Code Duplication

Yegor Bugayenko

Lecture #11 out of 24
80 minutes

Motivating Example

Before (wrong):

```
printf("Hi,%s!",getName(42));
printf("Hi,%s!",getName(7));
printf("Hi,%s!",getName(55));
```

After (right):

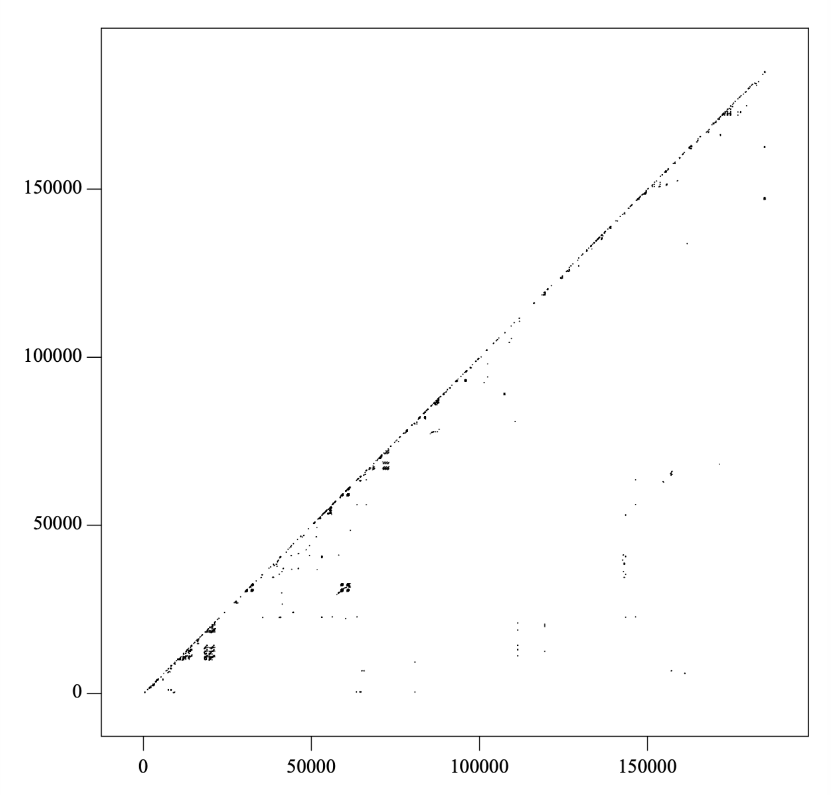
```
sayHello(42);
sayHello(7);
sayHello(556);

void sayHello(int id) {
  var n = getName(id);
  printf("Hi,%s!",n);
}
```

+2pt

"Two lines of code are considered to be identical if they contain the same sequence of characters after removing comments and white space; the semantics of the program statements are not analyzed."

— *A Program for Identifying Duplicated Code*, Brenda S. Baker, Computing Science and Statistics, 1993

Up to 38% of lines are involved in duplicates



The plots are dense near the main diagonal, implying that most copies tend to occur fairly locally, e.g. within the same file or module.

However, certain line segments occur away from the main diagonal; it would be interesting to investigate why the corresponding sections of code are duplicated.

Don't Repeat Yourself (DRY)

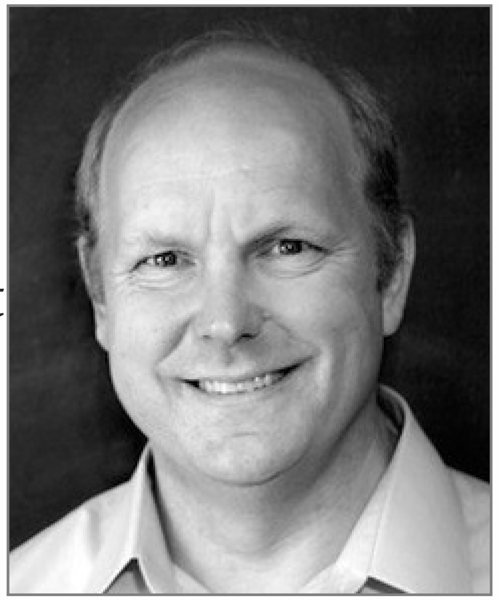
+2pt

"Every piece of knowledge must have a single, unambiguous, authoritative representation within a system."

— *The Pragmatic Programmer: From Journeyman to Master*, Andrew Hunt and David Thomas, Addison-Wesley, 1999

The Rule of Three

+2pt



"The first time you do something, you just do it. The second time you do something similar, you wince at the duplication, but you do the duplicate thing anyway. The third time you do something similar, you refactor."

— *Refactoring*, Martin Fowler and Kent Beck, Addison-Wesley, 1999

These tools can help detecting duplicate code:

1. ...

Read this:

A Program for Identifying Duplicated Code, Brenda S. Baker, Computing Science and Statistics, 1993

The Pragmatic Programmer: From Journeyman to Master, Andrew Hunt and David Thomas, Addison-Wesley, 1999 *Refactoring*, Martin Fowler and Kent Beck, Addison-Wesley, 1999