Analysis of developer expertise of APIs

Hakan Aksu

Ralf Lämmel
Software Languages Team
University of Koblenz-Landau

Preamble

- MSc thesis
- Initial stage – work in progress

Motivation

- Many software projects in IT companies
- They use:
  - Various languages
  - Various technologies
  - Various problem Domains
  - …

Motivation

- BUT how does an executive or project manager know which skills a developer has (when hiring or assigning)?
  - Interviews
  - Questionnaires
  - Assignments
  - Publicly available Information (e.g. on topcoder or on stackoverflow)

→ „problematic“ methods

Objective

- A new technique to determine the developer skills
  - Leverages previous work experience of developers in a systematic manner
  - We analyze existing evidence for developer expertise based on the version history of existing projects

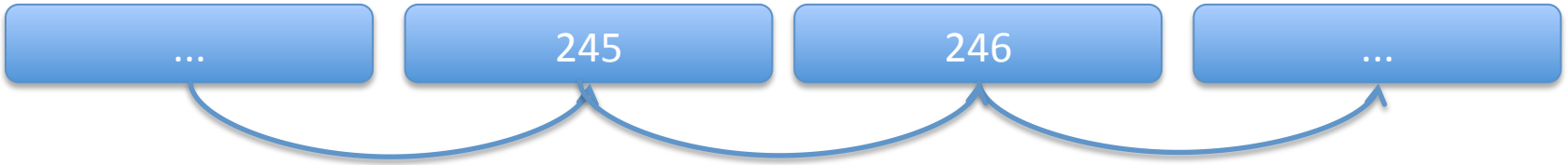
Milestones (1)

We review related work and best practices of MSR (mining software repositories)

** Chaturvedi, K.K., Singh, V.B., Singh, P.: Tools in Mining **
** Software Repositories. In: ICCSA (6). pp. 89–98. IEEE (2013) **

to agree on methods for:

- processing version history



- discovering traceability links between commits, code, and developers.

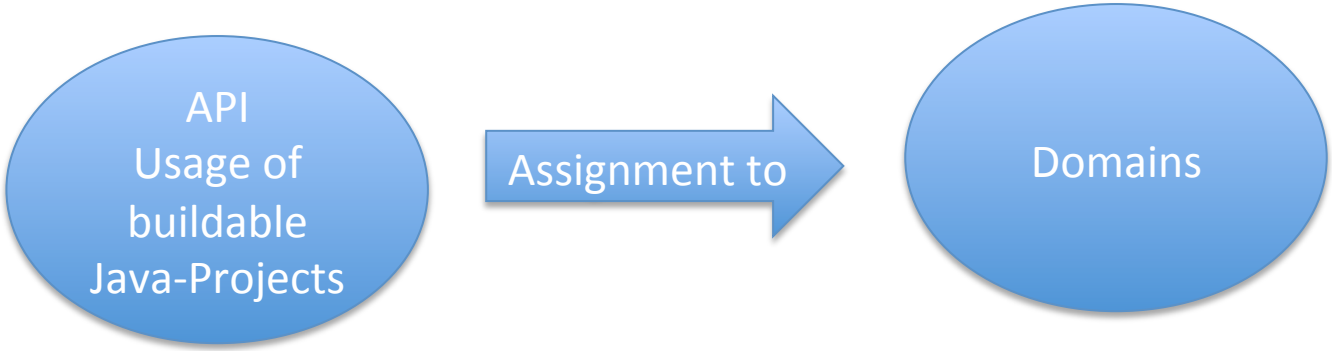
```
+ import java.io.FileReader;
+ import java.io.BufferedReader;
...
+ BufferedReader brTrace;
+ FileReader fReader;
...
+ fReader = new FileReader(...);
+ brTrace=new BufferedReader(fReader);
...
```

Milestone (2)

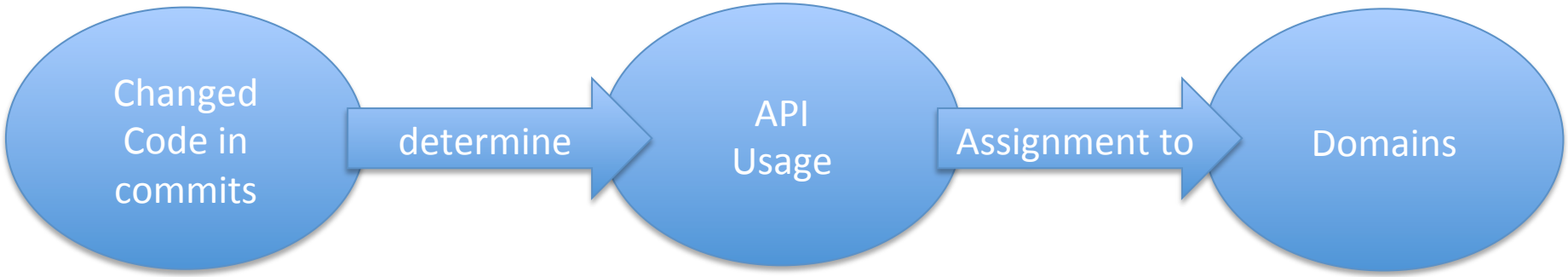
We leverage our prior work on API usage analysis and more related work to translate code changes into API usage data.

Prior work:

Roover, C.D., Lämmel, R., Pek, E.: Multi-dimensional exploration of API usage. In: ICPC. pp. 152–161. IEEE (2013)



New Challange:



Milestone (3)

We leverage best practices on corpus usage and engineering in MSR to select suitable open-source projects as the corpus to be used in our research.

Challenge: the analysis cannot generally assume all versions to be buildable (resolvable).

Pek, E.: **Corpus-based Empirical Research in Software Engineering**. Ph.D. thesis, University of Koblenz-Landau, Department of Computer Science (2014), available online at http://softlang.uni-koblenz.de/PekThesis.pdf

Tempero, E.D., Anslow, C., Dietrich, J., Han, T., Li, J., Lumpe, M., Melton, H., Noble, J.: **The Qualitas Corpus: A Curated Collection of Java Code for Empirical Studies**. In: APSEC. pp. 336–345. IEEE (2010)

Roover, C.D., Lämmel, R., Pek, E.: **Multi-dimensional exploration of API usage**. In: ICPC. pp. 152–161. IEEE (2013)

Milestone (4)

We identify techniques for

- Summarization and
- Visualization

To derive an

- Understandable and
- Informative

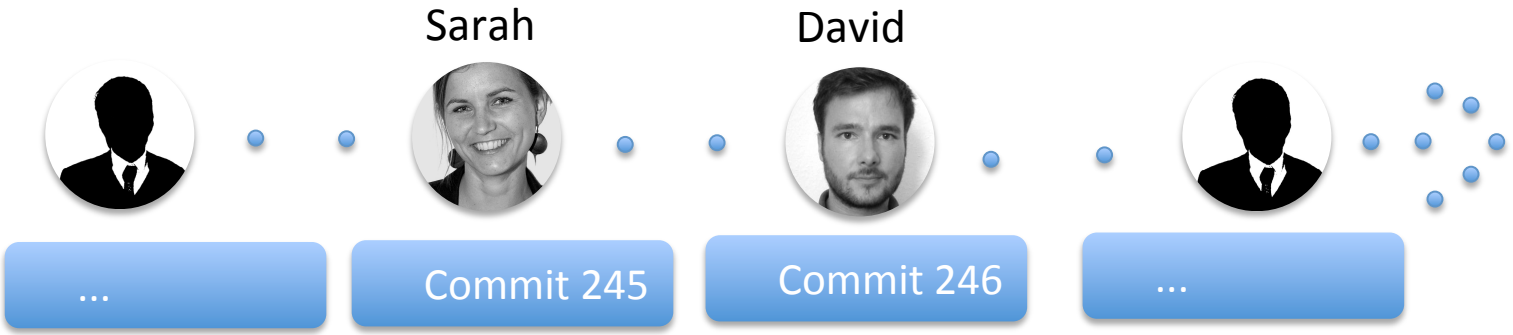
developer profile regarding

- API and
- domain expertise.

Conclusion

- Analysis of ...
  - API usage
  - Change along evolution
  - Developer activity

Conclusion




```
+ import java.io.FileReader;
+ import java.io.BufferedReader;
...
+ BufferedReader brTrace;
+ FileReader fReader;
...
+ fReader = new FileReader(...);
+ brTrace=new BufferedReader(fReader);
...
```

API: java.io

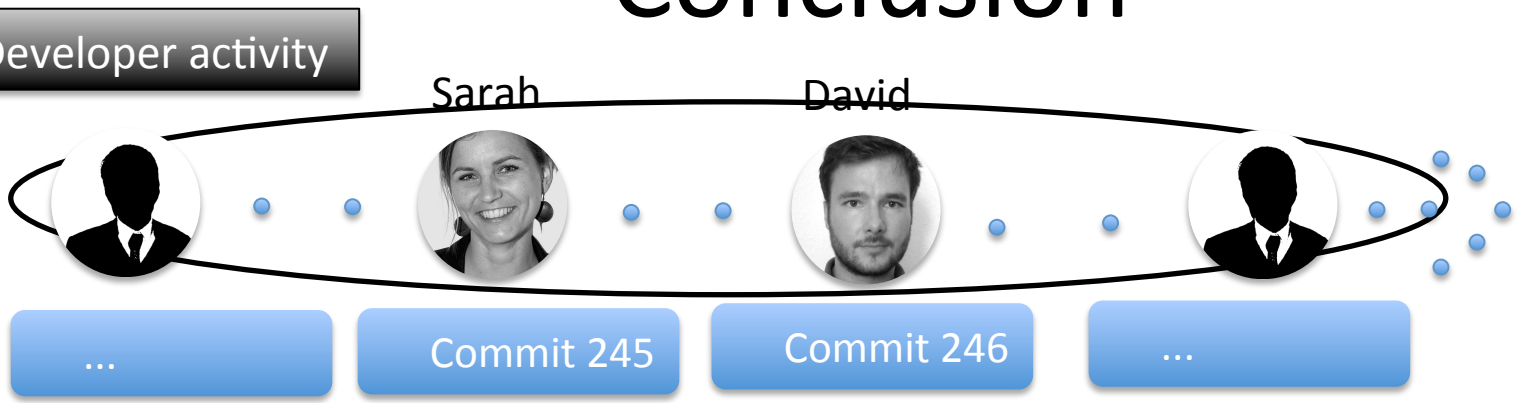
Domain: IO

```
+ import java.awt.*;
...
+ GridBagLayout grid = new GridBagLayout();
+ Button button;
...
+ import javax.swing.*;
...
+ JFrame f = new JFrame("Hello World");
+ f.setDefaultCloseOperation(JFrame.EXIT_ON_CLO
```

API: AWT & SWING

Domain: GUI

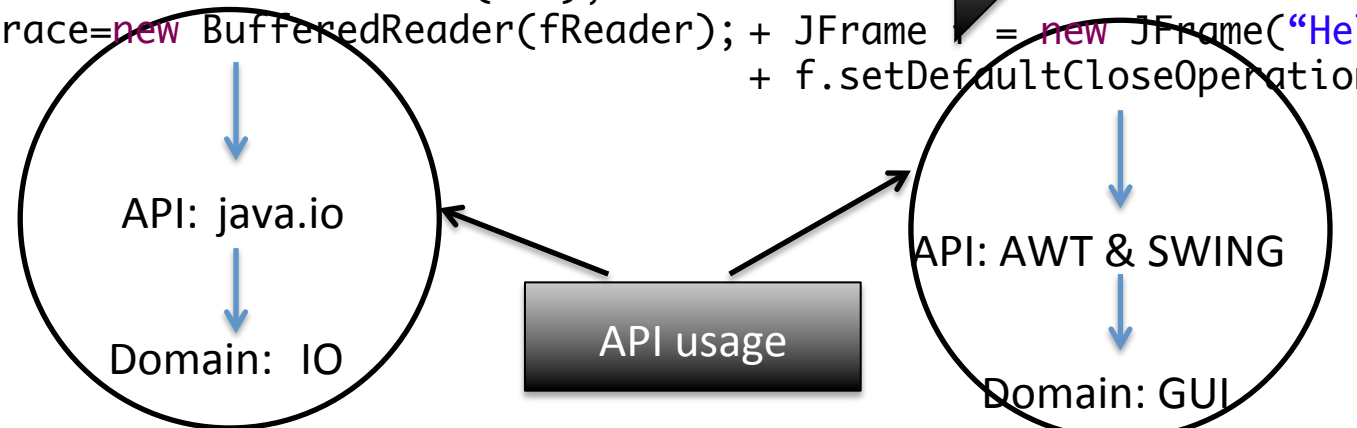
Conclusion




```
+ import java.io.FileReader;
+ import java.io.BufferedReader;
...
+ BufferedReader brTr
+ FileReader fReader;
...
+ fReader = new FileReader(...);
+ brTrace=new BufferedReader(fReader);
...
```

```
+ import java.awt.*;
...
+ GridBag yout grid = new GridBagLayout();
ton;
.swing.*;
...
+ JFrame = new JFrame("Hello World");
+ f.setDefaultCloseOperation(JFrame.EXIT_ON_CLO
```

Change along evolution



References

1. Canfora, G., Cerulo, L., Penta, M.D.: Identifying Changed Source Code Lines from Version Repositories. In: MSR. p. 14. IEEE (2007)

2. Chaturvedi, K.K., Singh, V.B., Singh, P.: Tools in Mining Software Repositories. In: ICCSA (6). pp. 89–98. IEEE (2013)

3. Lammel, R., Linke, R., Pek, E., Varanovich, A.: A Framework Profile of .NET. In: WCRE. pp. 141–150. IEEE (2011)

4. Linstead, E., Rigor, P., Bajracharya, S.K., Lopes, C.V., Baldi, P.: Mining Eclipse Developer Contributions via Author-Topic Models. In: MSR. p. 30. IEEE (2007)

5. Pek, E.: Corpus-based Empirical Research in Software Engineering. Ph.D. thesis, University of Koblenz-Landau, Department of Computer Science (2014), available online at http://softlang.uni-koblenz.de/PekThesis.pdf

6. Robbes, R.: Mining a Change-Based Software Repository. In: MSR. p. 15. IEEE (2007)

7. Roover, C.D., Lämmel, R., Pek, E.: Multi-dimensional exploration of API usage. In: ICPC. pp. 152–161. IEEE (2013)

8. Tempero, E.D., Anslow, C., Dietrich, J., Han, T., Li, J., Lumpe, M., Melton, H., Noble, J.: The Qualitas Corpus: A Curated Collection of Java Code for Empirical Studies. In: APSEC. pp. 336–345. IEEE (2010)

9. Wang, J., Dang, Y., Zhang, H., Chen, K., Xie, T., Zhang, D.: Mining succinct and high-coverage API usage patterns from source code. In: MSR. pp. 319–328. IEEE

(2013)

10. Xie, T., Pei, J.: MAPO: mining API usages from open source repositories. In: MSR.

pp. 54–57. IEEE (2006)

11. Yu, L., Ramaswamy, S.: Mining CVS Repositories to Understand Open-Source

Project Developer Roles. In: MSR. p. 8. IEEE (2007)

Questions?