# 项目结项报告

## 项目信息

- **项目 ID**：210010029
- **项目名称**：开源软件版本演进中的 License 冲突检测
- **项目地址**：

  gitlab：https://gitlab.summer-ospp.ac.cn/summer2021/210010029

  gitee：https://gitee.com/openeuler-competition/summer2021-13
- **方案描述**：

1. 构建开源许可证条款分析矩阵，条款包括但不限于链接、子证书授予、商用、分发、修改、专利授权、私用、使用商标、承担责任、提供担保、公开源码、防止协议与版权信息、使用网络分发、使用相同协议、声明变更等。

2. 定义冲突的概念以及产生冲突的条件，挖掘不同许可证之间兼容性判断的一般性规则，同时对选定开源许可证文本内容进行细致分析，构建开源许可证相容性矩阵，给出不同许可证之间的冲突情况以及相容性关系。

3. 开发开源许可证冲突查询工具，以 Web 应用形式实现，通过对项目使用的许可证以及项目要引入的许可证选择，可视化标识是否会产生冲突和相容性情况，并给出数据来源和准确性的标注。

- **时间规划**：

7.1 - 7.15（1-2 周）：参考现有资料，对开源许可证冲突以及产生冲突的条件进行定义，制定明确的标准，同时选定研究的许可证范围。

7.15 - 7.31（3-4 周）：对选定的许可证进行分析，构建许可证条款分析矩阵。根据定义的冲突产生条件，构建许可证相容性分析矩阵。

8.1 - 8.15（5-6 周）：完成开源许可证冲突查询工具开发，能够基于许可证条款分析矩阵和相容性分析矩阵，实现不同许可证冲突和相容性的查询以及可视化呈现。

8.16 - 8.31（7-8 周）：工具测试与完善。

9.1 - 9.15（9-10 周）：新增 20 余种许可证，完成新增许可证的条款分析。并对新增的 Copyleft 许可证内容进行细致分析，完成许可证相容性分析矩阵的扩充，并将该变化同步至开源许可证冲突查询工具。

9.15 - 9.22（11 周）：项目总结与材料撰写。

## 项目总结

- **项目产出：**

（1）开源许可证条款分析矩阵：该成果主要通过 Excel 图表形式呈现，该图表搜集了 60 种世界范围内流行的开源许可证，对每个许可证的 SPDX 标识符、类别、FSF 认证情况、OSI 认证情况、13 项条款的声明情况进行了详细标识。

（2）许可证相容性分析矩阵：该成果主要通过 Excel 图表形式呈现，基于对冲突和相容性的定义和理解，借助许可证相与否的判断规则，对 33 种 Copyleft 类型的开源许可证的相容性相关问题进行了细致分析和解读，构建形成流行开源许可证相容性分析图表，用于便捷查询不同许可证之间的冲突和相容情况。

（3）《开源许可证相关知识与相容性解读》白皮书撰写：归纳了开源许可证领域内的各类相关知识，主要对不同开源许可证之间的冲突问题进行了细致阐述，总结出判断开源许可证相容与否的一般性规则和各类例外情况。

（4）开源许可证冲突查询工具：以 Web 应用的方式实现不同开源许可证之间冲突查询以及兼容与否的情况，效果如下图所示。

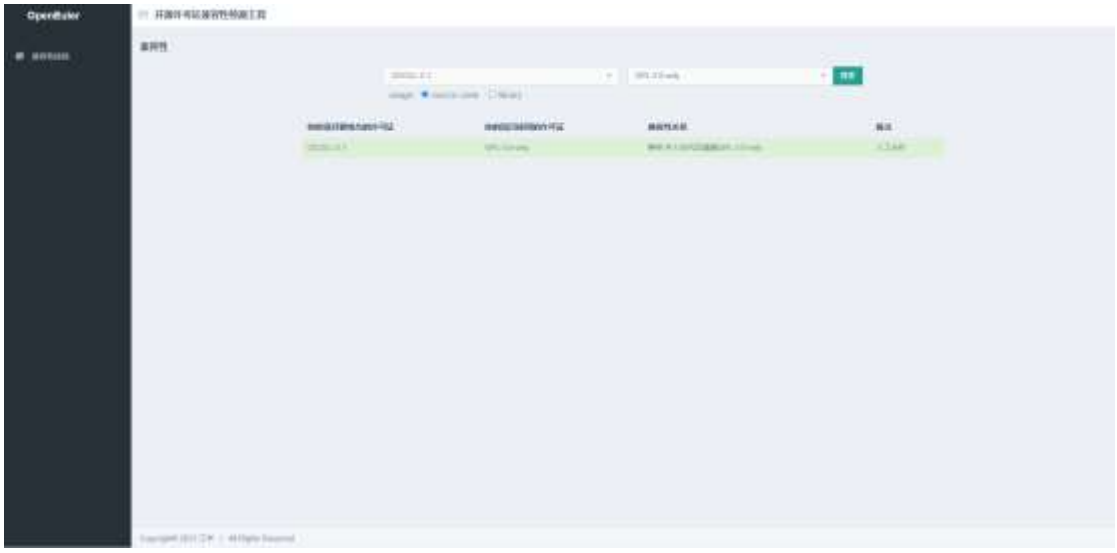

图 1 工具 UI

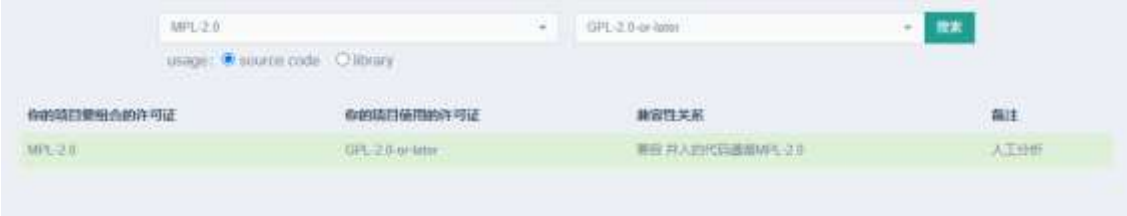


图 2 查询结果

- **方案完成情况：**

项目初设的基本目标均已达成，包含对常见开源许可证的条款以及冲突情况分析，同时基于开源许可证的条款分析矩阵和相容性分析列表，开发完成开源许可证冲突查询工具。

项目设立的挑战性目标遗憾未能完全实现，包括以 AI 和 NLP 相关技术实现许可证条款和冲突的自动化识别和抽取。此项任务将在结项后继续开展，同社区导师共同探索相关的解决方法。

- **遇到的问题与体会：**

（1）开源许可证相关资源和参考资料有限

国内有关开源合规的工作和研究相对较少，可供参考的资料更是少之又少。本次项目的主要参考资料为《开源软件知识产权风险防控研究报告》(2019)、《开源许可证兼容性指南》(2020)，这两份资料对此次项目的开展具有极高的参考和指导价值，但随着项目的进行，我对开源许可证的认识也在不断加深，因此也产生了很多同资料内容不符的见解和想法，加之不少许可证原文内容也与资料表述不相符，所以很多疑惑难以获得解答和考证，而主要解决途径就是与导师以及 Compliance SIG 组的其他老师共同探讨交流。

国外有关开源许可证的资源更为丰富，相关标准、案例、法规以及工具也更为健全。SPDX、Choosealicense、TLDRLegal、JLA、FSF、OSI、Dejacode 等机构和网站的相关资源也对我提供了极大的帮助。

就目前现有的国内外资源和技术方案而言，都未能提供便捷方式对两个开源许可证的冲突和相容性情况做出准确判断。尽管如 GPL 类许可证、EUPL 许可证等都在发布网站上给出了其兼容和不兼容的许可证列表、JLA 提供了兼容性检查工具，但存在着许可证范围有局限、兼容性判断结论有误等问题，因此本项目实际上是一项开创性的工作，虽然目前仅完成部分流行开源许可证的分析，但终极目标是能够实现所有许可证的冲突和兼容性分析的准确判断。

（2）开源许可证冲突识别流程的建立

我个人理解的许可证冲突识别过程应该包含以下几个环节。对于一个开源软件包，首先需要对其进行软件成分分析，主要目的是获取和识别该软件包通过复制、引用、依赖等等各种方式引入的开源组件，与此同时识别各开源组件的许可证；其次，对各开源组件的使用方式进行划分，主要分为使用源代码方式或使用库方式这两种方式；最后，可以根据开源许可证相容性分析表，对处于同一进程空间的许可证进行冲突判断，此外还要对项目整体使用的许可证与各开源组件使用的许可证进行兼容性判断。开源许可证冲突判断之前进行的各项工作也十分重要且是具有挑战性的。对许可证的精准识别、对许可证使用方式的识别以及对许

可证所处的进程空间的识别，都会对许可证冲突的判断产生影响。如果冲突不能及时发现，该开源软件包将给社区带来极大的隐患。

（3）许可证条款和冲突的自动化抽取

若要实现对许可证文本内容中各条款和相容性有关表述的自动抽取，现有的比较行之有效的技术方案是自然语言处理的实体抽取技术，由于本人对自然语言技术鲜有涉猎，因此还未通过此方法进行尝试，此项挑战未能顺利达成。不过未来我会继续朝着该方向努力，继续在该项目上耕耘，争取能够做出好的成果。

**- 项目质量自评**

我个人对项目完成的质量比较满意，在项目进行过程中总结的经验、对开源许可证积累的认识和理解对于项目后续的开展都是十分有益的。

**- 与导师沟通交流情况**

定期参加社区 SIG 组例会，与导师能够及时沟通，导师一直很耐心地提供帮助和指导。