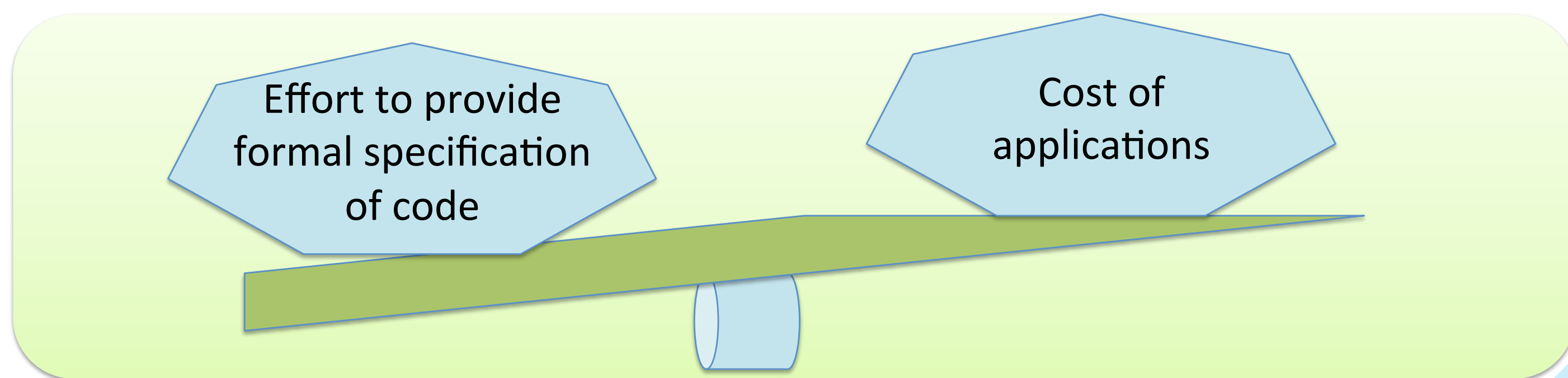


# Inferring Behavioral Specifications from Large-scale Repositories by Leveraging Collective Intelligence

Hridesh Rajan: hridesh@iastate.edu  
Gary T. Leavens: leavens@ecs.ucf.edu

Tien N. Nguyen: tien@iastate.edu  
Robert Dyer: rdyer@bgsu.edu

## Problem: Insufficient Code Specification and Cost Effectiveness of Specifying Applications



### Behavioral Interface Specifications:

writing non-trivial specification requires significant time and expertise.

### Problem 1, Insufficient Specification of Core Libraries:

Frequently used existing libraries lack formal specifications.

### Problem 2, Cost of Applications:

Specification of Applications depending on those libraries become highly expensive.

## Solution: A vision for Inference of Behavioral Interface Specifications

The overall problem of inferring specifications needs a three pronged approach by identifying widely used codes, similar rare codes and non-similar rare codes.

### Consensus-based Inference:

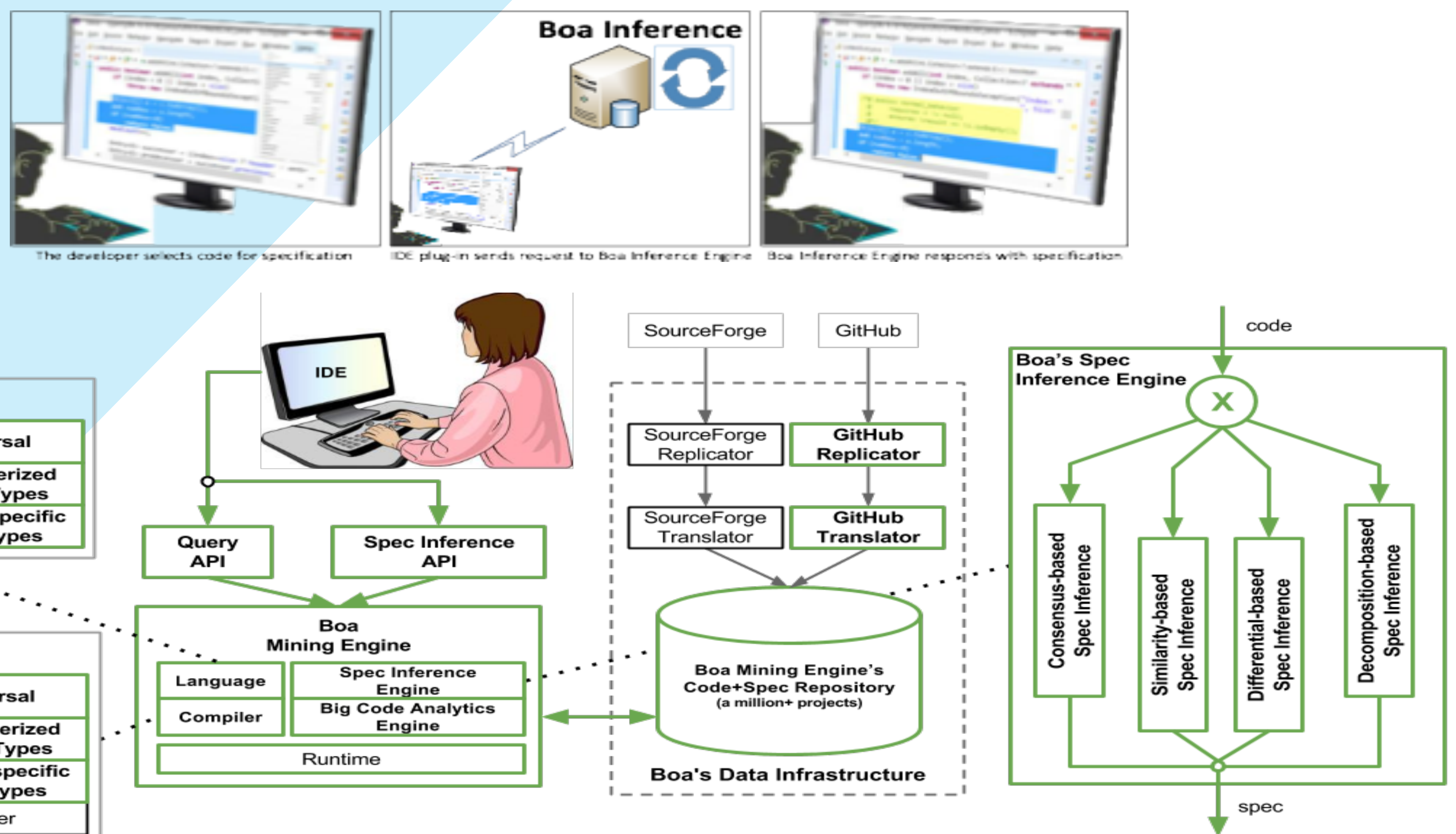
most clients for widely used code in big code work, broken client tend to get fixed eventually

### Similarity and Differential-based Inference:

Programmers reuse and clone good patterns of program design; similar code should have similar specifications

### Decomposition-based Inference:

function signatures in widely used codes can help quickly to separate (code, specification) pairs



## Publications

1. H. Rajan, T. N. Nguyen, G. T. Leavens and R. Dyer, "Inferring behavioral specifications from large-scale repositories by leveraging collective intelligence", in Proceedings of the 37th International Conference on Software Engineering, ser. ICSE'15, 2015.
2. H. A. Nguyen, R. Dyer, T. N. Nguyen, and H. Rajan, "Mining preconditions of APIs in large-scale code corpus," in 22nd International Symposium on Foundations of Software Engineering, ser. FSE'14, November 2014, pp. 166–177
3. R. Dyer, H. Rajan, H. A. Nguyen, and T. N. Nguyen, "Mining billions of AST nodes to study actual and potential usage of Java language features," in Proceedings of the 36th International Conference on Software Engineering, ser. ICSE'14, 2014.
4. R. Dyer, H. A. Nguyen, H. Rajan, and T. N. Nguyen, "Boa: A language and infrastructure for analyzing ultra-large-scale software repositories," in Proceedings of the 35th International Conference on Software Engineering, ser. ICSE'13, 2013, pp. 422–431.

The research and educational activities described on this poster has been supported in part by the US National Science Foundation (NSF) under grants 1513263, 1512947, 1423370, 1349153, 1117937, 1017334 and 0846059.

IOWA STATE UNIVERSITY  
OF SCIENCE AND TECHNOLOGY

BGSU®

Bowling Green State University



UNIVERSITY OF CENTRAL FLORIDA

Laboratory of Software Design  
<http://boa.cs.iastate.edu>