# Identifying Undesired Interaction between Variables Names

Irfan Ul Haq, Juan Caballero
IMDEA Software Institute, Madrid, Spain
[irfanul.haq, juan.caballero]@imdea.org

Michael D. Ernst
University of Washington , Washington, USA
mernst@cs.washington.edu

**Motivation:**
Sometimes, programmers make a mistake such as storing euros in a variable that should hold dollars, or using a file descriptor as an index into an array. Can we automatically identify such mistakes? Can we use semantics to identify undesired interactions automatically?

**Approach:**
The idea is to record variables which interact with each other in a group called ATI Cluster, and compare each variable name for semantic similarity with all the other variables names in same group. For example, in synthetic example, ‘itemPrice’ and ‘tax' depict monetary value and ‘miles’ shows distance concept. Hence, we want to predict if ‘miles' is semantically similar with ‘tax' and ‘itemPrice’.

**Synthetic Example:**

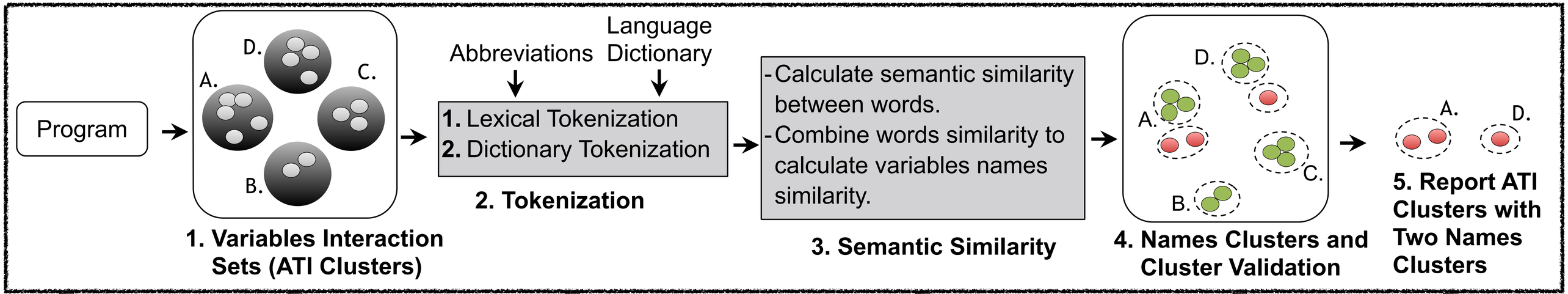
```
int tax = 0;
if (miles > 1000) {
        tax = itemPrice + miles;
}
```

Semantics/ Intention?

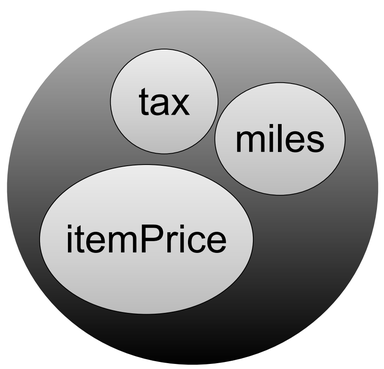
tax, itemPrice — Money

miles — Distance

*Example taken from “Dynamic Inference of Abstract Types” by Michael et al.




Synthetic Example




‘tax’ = [‘tax’]
‘itemPrice’ = [‘item’, ‘Price’]
‘miles' = [‘miles’]

| | tax | itemPrice | miles |
|---|---|---|---|
| **tax** | 1 | 0.27 | 0.07 |
| **itemPrice** | 0.27 | 1 | 0.04 |
| **miles** | 0.07 | 0.04 | 1 |