



# Apache Flink

## Dependency Extraction

Group Small

# In Today's Presentation

- Technique 1
- Technique 2
- Live Demo
- Technique 3
- Comparison Process
- Qualitative Analysis
- Quantitative Analysis
- Rationale for Differences
- Lessons Learned
- References



# Technique 1: Understand

- Provided in A1
- Parsing source code
- Static analysis
- Dependency graphs

```
$INSTANCE flink-1.17.1/tools/ci/flink-ci-tools/src/main/java/org/apache/tools/ci/licensecheck/NoticeFileChecker.java cFile
$INSTANCE flink-1.17.1/tools/ci/flink-ci-tools/src/main/java/org/apache/tools/ci/suffixcheck/ScalaSuffixChecker.java cFile
$INSTANCE flink-1.17.1/tools/ci/flink-ci-tools/src/main/java/org/apache/tools/ci/utis/dependency/DependencyParser.java cFile
$INSTANCE flink-1.17.1/tools/ci/flink-ci-tools/src/main/java/org/apache/tools/ci/utis/deploy/DeployParser.java cFile
$INSTANCE flink-1.17.1/tools/ci/flink-ci-tools/src/main/java/org/apache/tools/ci/utis/notice/NoticeContents.java cFile
$INSTANCE flink-1.17.1/tools/ci/flink-ci-tools/src/main/java/org/apache/tools/ci/utis/notice/NoticeParser.java cFile
$INSTANCE flink-1.17.1/tools/ci/flink-ci-tools/src/main/java/org/apache/tools/ci/utis/shade/ShadeParser.java cFile
$INSTANCE flink-1.17.1/tools/ci/flink-ci-tools/src/main/java/org/apache/tools/ci/utis/shared/Dependency.java cFile
$INSTANCE flink-1.17.1/tools/ci/flink-ci-tools/src/main/java/org/apache/tools/ci/utis/shared/DependencyTree.java cFile
$INSTANCE flink-1.17.1/tools/ci/flink-ci-tools/src/main/java/org/apache/tools/ci/utis/shared/ParserUtils.java cFile
cLinks flink-1.17.1/flink-annotations/src/main/java/org/apache/FlinkVersion.java flink-1.17.1/flink-annotations/src/main/java/org/apache/annotation/Public.java
cLinks flink-1.17.1/flink-annotations/src/main/java/org/apache/annotation/Experimental.java flink-1.17.1/flink-annotations/src/main/java/org/apache/annotation/Public.java
cLinks flink-1.17.1/flink-annotations/src/main/java/org/apache/annotation/Internal.java flink-1.17.1/flink-annotations/src/main/java/org/apache/annotation/Public.java
cLinks flink-1.17.1/flink-annotations/src/main/java/org/apache/annotation/PublicEvolving.java flink-1.17.1/flink-annotations/src/main/java/org/apache/annotation/Public.java
cLinks flink-1.17.1/flink-annotations/src/main/java/org/apache/annotation/docs/ConfigGroup.java flink-1.17.1/flink-annotations/src/main/java/org/apache/annotation/Internal.java
cLinks flink-1.17.1/flink-annotations/src/main/java/org/apache/annotation/docs/ConfigGroups.java flink-1.17.1/flink-annotations/src/main/java/org/apache/annotation/Internal.java
cLinks flink-1.17.1/flink-annotations/src/main/java/org/apache/annotation/docs/ConfigGroups.java flink-1.17.1/flink-annotations/src/main/java/org/apache/annotation/docs/ConfigGroup.java
cLinks flink-1.17.1/flink-annotations/src/main/java/org/apache/annotation/docs/Documentation.java flink-1.17.1/flink-annotations/src/main/java/org/apache/annotation/Internal.java
cLinks flink-1.17.1/flink-clients/src/main/java/org/apache/client/ClientUtils.java flink-1.17.1/flink-clients/src/main/java/org/apache/client/cli/ClientOptions.java
cLinks flink-1.17.1/flink-clients/src/main/java/org/apache/client/ClientUtils.java flink-1.17.1/flink-clients/src/main/java/org/apache/client/program/ContextEnvironment.java
cLinks flink-1.17.1/flink-clients/src/main/java/org/apache/client/ClientUtils.java flink-1.17.1/flink-clients/src/main/java/org/apache/client/program/PackageProgram.java
cLinks flink-1.17.1/flink-clients/src/main/java/org/apache/client/ClientUtils.java flink-1.17.1/flink-clients/src/main/java/org/apache/client/program/ProgramInvocationException.java
```

## Technique 2: Include

- Scripts
- Traverse source code folder
- Record include statements (import)

```
$INSTANCE flink-1.17.1/tools/ci/flink-ci-tools/src/test/java/org/apache/flink/tools/ci/licensecheck/JarFileCheckerTest.java cFile
$INSTANCE flink-1.17.1/tools/ci/flink-ci-tools/src/test/java/org/apache/flink/tools/ci/licensecheck/NoticeFileCheckerTest.java cFile
$INSTANCE flink-1.17.1/tools/ci/flink-ci-tools/src/test/java/org/apache/flink/tools/ci/utils/dependency/DependencyParserCopyTest.java cFile
$INSTANCE flink-1.17.1/tools/ci/flink-ci-tools/src/test/java/org/apache/flink/tools/ci/utils/dependency/DependencyParserTreeTest.java cFile
$INSTANCE flink-1.17.1/tools/ci/flink-ci-tools/src/test/java/org/apache/flink/tools/ci/utils/deploy/DeployParserTest.java cFile
$INSTANCE flink-1.17.1/tools/ci/flink-ci-tools/src/test/java/org/apache/flink/tools/ci/utils/notice/NoticeParserTest.java cFile
$INSTANCE flink-1.17.1/tools/ci/flink-ci-tools/src/test/java/org/apache/flink/tools/ci/utils/shade/ShadeParserTest.java cFile
$INSTANCE flink-1.17.1/tools/ci/flink-ci-tools/src/test/java/org/apache/flink/tools/ci/utils/shared/DependencyTreeTest.java cFile
dependency flink-1.17.1/flink-annotations/src/main/java/org/apache/flink/FlinkVersion.java DEPENDS_ON flink.annotation.Public
dependency flink-1.17.1/flink-annotations/src/main/java/org/apache/flink/annotation/docs/ConfigGroup.java DEPENDS_ON flink.annotation.Internal
dependency flink-1.17.1/flink-annotations/src/main/java/org/apache/flink/annotation/docs/ConfigGroups.java DEPENDS_ON flink.annotation.Internal
dependency flink-1.17.1/flink-annotations/src/main/java/org/apache/flink/annotation/docs/Documentation.java DEPENDS_ON flink.annotation.Internal
dependency flink-1.17.1/flink-clients/src/main/java/org/apache/flink/client/ClientUtils.java DEPENDS_ON flink.api.common.JobID
dependency flink-1.17.1/flink-clients/src/main/java/org/apache/flink/client/ClientUtils.java DEPENDS_ON flink.api.common.JobStatus
dependency flink-1.17.1/flink-clients/src/main/java/org/apache/flink/client/ClientUtils.java DEPENDS_ON flink.client.cli.ClientOptions
```



# Include Process

- Traverse source code folder
  - Read each .java file
  - Search for import statements
  - Record file name and its import statements
- 
- Convert to TA format
  - Reorganize to match Understand file structure

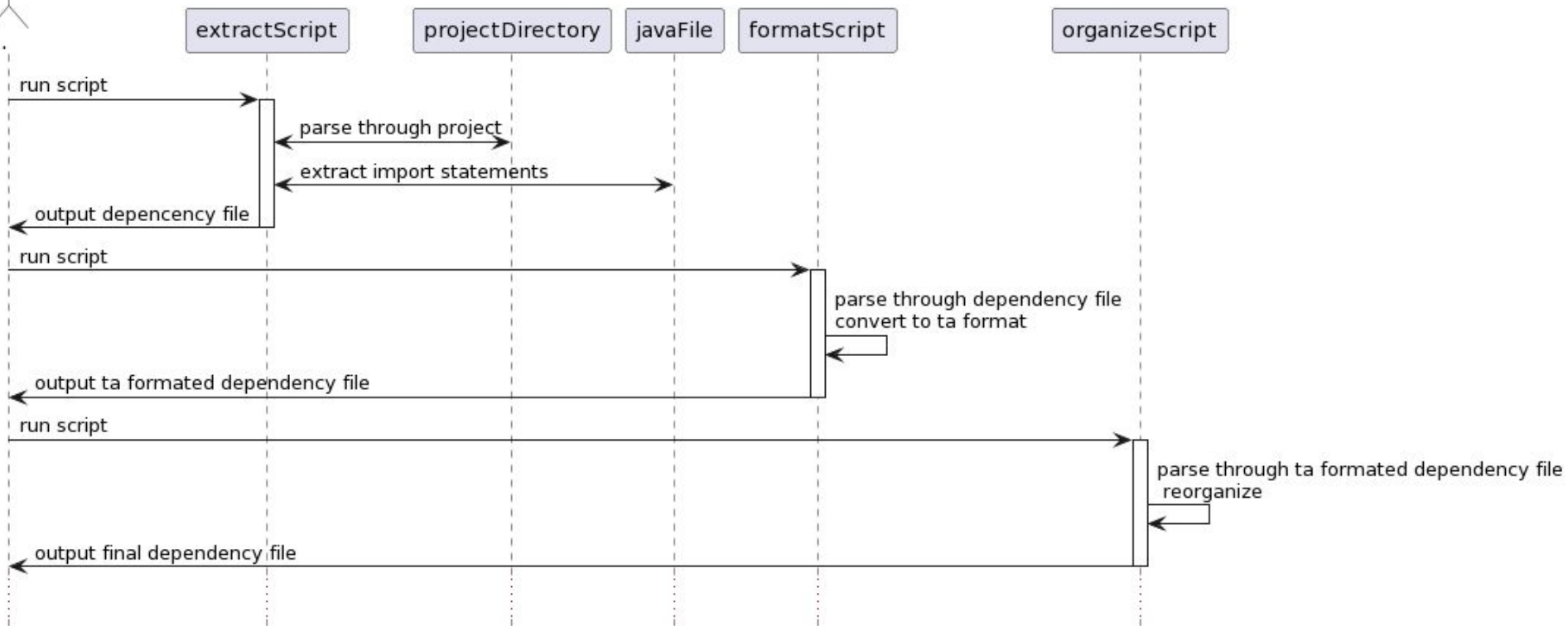
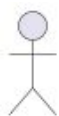
```
import os
import re

def extractDependencies(flink):
    flinkDependencies = {}

    pattern = re.compile(r'^\s*import\s+org\.apache\.([\w.]+)\s*;'')

    for root, __, files in os.walk(flink):
        for file in files:
            if file.endswith('.java'):
                filePath = os.path.join(root, file)
                relativePath = filePath.replace(flink, '').lstrip(os.path.sep)
                with open(filePath, 'r', encoding='utf-8') as java_file:
                    dependencies = []
                    lines = java_file.readlines()
                    for line in lines:
                        match = pattern.match(line)
                        if match:
                            dependency = match.group(1)
                            dependencies.append(dependency)
                    flinkDependencies[relativePath] = dependencies

    return flinkDependencies
```



Include Live Demo

# Technique 3: Issues

- srcML
  - Conversion worked
  - Could not read xml
  - No output
- JDT
  - IDE was not recognizing JDT as a dependency
  - JDT installed through eclipse market
  - JDT installed manually
  - AST Parser could not be resolved
- JavaParser
  - Maven could not find dependency



# Technique 3: Theory

- Convert source code to XML
- Trace method invocation
- Trace new keywords
- Look for matching method definitions

```
<package>package <name><name>org</name></operator>.</operator><name>apache</name></operator>.</operator><name>flink</name></operator>.</operator><name>runtime</name></operator>.</operator></package>

<import>import <name><name>org</name></operator>.</operator><name>apache</name></operator>.</operator><name>flink</name></operator>.</operator><name>runtime</name></operator>.</operator><name>jobgraph</name></operator>.</operator><name>OperatorID</name></operator>.</operator></import>

<import>import <name><name>javax</name></operator>.</operator><name>annotation</name></operator>.</operator><name>Nullable</name></operator>.</operator></import>

<import>import <name><name>java</name></operator>.</operator><name>io</name></operator>.</operator><name>Serializable</name></operator>.</operator></import>
<import>import <name><name>java</name></operator>.</operator><name>util</name></operator>.</operator><name>Optional</name></operator>.</operator></import>

<comment type="block" format="javadoc">/** Formed of a mandatory operator ID and optionally a user defined operator ID. */</comment>
<class><specifier>public</specifier> class <name>OperatorIDPair</name> <super_list><implements>implements <super><name>Serializable</name></super></implements></super_list> <block>{

    <decl_stmt><decl><type><specifier>private</specifier> <specifier>static</specifier> <specifier>final</specifier> <name>long</name></type> <name>serialVersionUID</name> <init>= <expr><literal type="number">1L</literal></expr></init></decl></decl_stmt>

    <decl_stmt><decl><type><specifier>private</specifier> <specifier>final</specifier> <name>OperatorID</name></type> <name>generatedOperatorID</name></decl></decl_stmt>
    <decl_stmt><decl><type><specifier>private</specifier> <specifier>final</specifier> <name>OperatorID</name></type> <name>userDefinedOperatorID</name></decl></decl_stmt>

    <constructor><specifier>private</specifier> <name>OperatorIDPair</name>(<parameter_list>{
        <parameter><decl><type><name>OperatorID</name></type> <name>generatedOperatorID</name></decl></parameter>; <parameter><decl><type><annotation>@<name>Nullable</name></annotation> <name>OperatorID</name></type> <name>userDefinedOperatorID</name></decl></parameter>}</parameter_list>
        <expr_stmt><expr><name>this</name></operator>.</operator><name>generatedOperatorID</name></operator> <operator>=</operator> <name>generatedOperatorID</name></expr></expr_stmt>
        <expr_stmt><expr><name>this</name></operator>.</operator><name>userDefinedOperatorID</name></operator> <operator>=</operator> <name>userDefinedOperatorID</name></expr></expr_stmt>
    }</block_content></block></constructor>

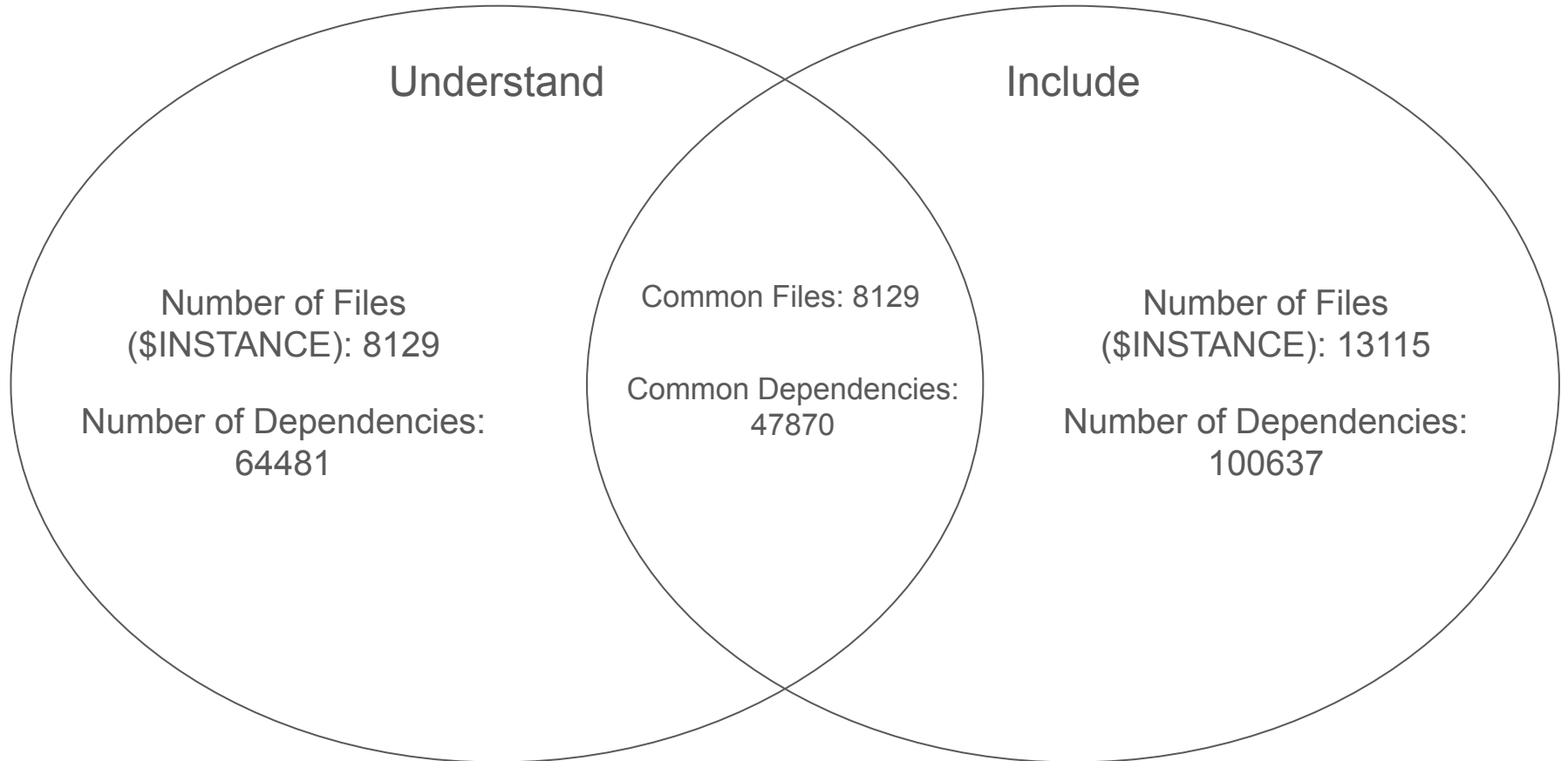
    <function><type><specifier>public</specifier> <specifier>static</specifier> <name>OperatorIDPair</name></type> <name>of</name>(<parameter_list>{
        <parameter><decl><type><name>OperatorID</name></type> <name>generatedOperatorID</name></decl></parameter>; <parameter><decl><type><annotation>@<name>Nullable</name></annotation> <name>OperatorID</name></type> <name>userDefinedOperatorID</name></decl></parameter>}</parameter_list>
        <return>return <expr><operator>new</operator> <call><name>OperatorIDPair</name></argument_list>{<argument><expr><name>generatedOperatorID</name></expr></argument>; <argument><expr><name>userDefinedOperatorID</name></expr></argument>}</argument_list></call></expr></return>
    }</block_content></block></function>
```

# Comparison Process

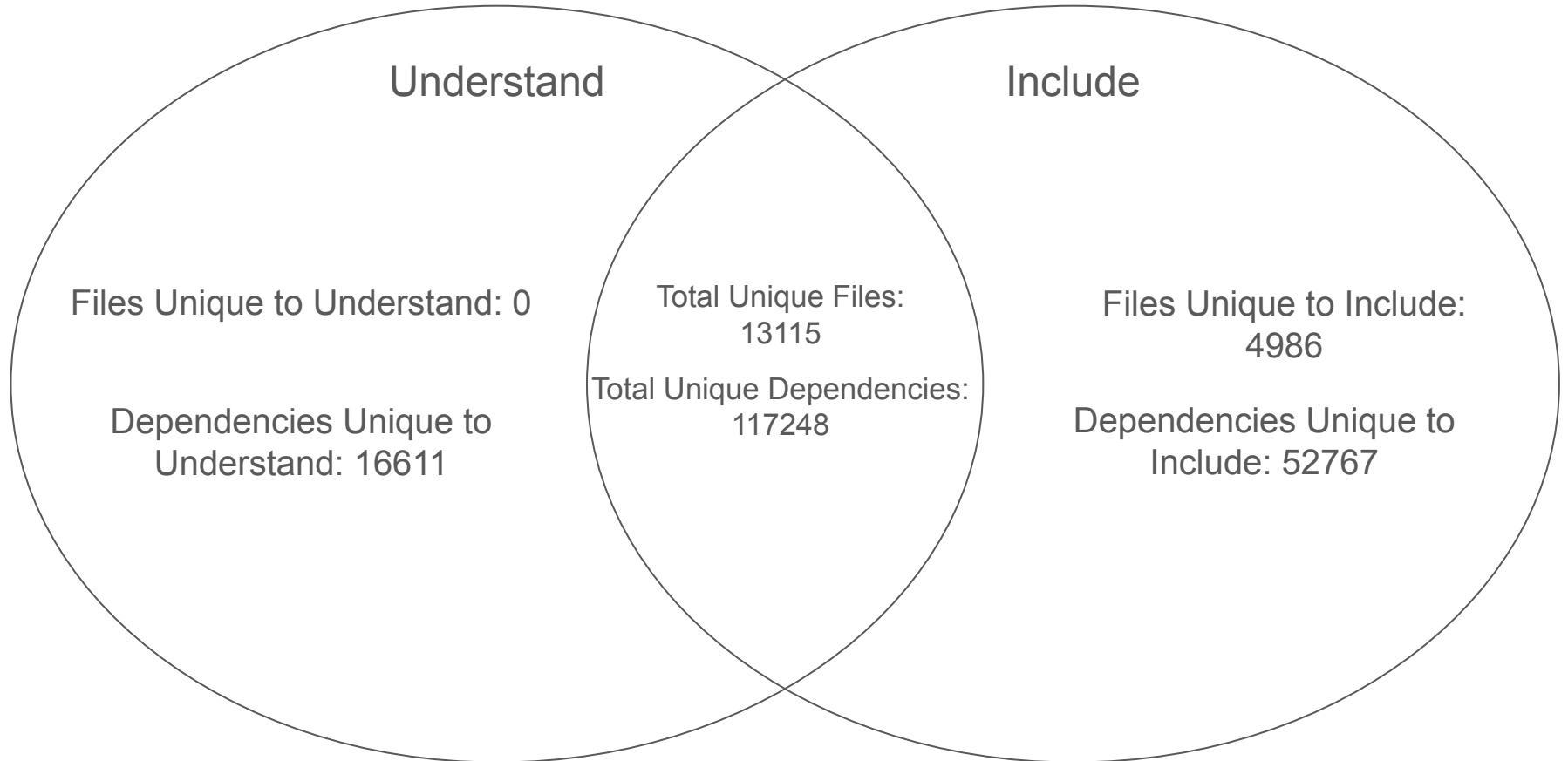
- Format include dependency file to match TA format
- Script to compare the two methods
- Manual analysis of discrepancies

```
> def understandCount(filePath): ...  
  
> def includeCount(filePath): ...  
  
> def understandInstances(understandFile): ...  
  
> def includeInstances(includeFile): ...  
  
> def commonFiles(understandFile, includeFile): ...  
  
> def commonDependencies(understandFile, includeFile): ...  
  
> def extractSmamples(understandFile, includeFile): ...  
  
understandFile = 'UnderstandDependencyFile.raw.ta.txt'  
includeFile = 'reorganized_ta_format.txt'
```

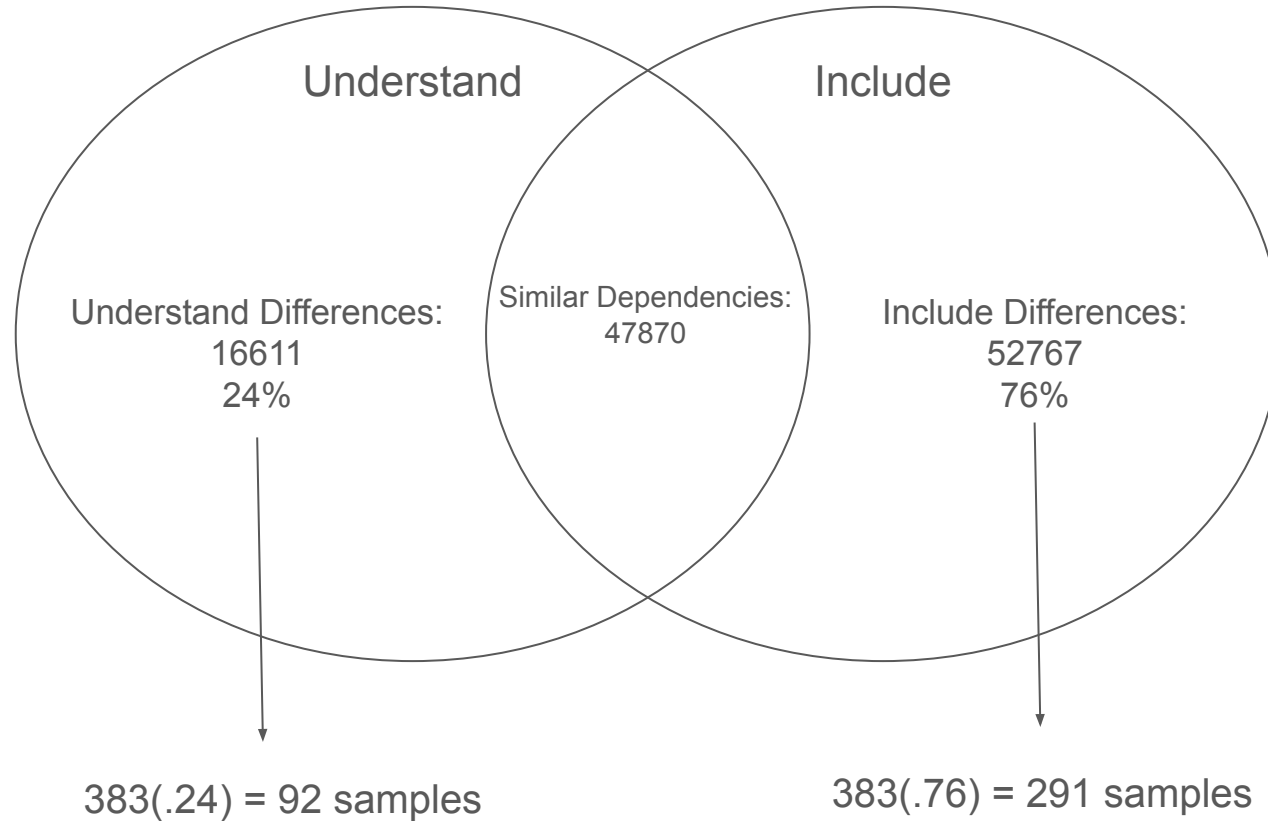
# Quantitative Analysis



# Quantitative Analysis



# Qualitative Analysis



## Determine Sample Size

Confidence Level: ☒ 95% ☐ 99%

Confidence Interval:

Population:

Calculate

Clear

Sample size needed:

# Qualitative Analysis

Precision and Recall For Understand:

Precision:  $5/(88+5) = 0.052 = 5\%$

Recall:  $(5/(5+287)) = 0.017 = 1.72\%$

Precision and Recall For Include:

Precision:  $5/(287+5) = 0.052 = 5\%$

Recall:  $5/(5+88) = 0.017 = 1.72\%$

# Rationale For Differences

Understand Files: 8129

Include Files: 13115

- Include contains all files from understand
- It is clear include looks at more files than understand
- A lot of these are tests

```
flink-1.17.1/flink-runtime/src/test/java/org/apache/runtime/operators/testutils/MockInputSplitProvider.java cFile
flink-1.17.1/flink-formats/flink-protobuf/src/test/java/org/apache/formats/protobuf/RepeatedProtoToRowTest.java cFile
flink-1.17.1/flink-yarn/src/test/java/org/apache/yarn/TestingRegisterApplicationMasterResponse.java cFile
flink-1.17.1/flink-core/src/test/java/org/apache/types/parser/UnquotedStringValueParserTest.java cFile
flink-1.17.1/flink-connectors/flink-connector-hive/src/test/java/org/apache/connectors/hive/HiveTemporalJoinITCase.java cFile
flink-1.17.1/flink-optimizer/src/test/java/org/apache/optimizer/dataproperties/GlobalPropertiesFilteringTest.java cFile
flink-1.17.1/flink-runtime/src/test/java/org/apache/runtime/checkpoint/OperatorSubtaskStateTest.java cFile
flink-1.17.1/flink-connectors/flink-connector-aws-base/src/test/java/org/apache/connector/aws/util/AWSAsyncSinkUtilTest.java cFile
flink-1.17.1/flink-connectors/flink-connector-hbase-1.4/src/test/java/org/apache/connector/hbase1/HBaseTablePlanTest.java cFile
flink-1.17.1/flink-tests/src/test/java/org/apache/test/plugin/PluginTestBase.java cFile
flink-1.17.1/flink-tests/src/test/java/org/apache/test/runtime/SelfJoinDeadlockITCase.java cFile
flink-1.17.1/flink-runtime/src/test/java/org/apache/runtime/operators/testutils/MatchRemovingJoiner.java cFile
flink-1.17.1/flink-table/flink-table-common/src/test/java/org/apache/table/utils/TableTestMatchers.java cFile
flink-1.17.1/flink-table/flink-sql-gateway/src/test/java/org/apache/table/gateway/service/SqlGatewayServiceITCase.java cFile
```



# Lessons Learned

- Different technique can yield significantly different results
- Using multiple methods will result in a better dependency extraction
  - Method invocation
  - Import statement
  - Dependency graphs

Things we regret or would do differently

- Familiarize with srcML
- Better understanding of Understand



Thank You

# References

<https://www.surveysystem.com/sscalc.htm>

<https://plg.uwaterloo.ca/~holt/papers/ta-intro.htm#:~:text=TA%2C%20the%20Tuple%2DAttribute%20language,of%20these%20nodes%20and%20edges.>

<https://projects.eclipse.org/projects/eclipse.jdt>

<https://stackoverflow.com/>

<https://www.srcml.org/>

EECS4314 Lectures

ChatGPT