for reference.

This survey contains 20 questions and will take ~15 minutes to complete. There are optiona feedback forms intended to allow you to elaborate on your answers if appropriate.

Thank you for taking the time to participate in this survey!

Part I - Background

Part I - Background

Q1.1 What is your primary profession?*

- O Student (Bachelor's)
- O Student (Masters)
- O Student (PhD)
- O Software Developer

Q1.2 How many years of experience do you have with Java?*



Part II - Round 1

Part II - Summaries

Round 1

In this section, you will be asked to evaluate the quality of the summaries created by different tools.

First, we present you the original automatically generated test case in each round, follower four summaries A, B, C, and D.

Original test

```
public void test10() throws Throwable {
    ArrayIntList arrayIntList0 = new ArrayIntList();

try {
    arrayIntList0.add(0, 0);
    arrayIntList0.add(0, 1);
    arrayIntList0.add(0, 2);
    assertEquals(3, arrayIntList0.size());
    arrayIntList0.removeElementAt((1));
    fail("Expecting exception: IndexOutOfBoundsException");
} catch (IndexOutOfBoundsException e) {
    //
    // Should be at least 0 and less than 0, found -1
    //
}
```

Method being tested

```
public int removeElementAt(int index) {
    checkRange(index);
    incrModCount();
    int oldval = _data[index];
    int numtomove = _size - index - 1;
    if(numtomove > 0) {
        System.arraycopy(_data,index+1,_data,index,numtomove);
    }
    __size--;
    return oldval;
}

private final void checkRange(int index) {
    if(index < 0 || index >= _size) {
        throw new IndexOutOfBoundsException("Should be at least 0 and less than " + _size + ", found "
    }
}
```

Method called

```
public void add(int index, int element) {
    checkRangeIncludingEndpoint(index);
    incrModCount();
    ensureCapacity(_size+1);
    int numtomove = _size-index;
    System.arraycopy(_data,index,_data,index+1,numtomove);
    _data[index] = element;
    _size++;
}
```

Here below are the four summaries A, B, C, and D.

A:

```
* OVERVIEW: The test case "test10" covers around 2.0% (low percentage)
      * statements in "ArrayIntList"
     @Test
     public void test10() throws Throwable {
             // The test case instantiates a "ArrayIntList" with the defaul
             // configuration (initial capacity is 8)
             ArrayIntList arrayIntList();
11
             try {
                 // The next method call removes the element at index -1 of
12
                 // "arrayIntList0"
13
                 arrayIntList0.add(0, 0);
15
                 arrayIntList0.add(0, 1);
16
                 arrayIntList0.add(0, 2);
17
                 assertEquals(3, arrayIntList0.size());
                 arrayIntList0.removeElementAt((1));
18
19
                 fail("Expecting exception: IndexOutOfBoundsException");
20
             } catch (IndexOutOfBoundsException e) {
21
22
24
25
```

B:

```
* 1. Creates a new ArrayIntList
      * 2. Adds to "arrayIntList0" 3 times and checks if its size is
      * 3. Expects an IndexOutOfBoundsException when calling
           removeElementAt on "arrayIntList0" with argument 1
     @Test
     public void test10() throws Throwable {
         ArrayIntList arrayIntList0 = new ArrayIntList();
11
         try {
12
           arrayIntList0.add(0, 0);
13
           arrayIntList0.add(0, 1);
           arrayIntList0.add(0, 2);
15
           assertEquals(3, arrayIntList0.size());
           arrayIntList0.removeElementAt(1);
17
           fail("Expecting exception: IndexOutOfBoundsException");
18
          } catch (IndexOutOfBoundsException e) {
19
           // Should be at least 0 and less than 0, found -1
21
           //
22
23
```

C:

D:

Q2.1 How would you rate the **content** of the four summaries?*

	(1) Missing much important information or information mostly incorrect	(2)	(3) Missing some important information or information is somewhat incorrect	(4)	(5) Not missing any important information or no incorrect information
А	\circ	0	\circ	0	0
В	\circ	\circ	\circ	\circ	\circ
С	\circ	\circ	\circ	\bigcirc	\circ
D	\circ	\bigcirc	\circ	\circ	\circ

(Optional) Please elaborate your answer if appropriate.

Q2.2 How would you rate the **conciseness** of the four summaries?*

A B C D (Optional) Pleas	0 0 0	0 0 0	0 0	0	0		
C D	0 0	0	0	0	0		
D	0	0	0	0	0		
	0	0	\circ				
(Optional) Pleas				0	0		
	se elabora	te your	answer if a	ppropr	iate.		
Q2.3 How would summaries?*		he nat i	uralness of	the fo	ur		
	(1) The flow and tone are not smooth and		(3) The flow and tone are somewhat		(5) The flow and tone are smooth		
	and tone are not	(2)	and tone	(4)			
A	and tone are not smooth and appropriate	(2)	and tone are somewhat smooth and	(4)	and tone are smooth and		
	and tone are not smooth and appropriate	(2)	and tone are somewhat smooth and	(4) O	and tone are smooth and		
А В С	and tone are not smooth and appropriate	(2) O	and tone are somewhat smooth and	(4) O O	and tone are smooth and		

Q2.4 Considering conciseness, content, and nature (Multiple possible)	ralness, which summary do you prefer to
Summary A	
☐ Summary B	
Summary C	
Summary D	
None	
(Optional) Feel free to leave any comments for the	ne 4 proposed summaries here.
Q2.5 What aspects do you like the most and the corresponding tool with it. Examples are shown be your own.*	elow, but you are encouraged to come up
~ Example: I like [aspects] in summary B the most ~ Examples of aspects: using a numbered list, len included, in-line comments, explaining what it do	gth of summary, having the code coverc
Aspect(s) liking the most	
Aspect(s) liking the least	

Part II - Round 2

Round 2

As in *round 1*, we present you the original automatically generated test case in each round followed by the four summaries A, B, C, and D. Please note that a specific summary from rolike summary A, is **not** necessarily made with the same tool as summary A from this rounc

Original test

```
public void test3() throws Throwable {
   Rational rational0 = new Rational(1L, 3215L);
   Rational rational1 = rational0.abs();
   assertEquals(1L, rational0.numerator);
   assertEquals(3215L, rational0.denominator);
   assertEquals(3.11041E-4F, rational1.floatValue(), 0.01F);
}
```

Method being tested

```
public Rational abs() {

return new Rational((numerator < 0L) ? -numerator : numerator, (denominator < 0L) ? -denominator : denominator }

public Rational abs() {

return new Rational((numerator < 0L) ? -numerator : numerator, (denominator < 0L) ? -denominator : denominator }

public Rational abs() {
```

Here below are the four summaries A, B, C, and D.

A:

```
/**
2  * Tests the absolute value function of the Rational class to ensure it
3  * correctly handles positive values and returns the expected float representation.
4  */
5  @Test
6  public void test3() throws Throwable {
7    Rational rational0 = new Rational(1L, 3215L);
8    Rational rational1 = rational0.abs();
9    assertEquals(1L, rational0.numerator);
10    assertEquals(3215L, rational0.denominator);
11    assertEquals(3.11041E-4F, rational1.floatValue(), 0.01F);
12 }
```

B:

```
* OVERVIEW: The test case "test3" covers around 6.0% (low percentage)
      * statements in "Rational"
     @Test
     public void test3() throws Throwable {
         // The test case instantiates a "Rational" with numerator equal to
         // and denominator equal to 3215L.
         // The execution of this constructor implicitly covers the followin
10
         // conditions:
11
         // - the condition " denominator equals to 0L" is FALSE;
12
         Rational rational0 = new Rational(1L, 3215L);
13
         // The test case declares an object of the class "Rational" whose v
         // is equal to the absolute value of "rational0"
14
15
         Rational rational1 = rational0.abs();
16
         // Then, it tests:
17
         // 1) whether the numerator of rational0 is equal to 1L;
18
         assertEquals(1L, rational0.numerator);
19
         // 2) whether the denominator of rational0 is equal to 3215L;
20
         assertEquals(3215L, rational0.denominator);
21
         // 2) whether the float value of "rational1" is equal to 3.11041E-4
22
         // with delta equal to 0.01F;
23
         assertEquals(3.11041E-4F, rational1.floatValue(), 0.01F);
24
```

C:

D:

Q2.6 How would you rate the **content** of the four summaries?

414	

	(1) Missing much important information or information mostly incorrect	(2)	(3) Missing some important information or information is somewhat incorrect	(4)	(5) Not missing an important informatio or no incorrect informatio
А	\circ	\circ	\circ	\circ	\circ
В	\circ	\bigcirc	\circ	\bigcirc	\bigcirc
С	\circ	\circ	\circ	\bigcirc	\circ
D	\circ	\circ	\bigcirc	\bigcirc	\circ

(Optional) Please elaborate your answer if appropriate.

Q2.7 How would you rate the **conciseness** of the four summaries?*

		(3)		
(1) Contains		Contains		(5)
mostly		some		Contains no
unnecessary information	(2)	unnecessary information	(4)	unnecessary information
\bigcirc	\circ	\circ	0	\circ

	(1) Contains mostly unnecessary information	(2)	(3) Contains some unnecessary information	(4)	(5) Contains no unnecessary information	
3	\circ	\circ	0	\bigcirc	0	
	\circ	\circ	\circ	\bigcirc	\circ	
	0	0	0	0	0	
ptional)	Please elabora	te you	r answer if a	ppropr	iate.	
28 How w	ould you rate tl	ne nat	ura iness ∩f	the fo	ır	
ummaries	•	ie iidi	urumess or	116 10	ui	
	(1) The flow and tone are not smooth and appropriate at all	(2)	(3) The flow and tone are somewhat smooth and appropriate	(4)	(5) The flow and tone are smooth and appropriate	
		\circ	\circ	\circ	\circ	
	\circ				_	
	0	\bigcirc	\circ	\circ	\circ	
	0 0	0	0	0	0	
	0 0 0	0	0	0	0	
A B C D	0 0 0	0 0	0	0	0 0	
B C D	O O O Please elabora	O O			O O iate.	
B C D	O O O Please elabora	o o te you			iate.	

23-06-2024, 18:33 Summary C Summary D None	Qualtrics Survey Software
(Optional) Feel free to leave any comment	ts for the 4 proposed summaries here.
Q2.10 What aspects do you like the most ar	nd the least in the four summaries? Please also
corresponding tool with it. Examples are shown own.* ~ Example: I like [aspects] in summary B the content of aspects: using a numbered	own below, but you are encouraged to come up
Aspect(s) liking the most	
Aspect(s) liking the least	

Part II - Round 3

Round 3

As in the previous rounds, we present you the original automatically generated test case ir round, but instead of followed by four summaries, these two last rounds will be followed by summaries A, B, and C. Please note that a specific summary from previous rounds, like sun is **not** necessarily made with the same tool as summary A from this round.

Original test

```
1  @Test(timeout = 4000)
2  public void testEncodedPath() throws Throwable {
3    KeycloakUriBuilder uri = KeycloakUriBuilder.fromPath("x");
4    HashMap<String, Integer> map = new HashMap<String, Integer>();
5    URI result = uri.buildFromEncodedMap(map);
6    assertEquals("x", result.getRawPath());
7 }
```

The summaries A, B, and C are below.

A:

B:

```
/**
2  * Tests that KeycloakUriBuilder correctly builds a URI from an encoded map,
3  * ensuring that the raw path remains unchanged.
4  */
5  @Test(timeout = 4000)
6  public void testEncodedPath() throws Throwable {
7   KeycloakUriBuilder uri = KeycloakUriBuilder.fromPath("x");
8   HashMap<String, Integer> map = new HashMap<String, Integer>();
9   URI result = uri.buildFromEncodedMap(map);
10  assertEquals("x", result.getRawPath());
11 }
```

C:

Q2.11 How would you rate the **content** of the three summaries?*

Q2.13 How would you rate the **naturalness** of the three summaries?*

Q2.15 What aspects do you like the most and the least in the four summaries? Please also corresponding tool with it. Examples are shown below, but you are encouraged to come up your own.*

- ~ Example: I like [aspects] in summary B the most.
- ~ Examples of aspects: using a numbered list, length of summary, having the code coverc included, in-line comments, explaining what it does, step-by-step explanation of how it wo

Aspect(s) liking the most	
Aspect(s) liking the least	

Part II - Round 4

Round 4

As in the previous rounds, we present you the original automatically generated test case in round. This last round will be followed by three summaries A, B, and C as it was in round 3. note that a specific summary from previous rounds, like summary A, is **not** necessarily ma the same tool as summary A from this round.

Original test

Method being tested

```
public void ensureCapacity(int mincap) {
    incrModCount();
    if(mincap > _data.length) {
        int newcap = (_data.length * 3)/2 + 1;
        int[] olddata = _data;
        _data = new int[newcap < mincap : newcap];
        System.arraycopy(olddata,0,_data,0,_size);
    }
}</pre>
```

Methods called

The summaries A, B, and C are below.

A:

B:

```
/**
2  * Create a new instance of ArrayIntList
3  * Ensure the list has enough capacity for 9 elements
4  * Assert that the size of the list is 0 (empty)
5  */
6  @Test
7  public void test8() throws Throwable {
8     ArrayIntList arrayIntList0 = new ArrayIntList();
9     arrayIntList0.ensureCapacity(9);
10     assertEquals(0, arrayIntList0.size());
11 }
```

C:

```
/**
2 * Ensures that the capacity of an ArrayIntList can be increased to a specified value without affecting its s.
3 */
4 @Test
5 public void test8() throws Throwable {
6 ArrayIntList arrayIntList0 = new ArrayIntList();
7 arrayIntList0.ensureCapacity(9);
8 assertEquals(0, arrayIntList0.size());
9 }
```

Q2.16 How would you rate the **content** of the three summaries?*

	(1) Missing		(3) Missing		
	much		some		(5) Not
	important		important		missing any
	information		information		important
	or		or		information
	information		information		or no
	mostly		is somewhat		incorrect
	incorrect	(2)	incorrect	(4)	information
А	\circ	\circ	\bigcirc	\circ	\circ
В		\circ	\bigcirc	\bigcirc	\circ

Q2.18 How would you rate the **naturalness** of the three summaries?*

23-06-2024, 18:33		Qualtrics Survey Software						
	(1) The flow and tone are not smooth and appropriate at all	(2)	(3) The flow and tone are somewhat smooth and appropriate	(4)	(5) The flow and tone are smooth and appropriate			
А	\circ	\bigcirc	\circ	\bigcirc	\circ			
В	\circ	\bigcirc	\bigcirc	\bigcirc	\bigcirc			
С	0	0	0	0	0			
(Optional) Pl	lease elabora	te youı	answer if a	ppropr	iate.			
Q2.19 Consid (Multiple pos	•	ness, c	ontent, and	natura	Iness, which s	summary do	you prefer to	
Summary A								
Summary B								
☐ Summary C☐ None								
(0.45	<i>f</i> u -			6 <u>1</u> 1	2			
(Optional) Fe	eel free to leav	ve any	comments	for the	3 proposed s	summaries r 	nere. 	

Q2.20 What aspects do you like the most and the least in the four summaries? Please also the corresponding tool with it. Examples are shown below, but you are encouraged to com with your own.*

- ~ Example: I like [aspects] in summary B the most.
- ~ Examples of aspects: using a numbered list, length of summary, having the code coverc included, in-line comments, explaining what it does, step-by-step explanation of how it wo

Powered by Qualtrics