CallMeMaybe: Using NLP to Automatically Generate Unit Test Cases Respecting Temporal Constraints

Arianna Blasi[♦] · Alessandra Gorla[♥] · Michael D. Ernst[♠] · Mauro Pezzè[♦♣]

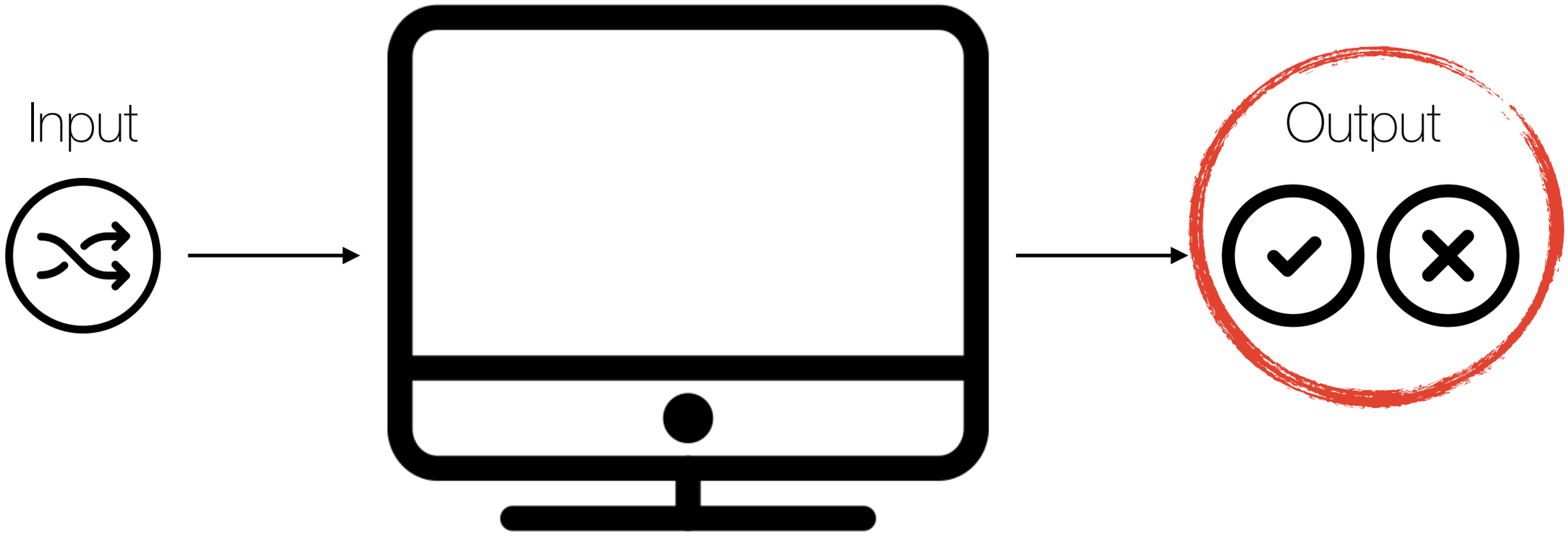
[♦]USI Università della Svizzera italiana, Switzerland

[♣]SIT Institute of Technology, Switzerland

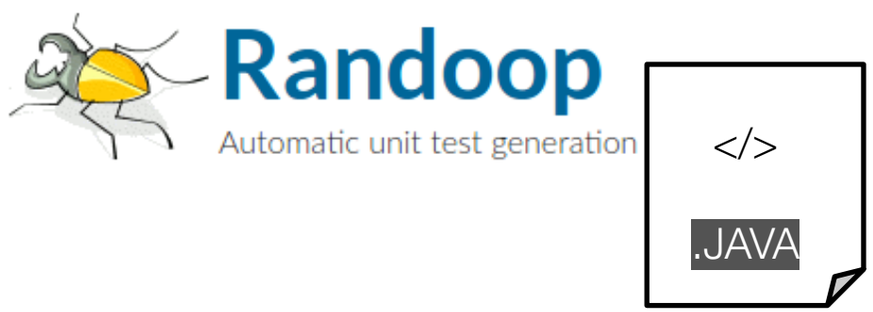
[♥]IMDEA Software Institute, Spain

[♠]University of Washington Seattle, USA

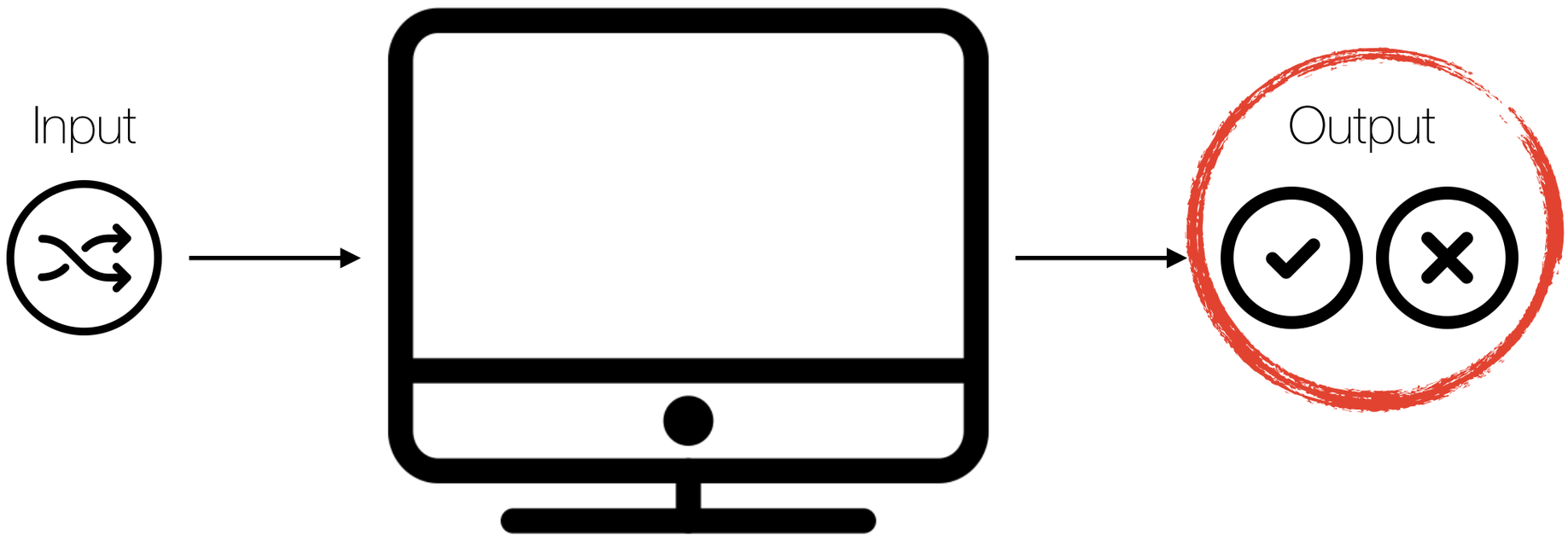
Automatic Test Case Generation



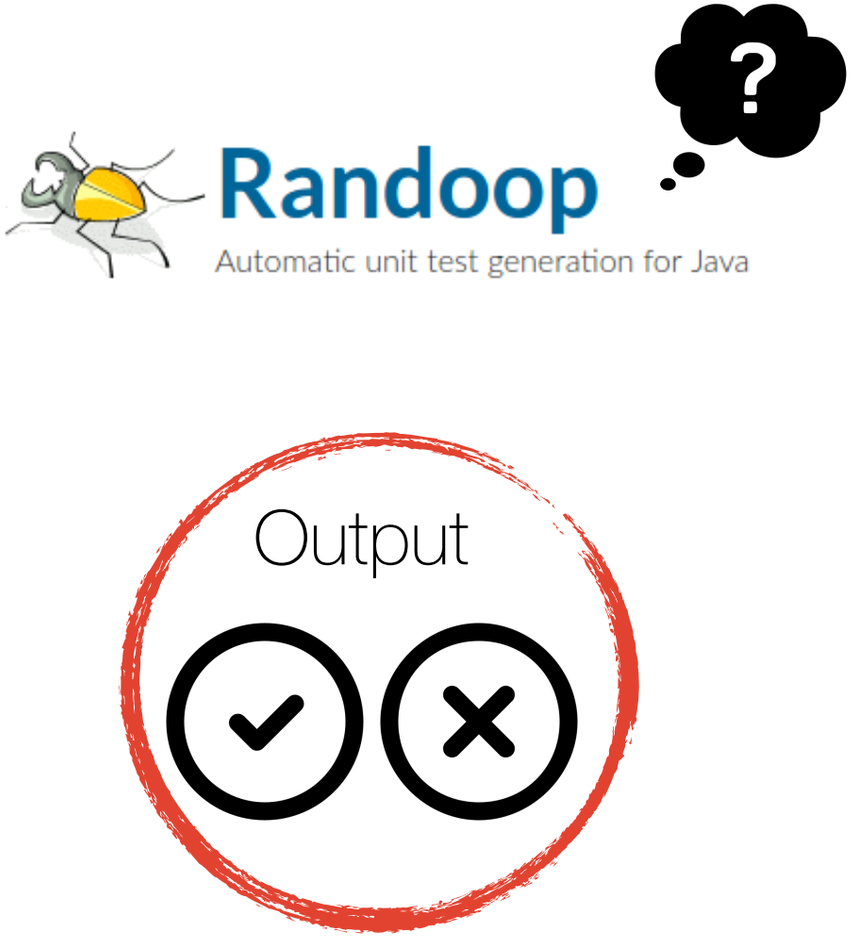
Automatic Test Case Generation



- Randomly select **method calls + inputs**
- Assess **correctness** of the outcome



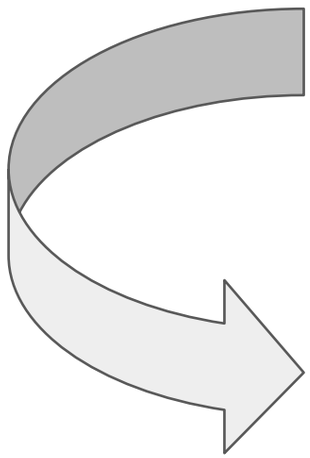
Automatic Test Case Generation



Where can the generator gather ***knowledge*** *about the semantics of the SUT?*

How can a user check the output of ***thousands*** *of automatically generated test cases?*

```
public IteratorEnumeration() {…}
```



THE APACHE SOFTWARE FOUNDATION

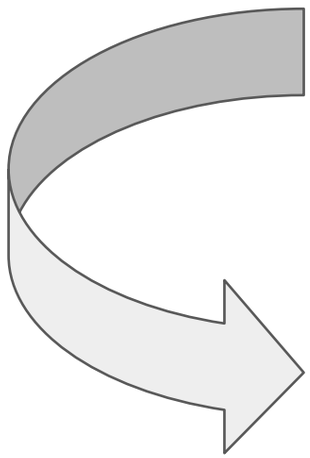
```
@Test
public void test001(){
    IteratorEnumeration<String> strItEn0 =
                new IteratorEnumeration<String>();

    String str1 = strItEn0.nextElement();
}
```

Randoop

Automatic unit test generation for Java

```
public IteratorEnumeration() {…}
```

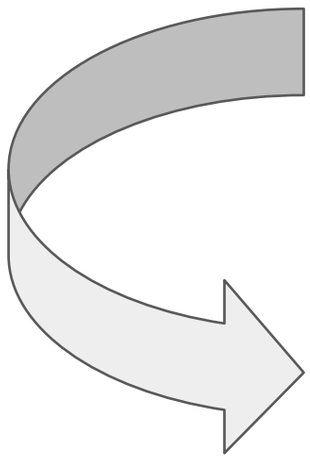


```
@Test
public void test001(){
    IteratorEnumeration<String> strItEn0 =
                new IteratorEnumeration<String>();

    String str1 = strItEn0.nextElement();
}
```

NullPointerException

```
public IteratorEnumeration() {…}
```



```
@Test
public void test001(){
    IteratorEnumeration<String> strItEn0 =
            new IteratorEnumeration<String>();

    String str1 = strItEn0.nextElement();
}
```

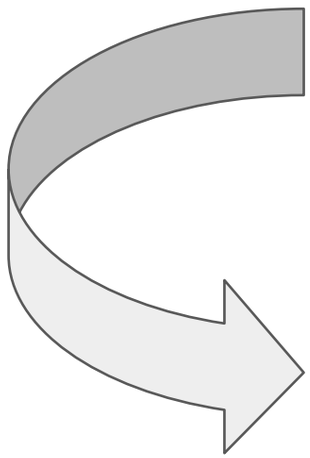
NullPointerException

Your program has a bug!

Randoop

Automatic unit test generation for Java

```
public IteratorEnumeration() {…}
```



```
@Test
public void test001(){
    IteratorEnumeration<String> strItEn0 =
                new IteratorEnumeration<String>();

    String str1 = strItEn0.nextElement();
}
```

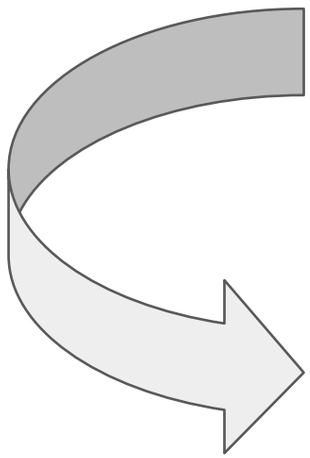
NullPointerException

...

```
/**
* Constructs a new IteratorEnumeration that will
* not function until setIterator(Iterator) is
* invoked.
*/

public IteratorEnumeration() {…}
```

THE APACHE SOFTWARE FOUNDATION



```
@Test
public void test001(){
    IteratorEnumeration<String> strItEn0 =
                new IteratorEnumeration<String>();

    String str1 = strItEn0.nextElement();
}
```

✓ **NullPointerException**

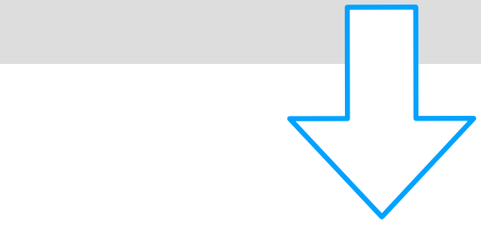
Randoop

Automatic unit test generation for Java

Alright: expected.

Improving Test Case Generators

```
[
  {
    "operationSignature": "org.apache.commons.collections4.iterators.IteratorEnumeration",
    "isConstructor": true,
    "mustPrecede": "setIterator(java.util.Iterator<? extends E>)",
    "mustFollow": "",
  }
]
```



Randoop

Automatic unit test generation for Java

Improving Test Case Generators

```
[
  {
    "operationSignature": "org.apache.commons.collections4.iterators.IteratorEnumeration",
    "isConstructor": true,
    "mustPrecede": "setIterator(java.util.Iterator<? extends E>)",
    "mustFollow": "",
  }
]
```

```
@Test
public void te
    IteratorEnu                                    tEn0 =
                new                                String>();

    String str1 = strItEn0.nextElement();
}
```

Randoop

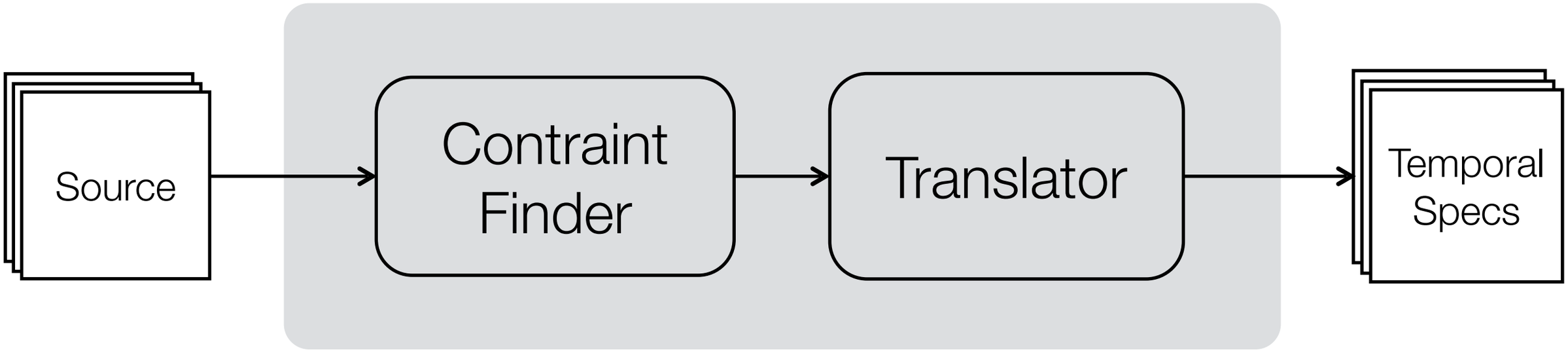
Automatic unit test generation for Java

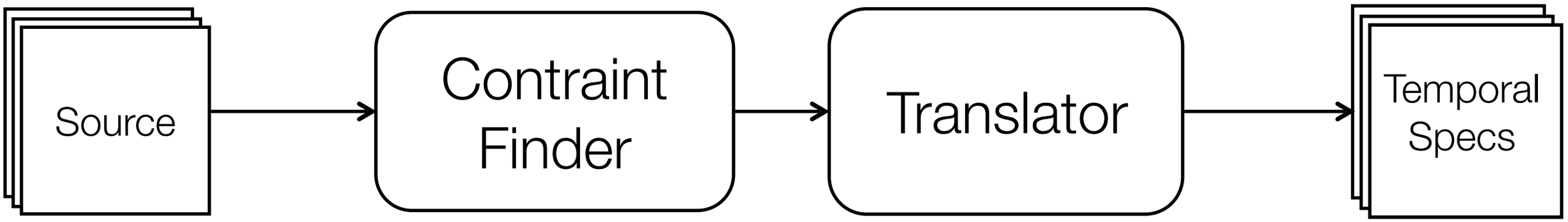
Bug revealing

Non-bug revealing

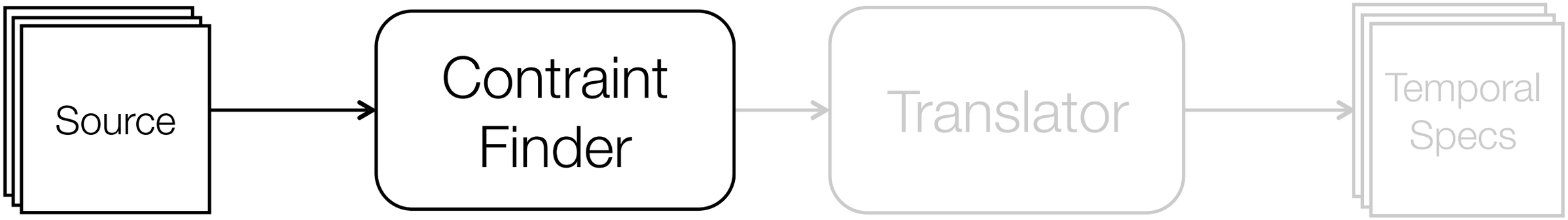
CallMeMaybe

Translating natural language method ordering information into Java code



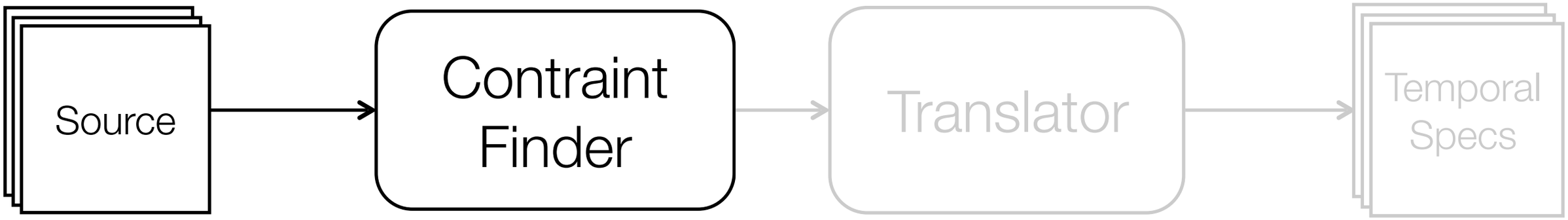



```
/**
* This method must be invoked before the thread
* is started.
*/
public final void setDaemon(boolean on)
```



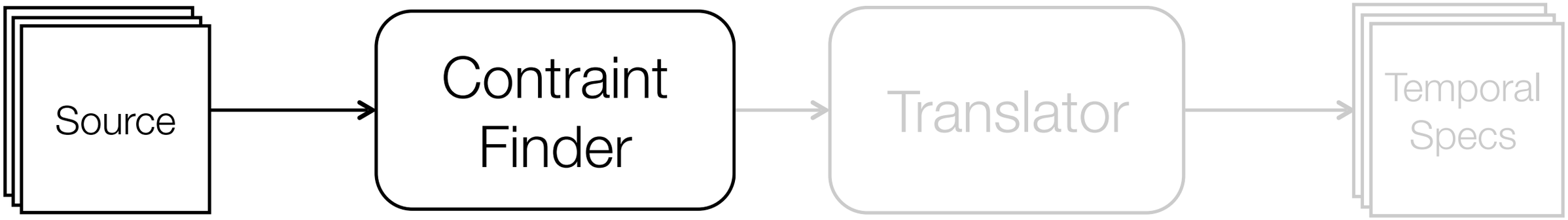

```
“ This method must be invoked before the thread is started ”
```

1. Subject relations
2. Adverbial relations



“ This method must be invoked before the thread is started ”

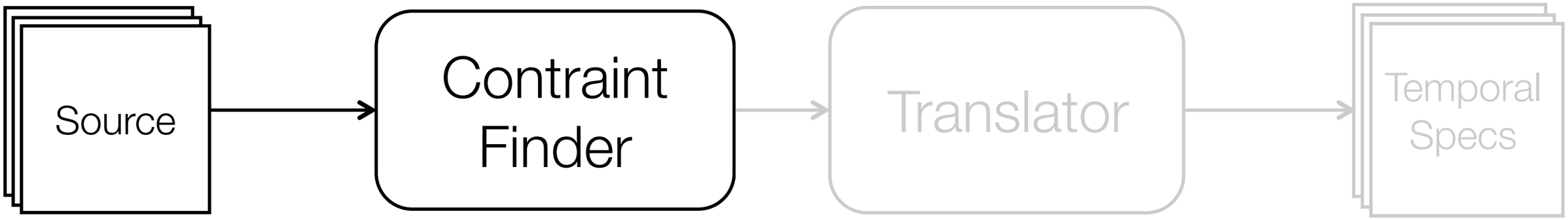
Subject	Predicate
“This method”	“be invoked”



" This method must be invoked before the thread is started "

Subject	Predicate
"this method"	"be invoked"
"thread"	"is started"

Stanford | NLP



" This method must be invoked before the thread is started "

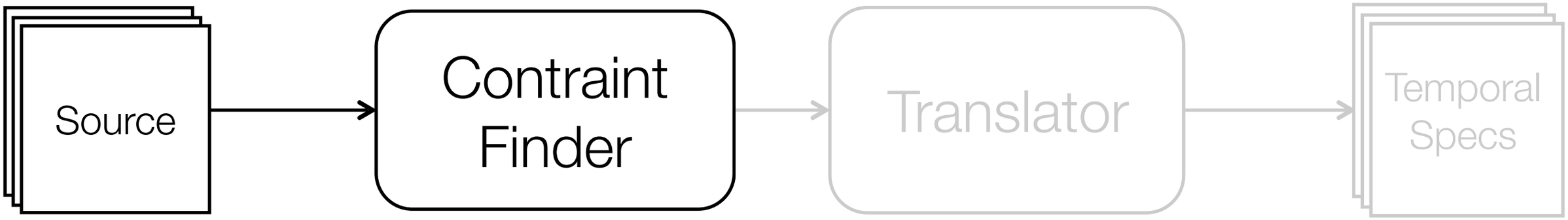
adverb: before

Governor	Adverb	Dependent
"invoked"	"before"	"started"

subject + verb: (this method, be invoked)

subject + verb: (thread, is started)

Stanford | NLP



" This method must be invoked **before** the thread is started "

adverb: before

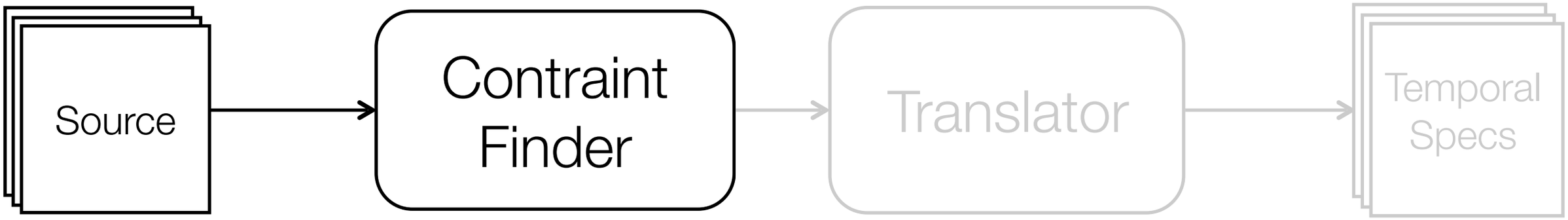
Governor	Adverb	Dependent
"invoked"	"before"	"started"

adverb (before): (invoked, started)

subject + verb: (this method, be invoked)

subject + verb: (thread, is started)

Stanford | NLP



" This method must be invoked **before** the thread is started "

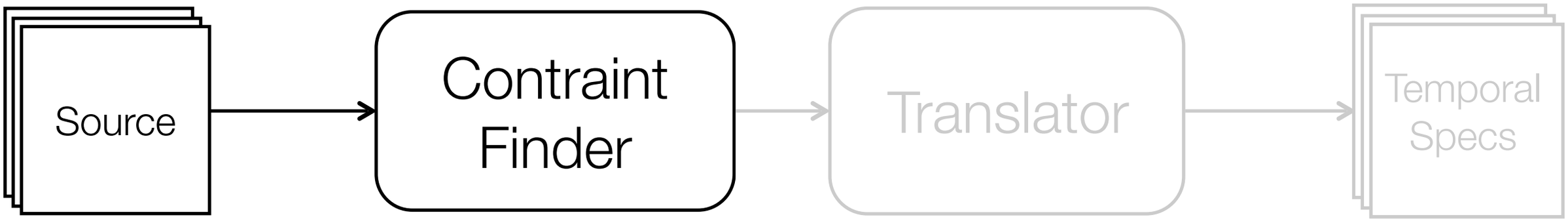
Governor	Adverb	Dependent
"invoked"	"before"	"started"

adverb (before): (invoked, started)

subject + verb: (this method, be invoked)

subject + verb: (thread, is started)

Stanford | NLP



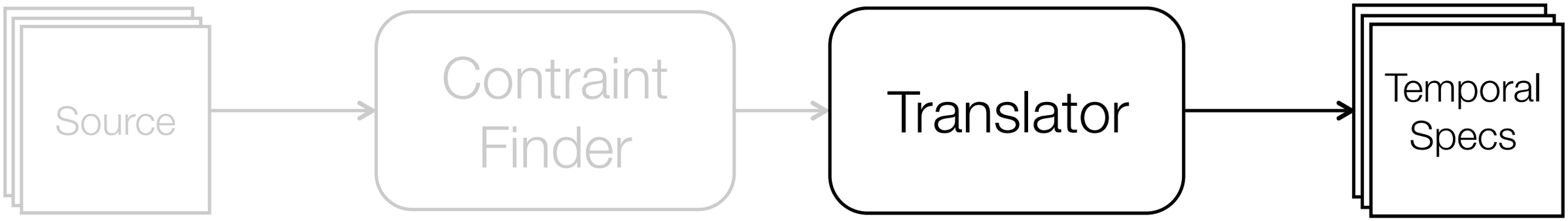
“ This method must be invoked before the thread is started ”

adverb (before): (invoked, started)

subject + verb: (this method, be invoked)

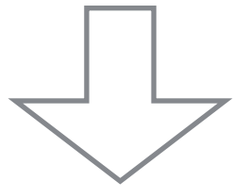
subject + verb: (thread, is started)

(this method, be invoked) **BEFORE** (thread, is started)

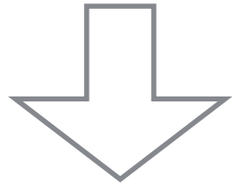


(this method, be invoked) **BEFORE** (thread, is started)

(this method, be invoked) **BEFORE** (thread, is started)

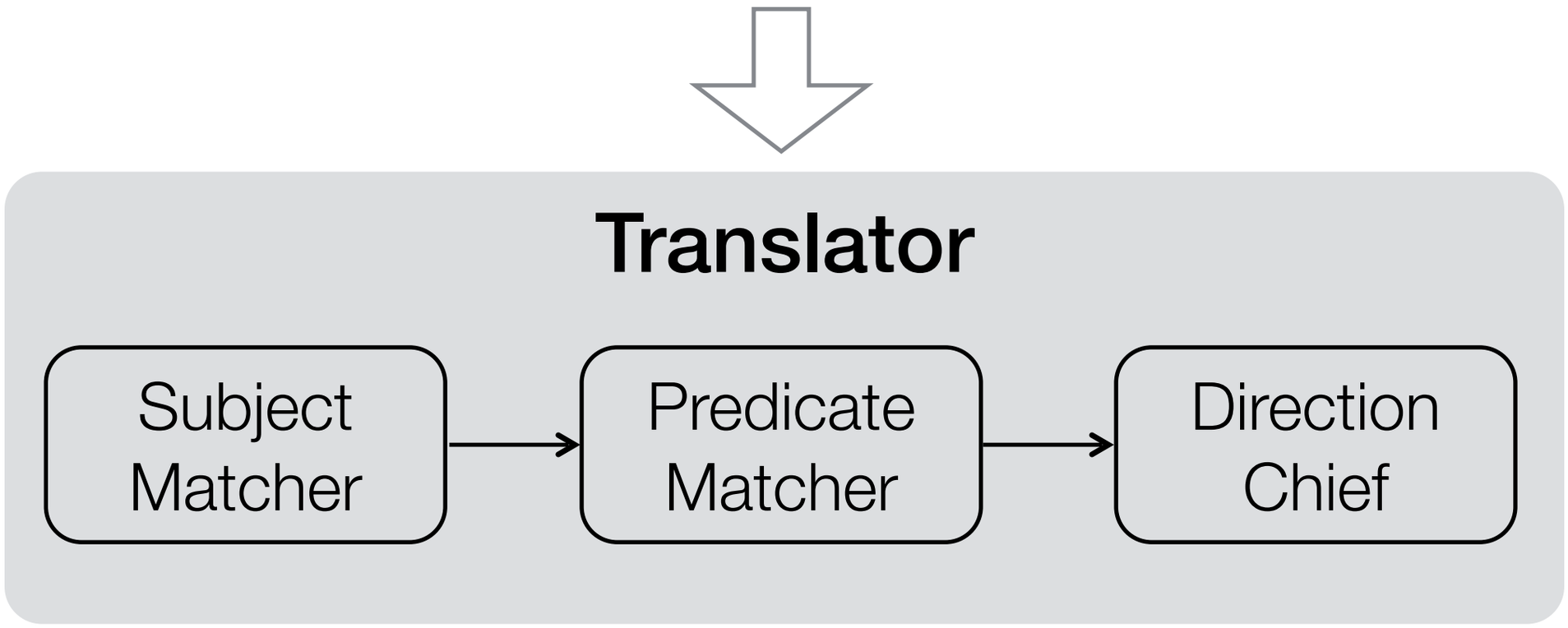


Translator



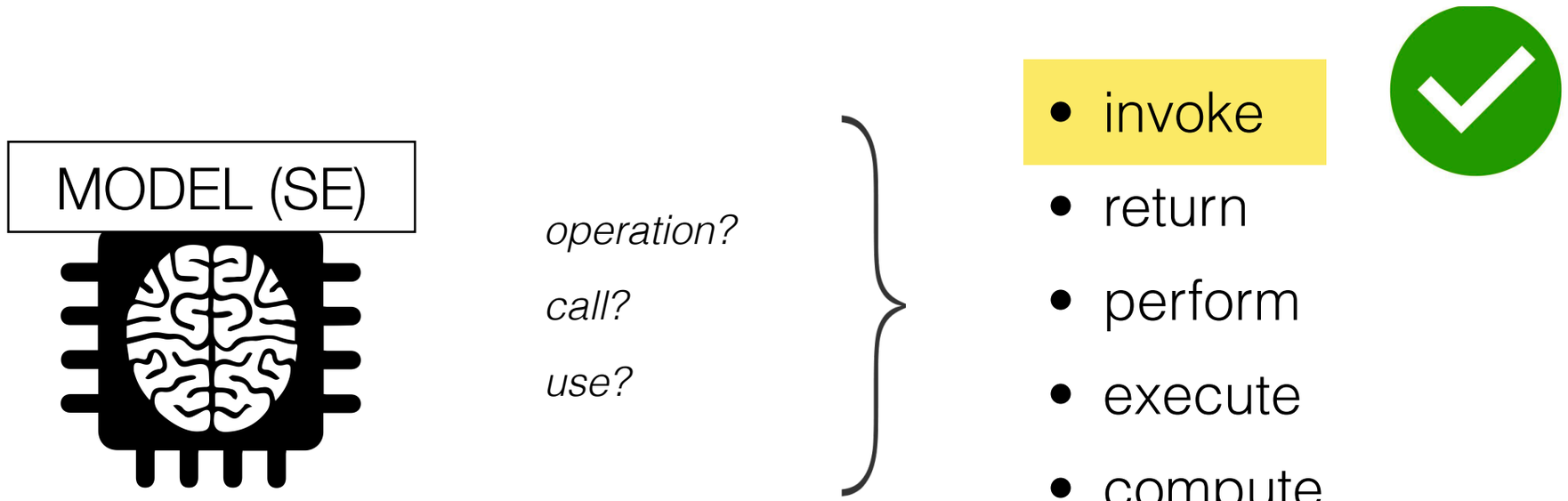
Temporal Specifications

(this method, be invoked) **BEFORE** (thread, is started)



Temporal Specifications

(this method, be **invoked**)



Efstathiou et al., *Word embeddings for the software engineering domain,* MSR 2018

(**this method**, be invoked)

~ **the documented method** should be invoked

```
receiverObjectID.setDaemon(args[0])
```

(**this method**, be invoked)

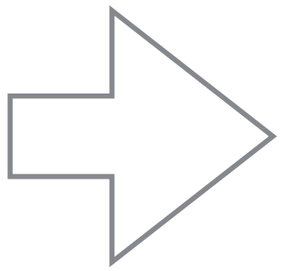
~ the documented method should be invoked

```
receiverObjectID.setDaemon(args[0])
```

(**thread**, is started)

Code Candidates

Formal Parameters
Class Name
Methods
Fields

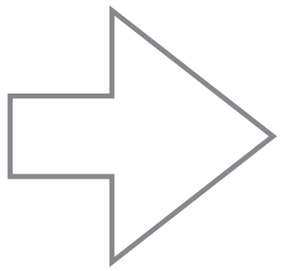


Candidate	Edit Distance
`on`	8
`Thread`	0
`start`	8
`sleep`	8
…	

(**thread**, is started)

Code Candidates

Formal Parameters
Class Name
Methods
Fields

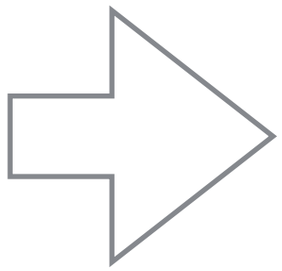


Candidate	Edit Distance
`on`	8
`Thread`	0
`start`	8
`sleep`	8
…	

(thread, is **started**)

Code Candidates

Formal Parameters
Class Name
Methods
Fields

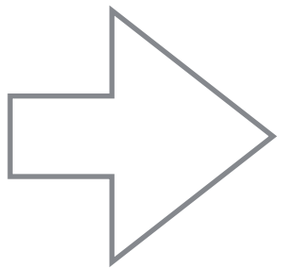


Candidate	Edit Distance
`on`	8
`Thread`	7
`start`	2
`sleep`	8
…	

(thread, is **started**)

Code Candidates

Formal Parameters
Class Name
Methods
Fields



Candidate	Edit Distance
`on`	8
`Thread`	7
`start`	2
`sleep`	8
...	

(thread, is started)

Code Candidates

Formal Parameters
Class Name
Methods
Fields

Candidate	Edit Distance
on	8
Thread	7
start	2
sleep	8
...	

```
receiverObjectID.start()
```


```
receiverObjectID.setDaemon(args[0])
```

BEFORE

```
receiverObjectID.start()
```


```
receiverObjectID.setDaemon(args[0])
```

BEFORE

```
receiverObjectID.start()
```

Sentence tense?

Previous observed examples?


```
receiverObjectID.setDaemon(args[0])
```

BEFORE

```
receiverObjectID.start()
```

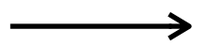
Modifier	Direction
BEFORE	⟶
PRIOR	⟶
(NOT) UNTIL	⟵
AFTER	⟵
ONCE	⟵


```
receiverObjectID.setDaemon(args[0])
```

BEFORE

```
receiverObjectID.start()
```

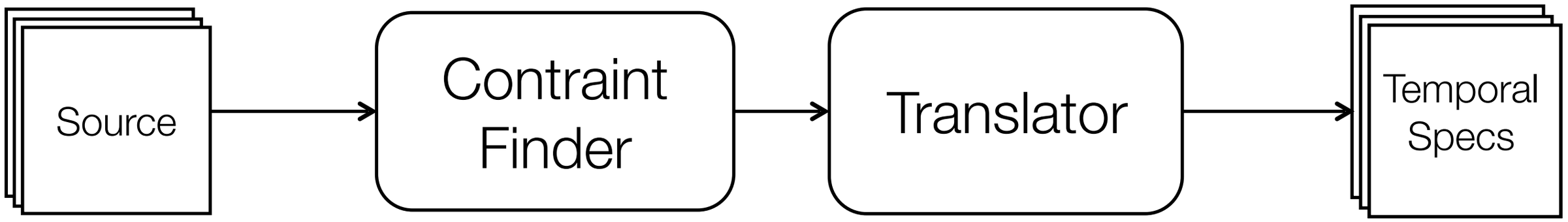
Modifier	Direction
BEFORE	⟶
PRIOR	⟶
(NOT) UNTIL	⟵
AFTER	⟵
ONCE	⟵



`setDaemon(args[0])`

```
[
  {
    "operationSignature": "setDaemon",
    "isConstructor": false,
    "mustPrecede": “receiverObjectID.start()",
    "mustFollow": "",
  }
]
```

CallMeMaybe




```java
/**
*  This method must be invoked before
   the thread
*  is started.
*/

public final void setDaemon(boolean on)
```


```
[
  {
    "operationSignature": "setDaemon",
    "isConstructor": false,
    "mustPrecede": “receiverObjectID.start()",
    "mustFollow": "",
  }
]
```

Evaluating CallMeMaybe

1. Can it **accurately** translate natural language to temporal specification?

2. Can it **improve** automatic test case generation?

Experimental setup

7 Popular Java systems

Experimental setup

89 Manually-written Java translations

Translation accuracy

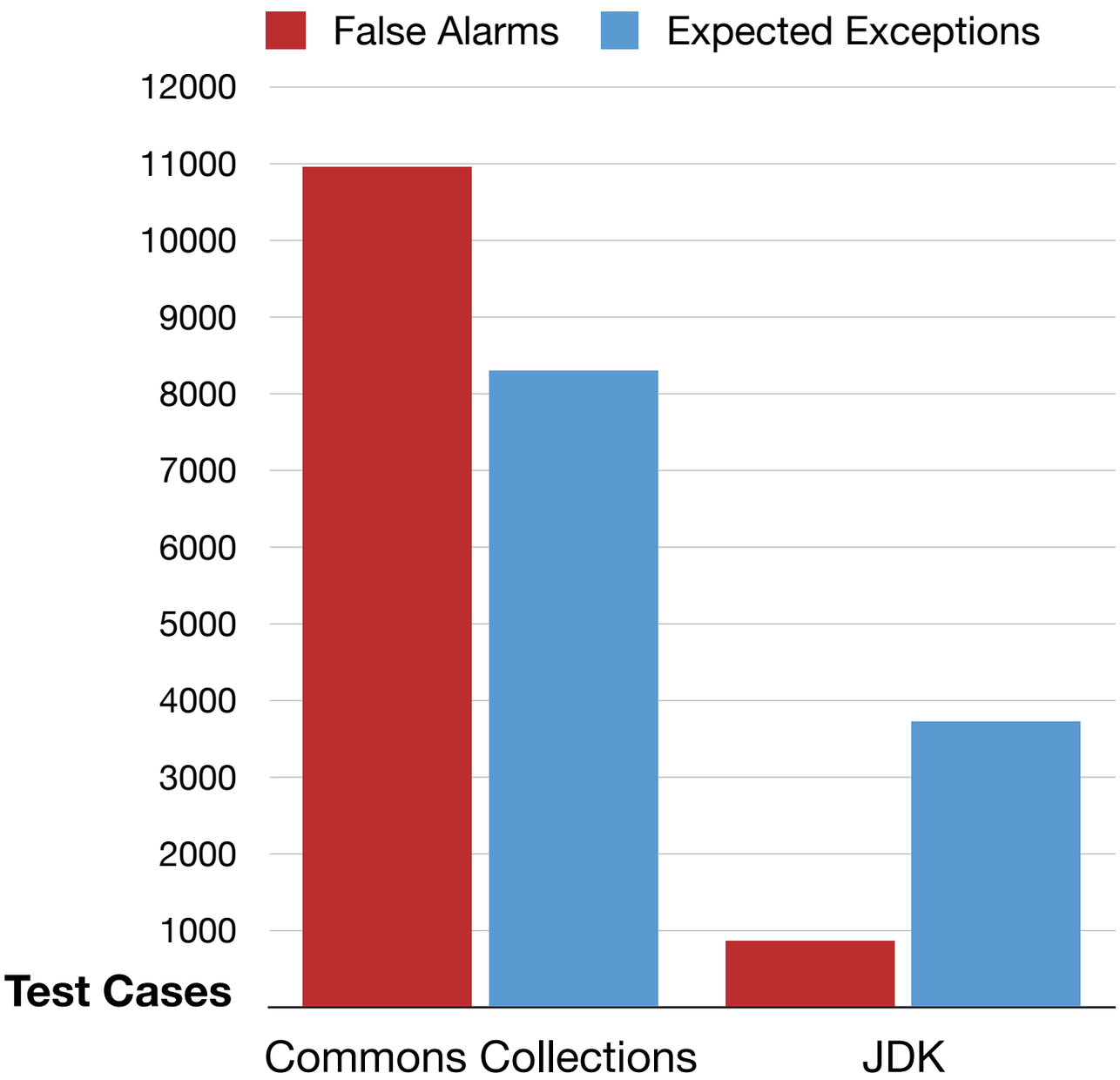
83 %

Precision

70 %

Recall

Improving Randoop



Randoop

Automatic unit test generation for Java

CallMeMaybe improves Random output:

Reveals **False Alarms**

in error test cases

Gives reasons for **Expected Exceptions**

in regression test cases

Improving Randoop

Reasons for **Expected Exceptions**
in regression test cases

Randoop
Automatic unit test generation for Java


```
@Test
public void test001(){
IteratorChain<java.io.Serializable> serializableIt0 = new
    IteratorChain<java.io.Serializable>(. . .);
        // The following exception was thrown during execution
        try {
            serializableIt0.remove();
            org.junit.Assert.fail("Expected exception
                java.lang.IllegalStateException; message:
                Iterator contains no elements");
        } catch (java.lang.IllegalStateException e) {
            // Expected exception.
            /* Violated CMM Constraint confirms this:
                 "You will normally use addIterator(Iterator)
                 to add some iterators after using this constructor." */
        }
```

Improving Randoop

False Alarms
in error-revealing test cases

Randoop
Automatic unit test generation for Java

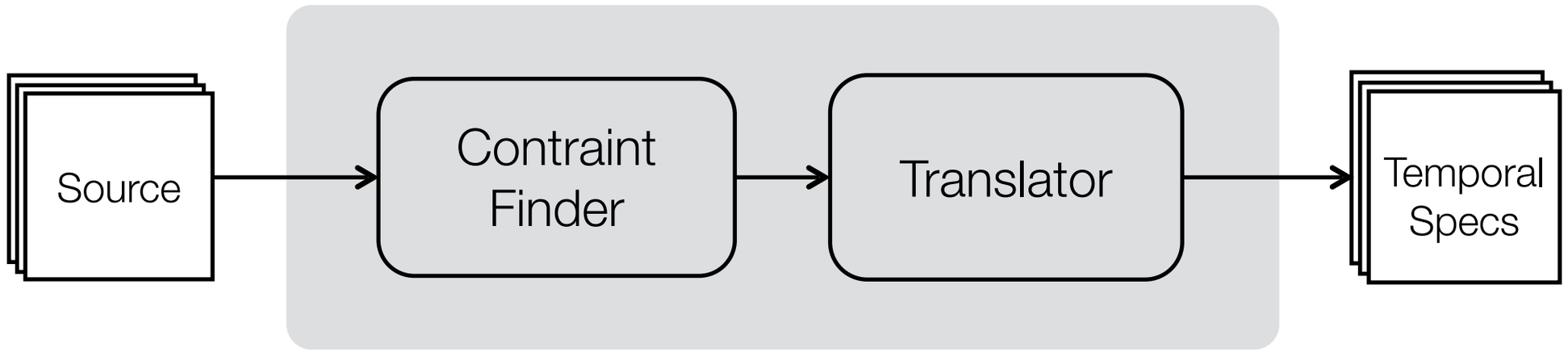
THE APACHE SOFTWARE FOUNDATION

```
@Test
public void test001(){
   IteratorEnumeration<String> strItEn0 =
            new IteratorEnumeration<String>();

   /* during test generation this statement threw an exception of type
      java.lang.NullPointerException in error

      But CMM Constraint was violated:
         "Constructs a new IteratorEnumeration that will not function
          until setIterator(Iterator) is invoked." */
   String str1 = strItEn0.nextElement();
}
```

Conclusions



- CallMeMaybe translates **natural language temporal** information into machine-readable specifications

- Its translations are **accurate** (83% precision ; 70% recall)

- Translations improve automatic test case generation by **revealing false alarms** and **explaining** exceptions

https://github.com/ariannab/callmemaybe