

Санкт-Петербургский государственный университет

# Разработка системы проверки совместимости лицензий ПО

Автор: Погребной Дмитрий Андреевич, 17.Б11

Научный руководитель: доц. каф. СП, к.т.н. Т. А. Брыксин

Консультанты: исследователь АНО ДПО «Научно-исследовательский и образовательный центр «ДжетБрейнс» Я. Д. Голубев,  
исследователь АНО ДПО «Научно-исследовательский и образовательный центр «ДжетБрейнс» В. Д. Танков

Рецензент: специалист по машинному обучению ООО «ИнтеллиДжей Лабс» Е. Г. Булычев

Санкт-Петербург, 2021

# Открытые лицензии ПО

- Лицензирование открытого ПО – сложная область со множеством нюансов
  - Более 450 различных открытых лицензий
  - Сложная система совместимости открытых лицензий<sup>1</sup>
- Нарушение лицензирования может привести к проблемам
  - Экстренная замена зависимости с нарушенной лицензией
  - Судебные разбирательства
  - Большие денежные штрафы

<sup>1</sup> Yaroslav Golubev, Maria Eliseeva, Nikita Povarov, and Timofey Bryksin. 2020. A Study of Potential Code Borrowing and License Violations in Java Projects on GitHub. In Proceedings of the 17th International Conference on Mining Software Repositories (MSR '20)

# Причины нарушения лицензий

- По результатам исследования<sup>1</sup> существует две основные причины нарушения лицензий открытого ПО
  1. Повторное использование стороннего кода ускоряет решение тривиальных задач
  2. Непонимание условий лицензий и различия между ними

<sup>1</sup> von Krogh G., Spaeth S., Haefliger S. *Knowledge Reuse in OpenSource Software: An Exploratory Study of 15 Open Source Projects*//*Proceedings of the 38th Annual Hawaii International Conference on System Sciences*. — 2005.

# Цель и задачи

**Цель:** Разработать плагин к IntelliJ IDEA для работы с лицензиями Java проектов, который должен позволить на основании лицензий компонент проекта определять допустимые лицензии для каждого модуля проекта и сообщать программисту о потенциальных нарушениях лицензирования

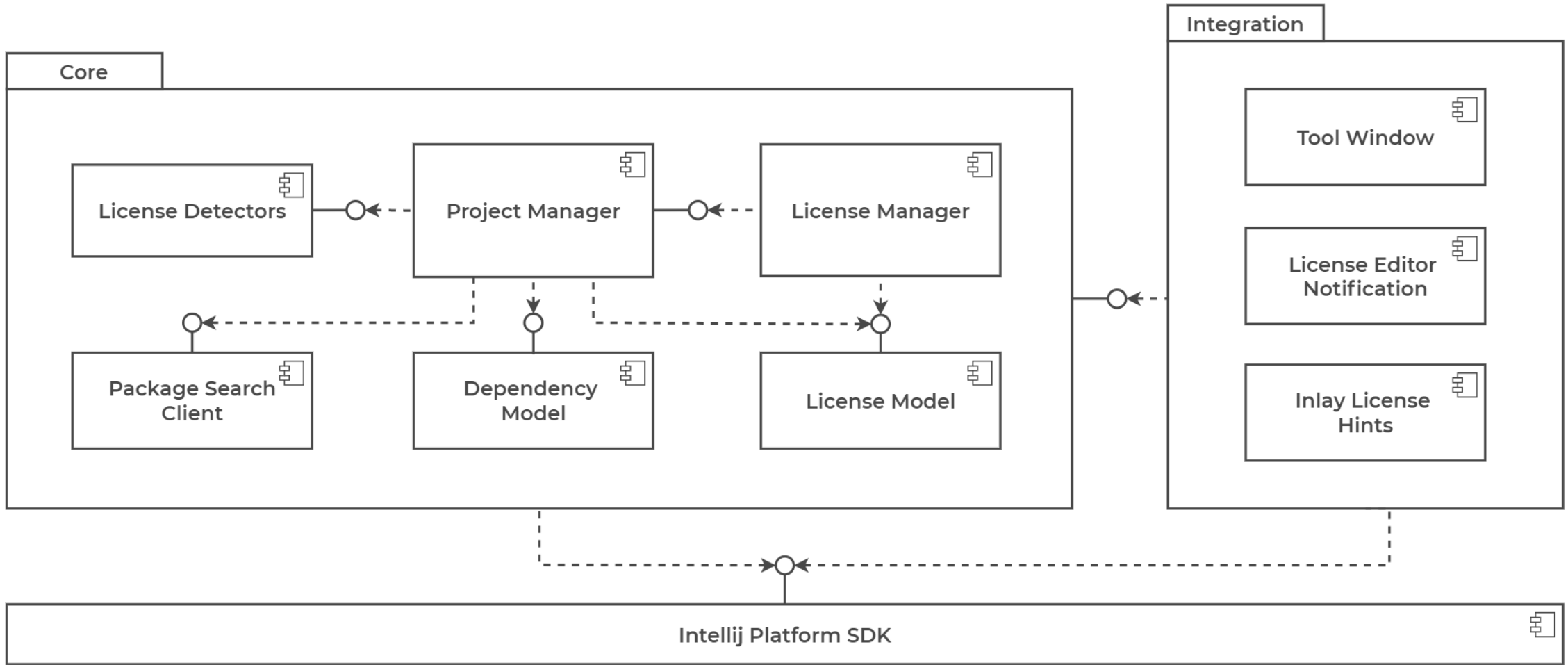
## **Задачи:**

- Сделать обзор предметной области
- Провести анализ существующих решений
- Разработать архитектуру плагина
- Реализовать алгоритм контроля лицензий проекта
  - Реализовать механизм извлечения информации о лицензиях компонент проекта
  - Реализовать механизм определения допустимых лицензий для модуля проекта
- Реализовать пользовательский интерфейс плагина
- Провести апробацию плагина

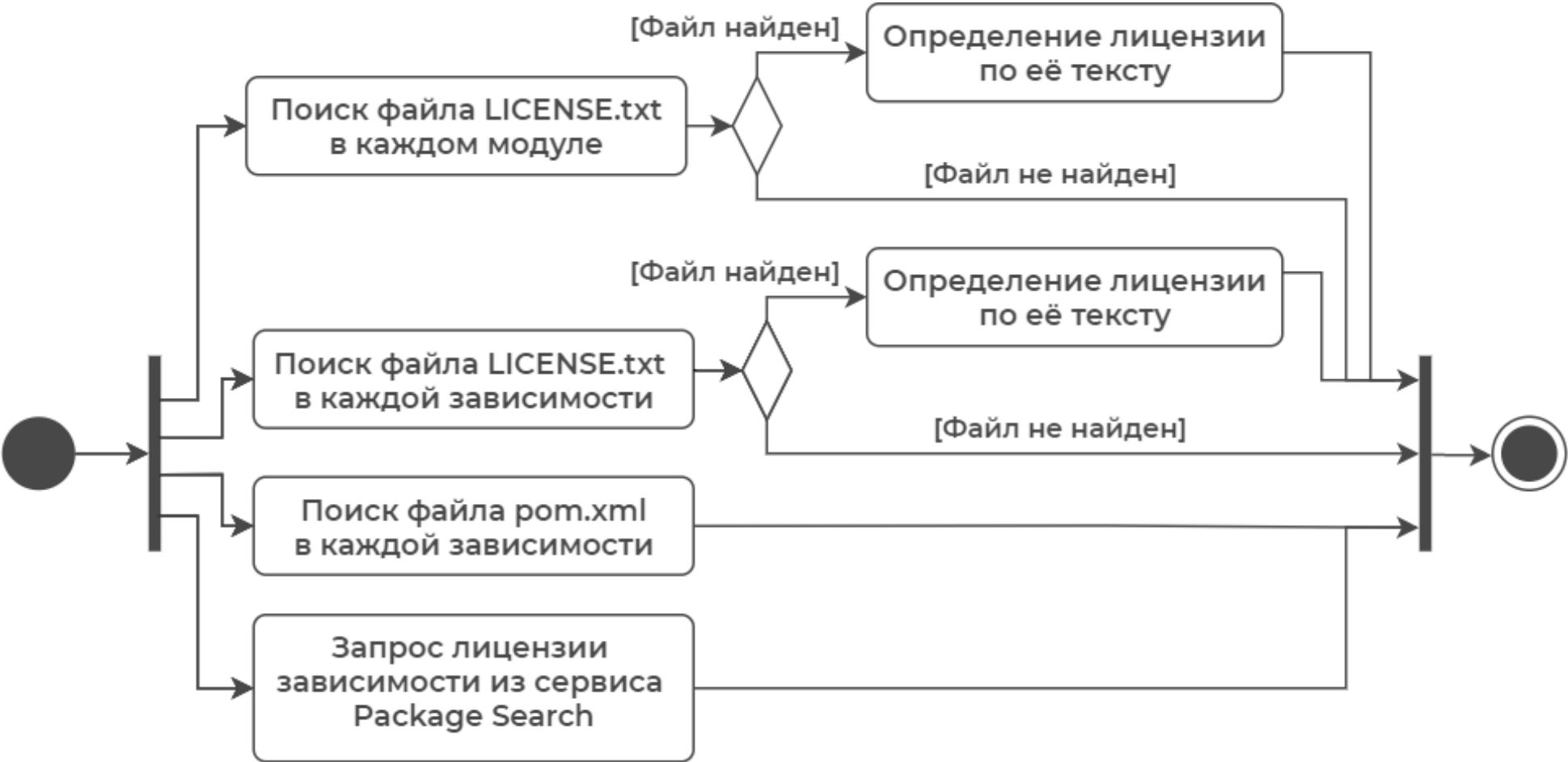
# Существующие решения

Инструмент	Лицензии	Тип	Рекомендации	Нарушения	Open-source
askalono	Проект, исходный код	Консольное приложение	-	-	+
FOSSology	Проект, зависимости, исходный код, бинарные файлы	Веб-сервис	+	+	-
go-license-detector	Проект, исходный код	Консольное приложение	-	-	+
license-checker	Проект, исходный код	Консольное приложение	-	-	+
licensee	Проект	Консольное приложение	-	-	+
scancode-toolkit	Проект, исходный код, бинарные файлы	Консольное приложение	-	-	+
FOSSA	Проект, зависимости, исходный код, бинарные файлы	Веб-сервис	+	+	-

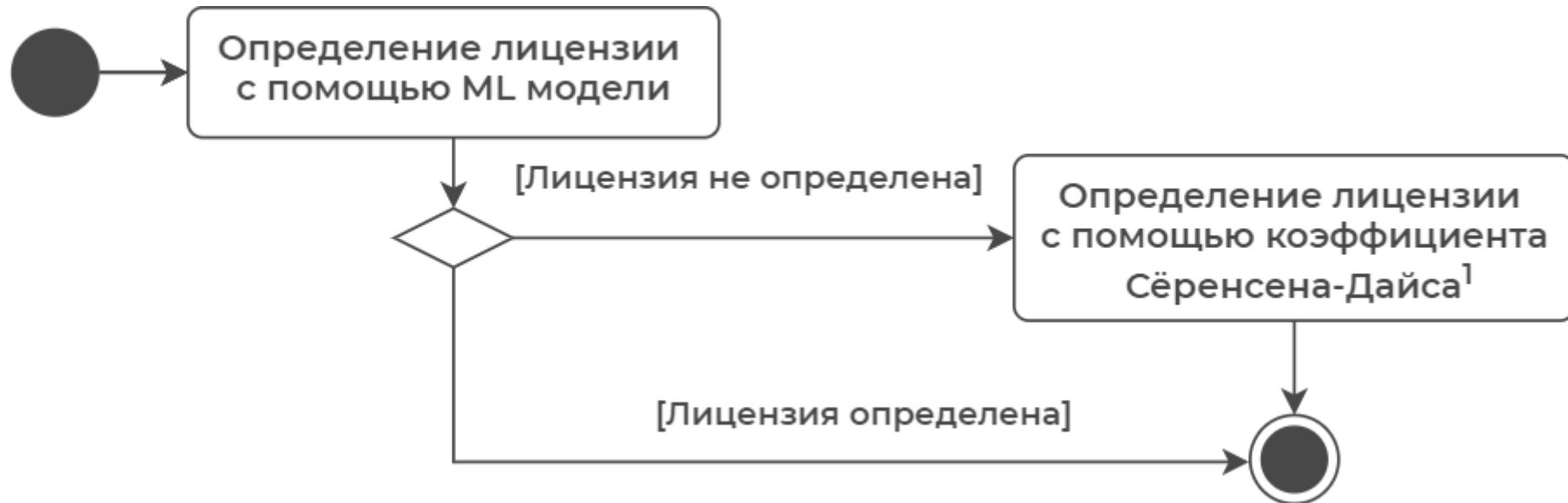
# Архитектура плагина



# Извлечение лицензий проекта



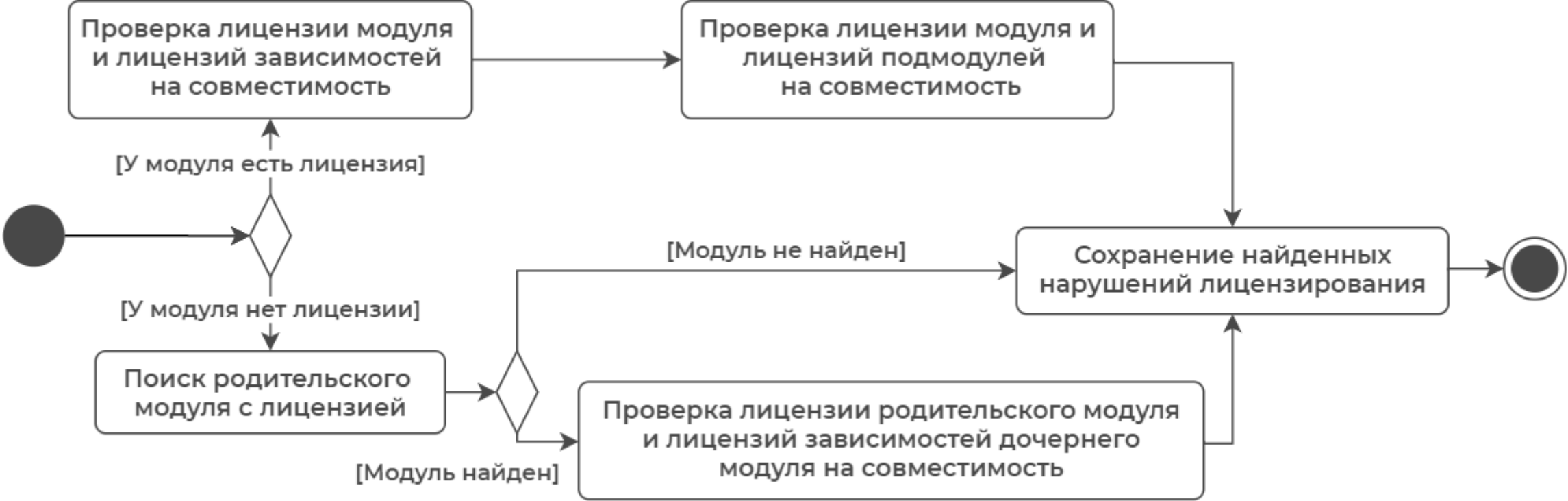
# Определение лицензии



<sup>1</sup> Sørensen, T. J. (1948). A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species and Its Application to Analyses of the Vegetation on Danish Commons. Kongelige



# Поиск нарушений лицензий



# Пользовательский интерфейс

The screenshot displays an IDE interface with two main components highlighted by red boxes:

- License Selection Dialog:** A dropdown menu is open over the text "The module license is". The menu lists various licenses, with "BSD 3-Clause 'New' or 'Revised' License" selected. Other visible options include Apache License 2.0, MIT License, GNU Affero General Public License v3.0 only, BSD 2-Clause "Simplified" License, Common Development and Distribution License 1.0, Eclipse Public License 1.0, GNU General Public License v2.0 only, GNU General Public License v3.0 only (checked), and ISC License.
- Project Licenses Panel:** A panel on the right titled "Project Licenses" shows details for the selected license. It includes a description of the "BSD 3-Clause 'New' or 'Revised' License", a table of permissions and limitations, and a list of "License Compatibility Issues".

Permissions	Limitations	Conditions
<input checked="" type="checkbox"/> Commercial use	<input type="checkbox"/> Liability	<input checked="" type="checkbox"/> License and copyright notice
<input checked="" type="checkbox"/> Modification	<input type="checkbox"/> Warranty	
<input checked="" type="checkbox"/> Distribution		
<input checked="" type="checkbox"/> Private use		

**License Compatibility Issues**

- Dependency licenses are incompatible with the **ExampleProject.main** module with license *BSD 3-Clause "New" or "Revised" License*
  - com.github.fracpete:remove-gpl** - GNU General Public License v3.0 only
- Dependency licenses are incompatible with the **ExampleProject.test** module with license *BSD 3-Clause "New" or "Revised" License*
  - com.github.fracpete:remove-gpl** - GNU General Public License v3.0 only

# Tool Window – Project License

The screenshot shows an IDE window titled "ExampleProject [C:\Users\pogre\Desktop\ExampleProject] - LICENSE.txt [ExampleProject]". The main editor displays the content of "LICENSE.txt" with a dropdown menu set to "BSD 3-Clause 'New' or 'Revised' License". The tool window on the right, titled "Project Licenses", shows the details for the selected license:

**Project Licenses:** Project License Package Licenses

**Project ExampleProject**  
C:/Users/pogre/Desktop/ExampleProject

**Root Module License**

**BSD 3-Clause "New" or "Revised" License**  
A permissive license similar to the BSD 2-Clause License, but with a 3rd clause that prohibits others from using the name of the project or its contributors to promote derived products without written consent.

Permissions	Limitations	Conditions
✓ Commercial use	⊘ Liability	ⓘ License and copyright notice
✓ Modification	⊘ Warranty	
✓ Distribution		
✓ Private use		

**License Compatibility Issues**

1. Dependency licenses are incompatible with the **ExampleProject.main** module with license *BSD 3-Clause "New" or "Revised" License*
  - **com.github.fracpete:remove-gpl** - GNU General Public License v3.0 only
2. Dependency licenses are incompatible with the **ExampleProject.test** module with license *BSD 3-Clause "New" or "Revised" License*
  - **com.github.fracpete:remove-gpl** - GNU General Public License v3.0 only

At the bottom of the IDE, the status bar shows "1:1 CRLF UTF-8 4 spaces 599 of 2048M".

# Tool Window – Package Licenses

The screenshot shows an IDE window titled "ExampleProject [C:\Users\pogre\Desktop\ExampleProject] - LICENSE.txt [ExampleProject]". The main editor displays the content of "LICENSE.txt" with the following text:

```
The module license is  Update text Show diff
```

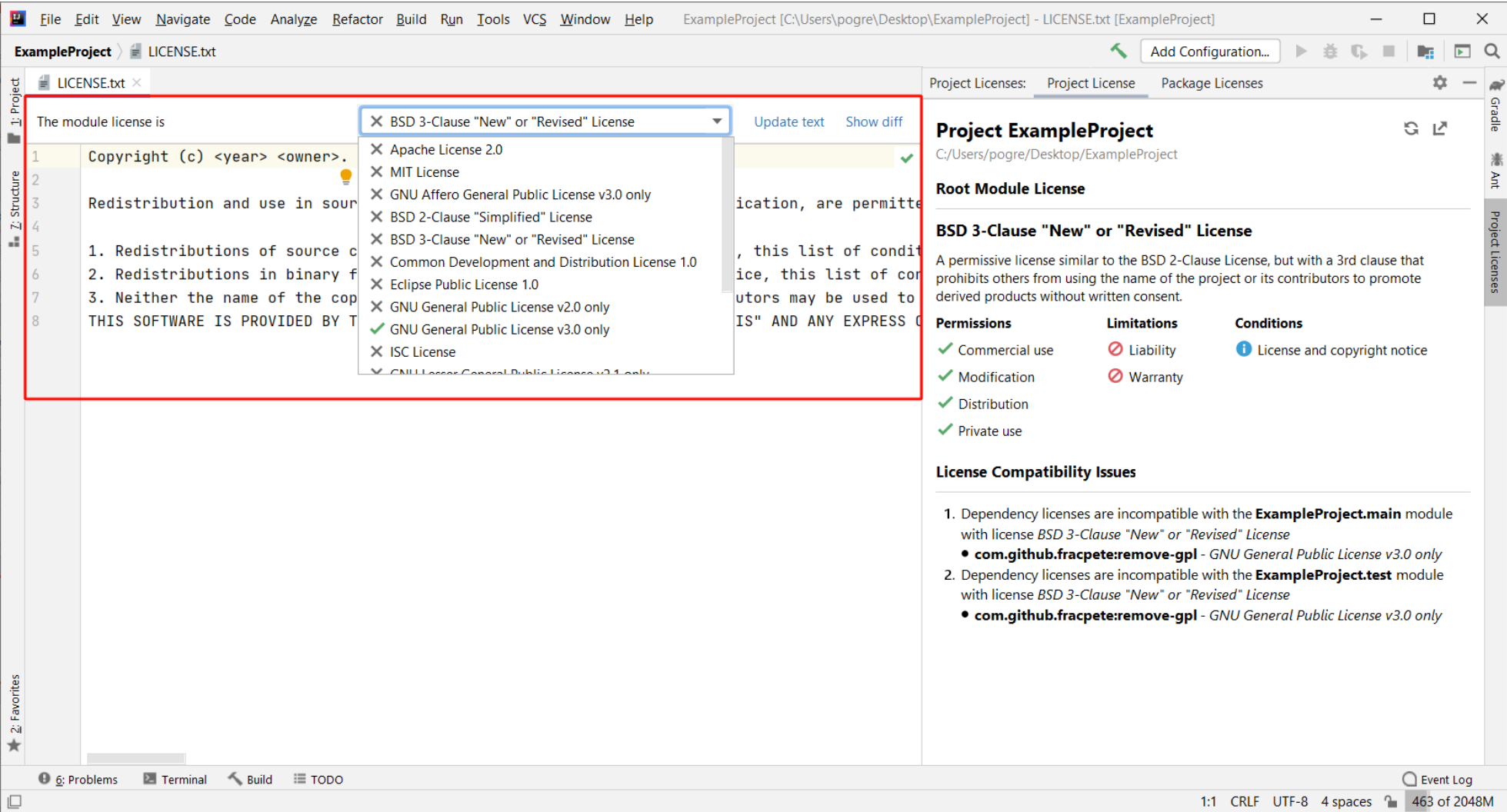
```
1 Copyright (c) <year> <owner>. All rights reserved. ✓
2
3 Redistribution and use in source and binary forms, with or without modification, are permitted under the following conditions:
4
5 1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
6 2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the
7 3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from
8 THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE
```

The "Package Licenses" tool window is open on the right, showing a list of project dependencies and their licenses:

Project Dependency	License
JUnit junit:junit	Eclipse Public License 1.0
mockito-core org.mockito:mockito-core	MIT License
remove-gpl com.github.fracpeter:remove-gpl	GNU General Public License v3.0 only
ktor io.ktor:ktor	Apache License 2.0
org.hamcrest:hamcrest-core	BSD 3-Clause "New" or "Revised" License
net.bytebuddy:byte-buddy	Apache License 2.0
Byte Buddy agent net.bytebuddy:byte-buddy-agent	Apache License 2.0
org.objenesis:objenesis	Apache License 2.0
SLF4J API Module org.slf4j:slf4j-api	MIT License
argparse4j net.sourceforge.argparse4j:argparse4j	MIT License
org.jetbrains.kotlin:kotlin-stdlib	
Kotlin Stdlib Common org.jetbrains.kotlin:kotlin-stdlib-common	
org.jetbrains:annotations	

The bottom status bar shows: 1:1 CRLF UTF-8 4 spaces 629 of 2048M

# License Editor Notification

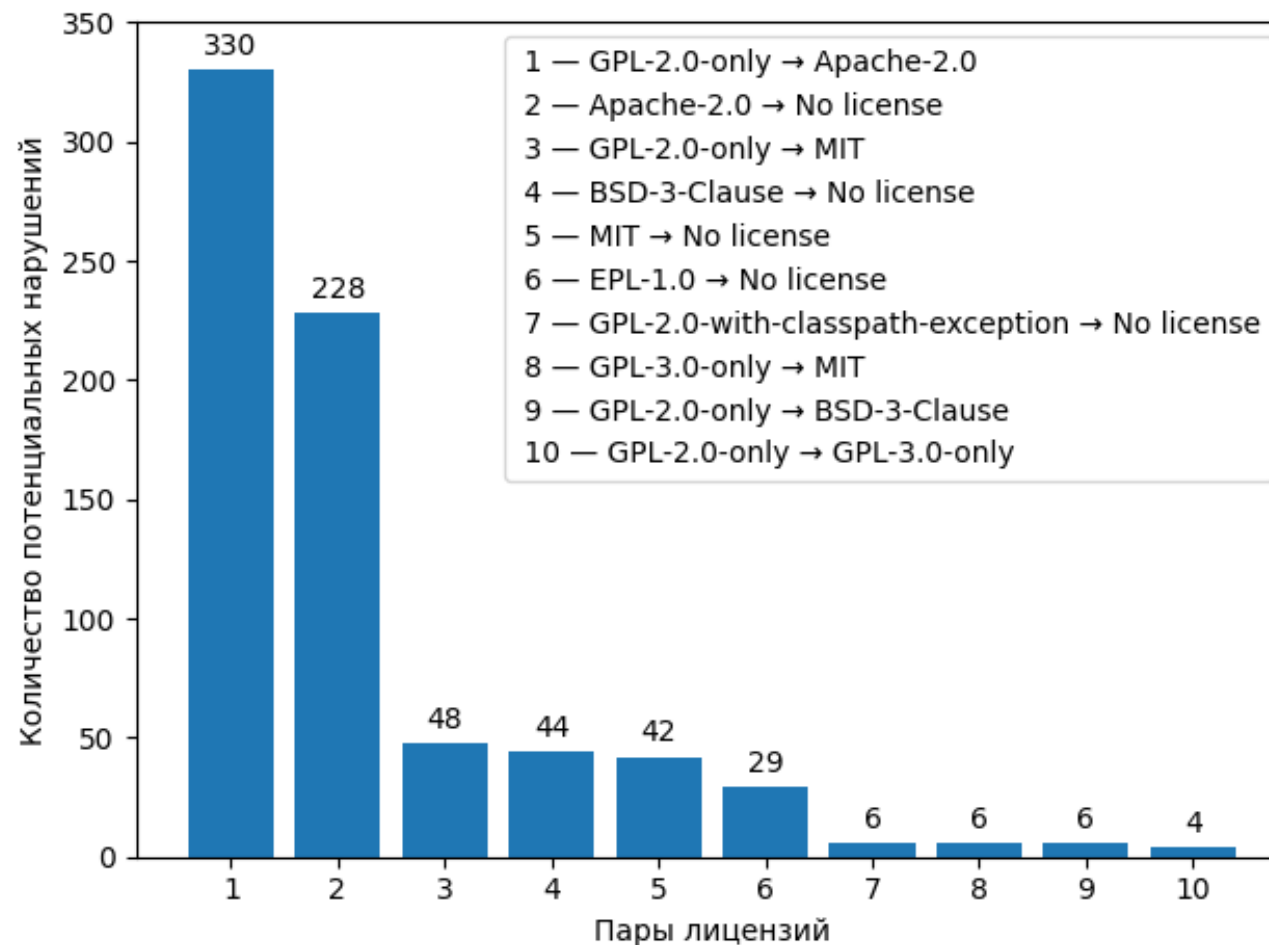


# Inlay License Hints

```
12
13 ▶ dependencies {
14     implementation group: 'junit', name: 'junit', version: '4.12' Eclipse Public License 1.0
15     implementation(group: 'org.mockito', name: 'mockito-core', version: '3.6.0') MIT License
16     implementation('org.eclipse.che.plugin:org.eclipse.search:6.12.2') Eclipse Public License 2.0 (Unsupported)
17     implementation 'io.openexchange:openexchange-boot-batch:1.0.4' Apache License 2.0
18     implementation 'com.github.fracpete:remove-gpl:0.0.3' GNU General Public License v3.0 only
19     implementation 'ch.qos.logback:logback-classic:1.3.0' Eclipse Public License 1.0
20     implementation 'org.sitoolkit.ad:sit-ad:0.5.2' Apache License 2.0
21     implementation('io.ktor:ktor:1.4.0') Apache License 2.0
22     implementation 'jpl:jpl:7.4.0' BSD 2-Clause "Simplified" License
23 }
24
```

# Апробация

- 100 самых звездных Java-проектов на GitHub
- Всего обнаружено 763 потенциальных нарушений лицензирования
- 314 потенциальных нарушений при использовании библиотек в модулях без лицензии (No license)



# Результаты

- Сделан обзор предметной области
- Проведен анализ существующих решений
- Разработана архитектура плагина
- Реализован алгоритм контроля лицензий
  - Реализован механизм извлечения информации о лицензиях компонент проекта
  - Реализован механизм определения допустимых лицензий для модуля проекта
- Реализован пользовательский интерфейс плагина
- Проведена апробация плагина

Репозиторий: <https://github.com/JetBrains-Research/license-detector-plugin>



# Точность ML модели

Predicted \ Actual	Apache-2.0	BSD-2-Clause	BSD-3-Clause	GPL-2.0-only	GPL-3.0-only	ISC	LGPL-2.1-only	LGPL-3.0-only	MIT	Total
Apache-2.0	168	1	2	0	0	0	0	1	5	177 94.92% / 5.08%
BSD-2-Clause	0	11	1	0	0	0	0	0	0	12 91.67% / 8.33%
BSD-3-Clause	0	0	66	1	0	0	0	0	5	72 91.67% / 8.33%
GPL-2.0-only	0	0	0	15	0	0	0	0	1	16 93.75% / 6.25%
GPL-3.0-only	0	0	0	0	30	0	0	1	1	32 93.75% / 6.25%
ISC	0	0	0	0	0	1	0	0	0	1 100% / 0%
LGPL-2.1-only	0	0	1	1	0	0	4	0	1	7 57.14% / 42.86%
LGPL-3.0-only	0	0	0	0	0	0	0	2	0	2 100% / 0%
MIT	2	0	0	0	0	0	0	0	454	456 99.56% / 0.44%
<b>Total</b>	170 98.82% 1.18%	12 91.67% 8.33%	70 94.29% 5.71%	17 88.24% 11.76%	30 100% 0%	1 100% 0%	4 100% 0%	4 50% 50%	467 97.22% 2.78%	<b>775</b> <b>96.90%</b> <b>3.10%</b>