XXXXX X—2022

团　　体　　标　　准

T/CIE xxx-2022

工业软件技术来源检测规范

Specification for technology source testing of industry software

2022-xx-xx 发布　　2022-xx-xx 实施

中国电子学会　发布

目录

前　　言 III
工业软件技术来源检测规范 1
1 范围 1
2 规范性引用文件 1
3 术语和定义 1
4 检测要求 2
4.1 检测目的 2
4.2 检测内容 2
4.3 检测过程 3
4.4 检测方法 3
4.5 样品及资料要求 3
5 知识产权材料审查 3
5.1.1 开源协议风险评估 3
5.1.2 授权协议评估 4
6 源代码成分检测 4
6.1.1 检测内容 4
6.1.2 基础环境 4
6.1.3 检测策略 5
7 二进制成分检测 6
7.1.1 检测内容 6
7.1.2 基础环境 6
7.1.3 检测策略 6
8 开源许可证风险检测 7
8.1.1 检测内容 7
8.1.2 基础环境 7
8.1.3 检测策略 7
附录 A（资料性附录）工业软件技术来源检测报告 1
一、 源代码成分检测 5
二、二进制成分检测 7
二、 测评环境和条件 9

前　　言

本标准按照 GB/T 1.1－2009 给出的规则起草。

本标准由XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX组织制定。

本标准由XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX解释并归口。

本标准主要起草单位：工业和信息化部电子第五研究所，XXXXXXXXXXX，XXXXXXXXXXX。

本标准主要起草人：XXX，XXX，XXX。

工业软件技术来源检测规范

1 范围

本标准规定了工业软件产品组成成分技术来源的检测要求、检测内容、检测方法及检测流程。

本标准适用于工业软件第三方检测机构、开发厂商、用户及相关人员进行工业软件技术来源的分析、测试及评价等。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 11457—2006 信息技术 软件工程术语

3 术语和定义

GB/T 11457—2006界定的以及下列术语和定义适用于本文件。

3.1

工业软件 industry software

工业软件指专用于或主要用于工业领域，为提高工业企业研发、制造、生产管理水平和工业装备性能的软件。

3.2

技术来源 technology source

工业软件产品组成成分的技术来源，一般分为开源成分、闭源成分、自主研发成分。

3.3

开源软件 open source software

软件版权所有人允许用户直接访问源代码，并通过开源许可协议将其复制、修改、再发布的权利向公众开放的计算机软件或开源组件。

3.4

开源组件 open source components

开源软件系统中最小可识别且本身不再包含另外组件的、组件信息可在公共网站获取且可独立分发、开发过程中带有版本号并且可组装的软件实体。

3.5

开源许可证 open source license

一种为计算机软件和其它产品的许可证，其允许源代码、蓝图或设计使用、修改和/或在确定的条款和条件共享。是一种法律许可，通过它，版权拥有人明确允许，用户可以免费地使用、修改、共享版权软件。是一种具有法律性质的合同，目的在于规范受著作权保护的软件的使用或者分发行为。

3.6

溯源对象 traceability object

技术来源检测所用样本库中的以软件、内核技术、开源组件、开源项目等对象构建的独立样本。

3.7

源代码特征指纹 characteristic fingerprint of source code

基于单向函数对软件源代码或其片段进行输出，形成的一系列特征值的集合。

3.8

闭源软件

源代码不公开的软件。

注：与开源软件相对。

4 检测要求

4.1 检测目的

通过检测识别工业软件产品组成成分的技术来源，并通过溯源分析，评估工业软件产品是否存在断供风险、开源许可证违规风险或传染性风险。

4.2 检测内容

工业软件技术来源检测内容包括知识产权材料审查、开源代码成分检测、闭源二进制成分检测、开源许可证风险检测。

4.3 检测过程

工业软件技术来源检测过程主要包括检测准备、检测实施、结果分析、检测总结四项活动。

检测准备：编制检测方案，确定检测范围，准备检测工具、样本库以及相应的测试环境。

检测实施：对被测软件样品进行处理，如生成样品指纹特征码；按照检测方案对被检测样品执行检测，获取检测结果；对检测结果的准确性和合理性等方面进行核查。

结果分析：对检测结果进行整理，并通过溯源分析，确定是否断供风险、许可证违规风险或传染性风险。

检测总结：根据检测结果和分析结果，编制工业软件技术来源检测报告。

4.4 检测方法

工业软件的代码规模有大有小，其技术来源通常涉及开源软件、开源代码、开源组件、第三方库以及二进制、开源许可证等，宜建立相应的样本库，并采用专业的软件成分检测工具进行自动化检测。

同时，为避免误报，在工具检测的基础上，应通过人工核查方法对相关结果进行确认。

4.5 样品及资料要求

所提交的被测样品及相关资料主要包括：

1) 工业软件技术来源检测申请表；
2) 源代码（或特征指纹，区分核心与非核心）、二进制文件；
3) 功能模块清单（区分核心与非核心）、样品文件目录对应功能模块清单（区分核心与非核心）；
4) 送测单位身份证明，如：营业执照、工作证和身份证信息等；
5) 企业代码知识产权承诺书（加盖公章）；
6) 送测软件的专利、著作权、开源协议合规性、授权协议等知识产权说明材料及相关证明材料（加盖公章）；
7) 企业及代码可持续性情况说明及相关证明材料（加盖公章）；
8) 企业履行代码风险和道德风险声明；
9) 如果有 License 授权限制，应提供授权文件，授权期限有效期不得低于 12 个月（以送样日期计算）；
10) 样品自主产权保证书；
11) 产品厂商认为与测试相关的其它资料。

5 知识产权材料审查

审查工业软件的知识产权相关材料，对工业软件在开源软件引入（含开源许可证）和商业授权引入进行识别分析。

5.1.1 开源协议风险评估

对工业软件涉及的开源软件所属开源协议的风险情况进行审核确认：

1）确认工业软件引入了哪些开源软件以及所属的开源协议；

2）分析并评估开源协议中存在的风险条款。

5.1.2 授权协议评估

评估工业软件涉及的商业授权协议、合作开发协议、许可协议等，是否存在相关授权风险，包括：

1）是否存在商业不友好授权许可；

2）授权协议的具体内容、版本、授权类型是否存在风险。

6 源代码成分检测

6.1.1 检测内容

源代码成分检测主要通过检测工具自动进行源代码文件元数据检测、开源成分溯源检测、依赖组件检测和源代码成分风险分析。

（1）源代码文件元数据检测

源代码文件元数据检测主要是对被测源代码进行静态特征提取，解析源代码文件中存在的版权声明、公司信息，形成被测源代码文件元数据清单。源代码文件元数据清单用于源代码文件风险分析。

（2）开源代码溯源检测

开源代码溯源检测主要是对被测源代码项目进行源代码成分分析，检测被测源代码中存在的开源代码文件或者代码片段，形成被测项目开源文件匹配清单，并依据开源文件匹配清单进行开源项目溯源分析，检测与被测项目具有相同开源文件的开源项目，形成开源溯源项目清单。被测项目开源文件匹配清单和开源溯源项目清单用于后续风险分析。

（3）依赖组件检测

依赖组件检测主要是对被测源代码项目进行依赖组件分析，检测被测源代码中存在的第三方依赖组件，形成被测项目依赖组件清单。依赖组件检测支持配置文件引入的依赖组件和制品包引入的依赖组件。被测项目依赖组件清单用于后续风险分析。

（4）源代码成分风险分析

源代码成分风险分析是在源代码文件元数据检测、开源代码溯源检测和依赖组件检测基础上，与检测元数据知识库进行查询、匹配和关联度分析，对源代码自身引入的开源成分及引用的第三方组件的风险进行评估，并在需要时对工业软件源代码检测结果进行人工审计和输出，确保源代码检测结果的准确性。

6.1.2 基础环境

（1）检测样本库

源代码成分检测需要构建检测样本库，主要包括开源项目特征库、开源组件库等。

（2）检测技术方法

源代码成分检测涉及的检测技术方法主要包括特征提取技术、特征相似度计算方法、成分溯源方法等。

6.1.3 检测策略

（1）源代码文件相似性判定

样品的源代码文件与样本库中样本文件进行特征相似度计算，可得到两者文件特征比对结果，通过设定特征相似度阈值来判定源代码文件与样本文件是否具备相似性，具备相似性则认为其为引入成分。可考虑的阈值参数包括但不限于相似代码行占文件总代码行比例、相似代码行数。

（2）源代码成分溯源

对与样本具有相似性的源代码成分或开源依赖成分进行来源追溯，追溯的来源对象包括开源项目、开源组件等。源代码成分溯源确定方法包括但不限于根据创建时间最早的匹配样本文件进行溯源、根据溯源结果排序并对匹配样本文件去重。

7 二进制成分检测

7.1.1 检测内容

二进制成分检测主要主要通过检测工具自动进行二进制文件资源数据提取、二进制成分检测和二进制成分风险分析。

（1）二进制资源提取

工业软件二进制资源提取主要是从被测二进制文件中提取二进制相应的资源数据，例如二进制文件的版权声明、公司信息，形成被测二进制资源数据清单。二进制资源数据清单用于源代码文件风险分析。

（2）二进制成分检测

工业软件二进制成分检测，主要是对被测二进制软件进行二进制成分分析，检测被测二进制代码中使用的第三方组件，形成被测二进制文件的二进制成分清单。被测项目开源文件匹配清单和开源溯源项目清单用于后续风险分析。

（3）二进制成分风险分析

二进制成分风险分析，是在工业软件二进制资源提取、工业软件二进制成分检测的基础上，与检测元数据知识库进行查询、匹配和关联度分析，对二进制自身及引用的第三方组件的风险进行评估，并在需要时对工业软件二进制检测结果进行人工审计和输出，确保二进制检测结果的准确性。

7.1.2 基础环境

（1）检测样本库

二进制成分检测需要构建检测样本库，主要包括闭源软件特征库、开源二进制组件库等。

（2）检测技术

二进制成分检测涉及的检测技术主要包括特征提取技术、特征相似度计算方法、成分溯源方法等。

7.1.3 检测策略

（1）二进制文件匹配

样品的二进制文件与样本库中样本文件进行特征相似度计算，可得到两者文件特征比对结果，通过设定特征相似度阈值来判定二进制文件与样本文件是否具备相似性，具备相似性则认为其为引入成分。可考虑的阈值参数包括但不限于特征相似比例。

（2）二进制成分溯源

对与样本具有相似性的二进制成分进行来源追溯，追溯的来源对象包括闭源软件、二进制组件等。二进制成分溯源确定方法包括但不限于根据创建时间最早的匹配样本文件进行溯源、根据溯源结果排序并对匹配样本文件去重。

8 开源许可证风险检测

8.1.1 检测内容

通过检测工具自动将所提取的源代码文件级许可证、项目级许可证，以及开源软件溯源清单和依赖组件清单所对应开源许可证，二进制成分清单所对应的开源许可证作为检测目标，与开源许可证传染性、合规性风险库库进行匹配，识别这些开源许可证的传染性和合规性风险。

8.1.2 基础环境

（1）检测样本库

开源许可证检测需要构建检测样本库，包括但不限于开源许可证库。

（2）检测技术

开源许可证检测涉及的检测技术包括但不限于开源许可证条款识别和翻译技术、风险条款识别方法、许可证篡改识别技术等。

8.1.3 检测策略

开源许可证风险检测通过检测许可证、许可证条款，识别许可证存在的风险条款，并可通过样品目录定位风险许可证的影响路径。

附录 A（资料性附录）工业软件技术来源检测报告

工业软件技术来源检测报告

委托单位： _____

软件名称： _____

软件版本： _____

软件类别： _____

测评机构： _____

测评时间： _____

声明

1、报告无测评机构章无效。

2、报告涂改无效。

3、本报告未经允许不得部分复制。

4、本报告只对送测产品版本有效。

5、本报告仅适用于工业软件技术来源检测。

检测结果概述

<table>
<tr><td>被测软件名称</td><td colspan="3"></td></tr>
<tr><td>被测软件版本</td><td colspan="3"></td></tr>
<tr><td>被测软件类别</td><td></td><td>检测时间</td><td></td></tr>
<tr><td>被测单位</td><td colspan="3"></td></tr>
<tr><td>检测地点</td><td colspan="3"></td></tr>
<tr><td>检测依据</td><td colspan="3"></td></tr>
<tr><td>检测结论</td><td colspan="3">本次检测是针对工业软件技术来源进行检测，主要包括知识产权材料审查、源代码成分检测、二进制成分检测以及开源许可证风险检测。本次检测结果如下所示：
（1）知识产权材料审查方面；
（2）源代码成分检测方面；
（3）二进制成分检测方面；
（4）开源许可证风险检测方面；</td></tr>
<tr><td colspan="2">拟制：

审核：

批准：</td><td colspan="2">中国赛宝实验室
（公章）

202X 年 XX 月 XX 日</td></tr>
</table>

知识产权材料审查

(1)商业授权

序号	软件/模块名称	授权方	授权时间	使用期限	使用范围	授权文件/合同名称 登记号
1						
2						
…						

(2)软件著作权

序号	软件名称	著作权人	开发完成日 期	首次发表日期	权利取得方式	权利范围	登记号
1							
2							
…							

(3)专利

序号	专利名称	专利号	专利申请日	专利权人	授权公告日	法律状态	是否参与标准建设	同族	功能实现描述
1									
2									
…									

(4)论文

提供与产品相关的论文清单（含作者、时间、题目、与产品对应的模块情况）

序号	论文	作者	国别	时间	发表刊物	功能实现描述
1						
2						

…						

一、 源代码成分检测

（1）被测软件基本信息

检测样品基本信息：

被测样品			**被测样品大小**	字节	
被测样品 SHA256 值					
文件数		**参与检测文件数**		**未参与检测文件数**	
码行数		**参与检测代码行数**		**未参与检测代码行数**	

（2）源代码成分检测结果

源代码开源成分检测结果：

代码成分匹配类型		**涉及文件个数**	**涉及代码行数**
开源	文件完全一致性		
	文件结构一致性		
	代码行一致性		
自主研发			
代码自主率			

（3）开源成分溯源及风险分析

开源成分项目溯源结果：

序号	溯源项目名称	版本	项目所属主体	项目托管平台	文件个数

溯源项目涉 A 风险分析结果：

序号	项目	项目所属主体国别	项目托管平台国别	项目捐献基金会/基金会国别	备注

依赖组件检测结果：

序号	组件名称	依赖管理	版本	匹配文件	许可证

（4）开源许可证检测及引入风险分析

开源许可证检测结果：

序号	许可证名称	引入次数	许可证颁发组织	许可证颁发组织类型
…				

开源许可证引入风险：

序号	许可证名称	许可证颁发组织	颁发组织国别	开放源代码风险	是否引入涉A风险
…					

注：由于风险依赖形势情况动态变化，风险分析结果仅做参考

二、二进制成分检测

（1）被测软件基本信息

检测样品基本信息：

被测样品			**被测样品大小**	字节	
被测样品SHA256值					
二进制文件总数		**参与检测二进制文件数**		**未参与检测二进制文件数**	

（2）二进制成分检测分析结果

样品二进制文件匹配结果：

序号	二进制文件名称	常量字符串	导出函数	匹配文件数

二进制非自主性成分结果：

二进制成分类型	涉及文件数量
自主性成分	
非自主性成分	
非自主性成分比率	

（3）非自主性成分溯源及风险分析

非自主性成分溯源结果：

序号	对象名称	文件数	对象类型	版本	所属主体	许可证	存储/收录/托管平台

非自主性成分风险分析结果：

序号	溯源对象名称	所属主体国别	存储/收录/托管平台国别	捐献基金会/基金会国别	备注

序号	溯源对象名称	所属主体国别	存储/收录/托管平台国别	捐献基金会/基金会国别	备注

二、 测评环境和条件

软件项：

序号	软件项名称	版本	用途
1	Windows 10	专业版	测试终端操作系统
…			

硬件项：

序号	硬件或固件项名称	型号	配置	用途
1	测试终端	联想 ThinkPad P14s Gen 2i	CPU：2.80GHz*4 内存：8GB 硬盘：1TB	测试客户端

序号	硬件或固件项名称	型号	配置	用途
…				