

# IRBL需求规格说明书

## 变更记录

修改人员	日期	变更原因	版本号
陈泔鎔	2021.3.9	更新功能需求	v0.1
陈俊杰	2021.3.9	更新总体描述	v0.2
陈泔鎔	2021.3.10	更新引言	v0.3
冯鑫泽	2021.3.10	更新非功能性需求	v1.0
陈泔鎔	2021.4.17	更新迭代二内容	v2.0
冯鑫泽	2021.6.10	更新迭代三内容	v3.0

## 目录

### IRBL需求规格说明书

变更记录

目录

1、引言

1.1 目的

1.2 范围

1.3 定义、首字母缩写和缩略语

1.4 参考文献

2、总体描述

2.1 产品前景

2.1.1 背景与机遇

2.1.2 业务需求

2.2 产品功能

2.3 用户特征

2.4 约束

2.5 架设和依赖

3、详细需求描述

3.1 对外接口需求

3.1.1 用户界面

3.1.2 通信接口

3.2 功能需求

3.2.1 文件预处理

3.2.2 缺陷报告的文件相似度排序

3.2.3 查看相似信息

- 3.3 性能需求
- 3.4 约束
- 3.5 质量属性
- 3.6 其他需求

# 1、引言

## 1.1 目的

本文档描述了IRBL-基于信息检索的缺陷定位系统的功能需求和非功能需求。后续展开开发实现与验证工作时，开发团队都会以此文档为依据。

## 1.2 范围

IRBL-基于信息检索的缺陷定位系统旨在智能分析错误报告与源代码文件，筛选出和bug相关性最高的源代码文件，帮助程序员更快更准确的定位bug。系统使用最常见的IR模型VSM进行索引和建模，首先对源文件进行预处理，然后计算经过预处理后的缺陷报告和每个源代码文件之间的相似程度，最后按照相似度得分高低对这些源代码文件进行降序排序，并输出相似度得分；系统同时提供指定缺陷报告与指定源代码文件的相似信息展示。

## 1.3 定义、首字母缩写和缩略语

1. IRBL: Information Retrieval Based Bug Localization
2. VSM: Vector Space Model
3. TF: Term-Frequency
4. IDF: Inverse Document Frequency

## 1.4 参考文献

1. 郭肇强,周慧聪,刘释然,李言辉,陈林,周毓明,徐宝文. 基于信息检索的缺陷定位:问题、进展与挑战.软件学报.  
<http://www.jos.org.cn/1000-9825/6087.htm>
2. R. K. Saha, M. Lease, S. Khurshid and D. E. Perry, "*Improving bug localization using structured information retrieval*," 2013 28th IEEE/ACM International Conference on Automated Software Engineering (ASE), Silicon Valley, CA, USA, 2013, pp. 345-355, doi: 10.1109/ASE.2013.6693093.
3. M. Rath and P. Mäder, "*Influence of Structured Information in Bug Report Descriptions on IR-Based Bug Localization*," 2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), Prague, Czech Republic, 2018, pp. 26-32, doi: 10.1109/SEAA.2018.00014.
4. Jaime Arguello. *Vector Space Model*. INLS 509: Information Retrieval
5. 骆斌,刘嘉,张瑾玉,黄蕾.《软件工程与计算（卷三） 团队与软件开发实践》
6. IEEE标准

## 2、总体描述

### 2.1 产品前景

#### 2.1.1 背景与机遇

对于程序开发人员来说，debug永远是一个灵人头疼的事，尤其是当项目规模比较大的时候。现在的集成开发环境大多只能提供静态语法检查以及程序编译与运行时的错误报告，且只会简单的报告导致程序停止运行的那一行代码。而现实中程序员遇到更多的是程序逻辑甚至业务逻辑上的bug，只能依据程序员对项目的了解与经验去定位bug所在。

为解决上述问题，IRBL系统将使用VSM技术智能分析错误报告与源代码文件，筛选出和bug相关性最高的源代码文件，帮助程序员更快更准确的定位bug，从而大幅提高程序员的工作效率。

#### 2.1.2 业务需求

BR1：系统上线半年后反馈系统定位准确的用户占80%以上

### 2.2 产品功能

SF1：对源代码进行包括去除关键词、去除停用词、词根还原以及拆分复合词的预处理

SF2：通过VSM对已经预处理好的源代码文件与错误报告文件进行相似度计算与排序

SF3：可以查看源代码文件与选定缺陷报告的相似度与具体相似信息（相似度与相似的代码行标注）

### 2.3 用户特征

程序开发人员	系统的使用者，希望上传项目文件与错误报告文件后，系统能帮助他们快速定位bug。对系统的准确度要求高。计算机水平高，对界面要求高。
--------	--

### 2.4 约束

CON1: 系统使用Web界面

CON2: 系统主要使用java语言开发

CON3: 系统使用迭代式开发

CON4: 开发中要求开发者提交计划、软件需求规格说明文档，设计文档与测试文档

### 2.5 架设和依赖

AE1: 源代码文件与错误报告文件已上传并已预处理好

AE2: 用户网络畅通

AE3: 不存在高并发场景

## 3、详细需求描述

3.1 对外接口需求

3.1.1 用户界面

展示缺陷报告列表

IRBL

缺陷序号	简述	操作
75739	Variant has no toString()	<a href="#">查看</a>
77948	NullPointerException in CLabel.findMnemonic	<a href="#">查看</a>
78548	[consistency] Button Selection fires before MouseUp	<a href="#">查看</a>
78559	[consistency] Slider fires two Selection events before MouseDown	<a href="#">查看</a>
79107	[consistency] setItems(String[]) with null elements	<a href="#">查看</a>
79481	Sash no longer draggable when too small	<a href="#">查看</a>
80120	CTabFolder layout puts top right item one pixel to far to the right	<a href="#">查看</a>
80276	CBannerLayout calls Control.update too often	<a href="#">查看</a>
78854	Memory leak in ClipboardProxy.getFunc()	<a href="#">查看</a>
81265	background of tree is incorrect when disabled	<a href="#">查看</a>
80506	[consistency] forceActive and setActive are inconsistent across platforms	<a href="#">查看</a>
58185	Eclipse fails to load oldstyle BMP files	<a href="#">查看</a>

Q 评估

展示评估指标

front-endfront-end+

←→localhost:8081/#/

☆👤⋮

IRBL

缺陷序号	简述	操作
75739		
77948		
78548		
78559		
79107	[consistency] setItems(String[]) with null elements	<a href="#">查看</a>
79481	Sash no longer draggable when too small	<a href="#">查看</a>
80120	CTabFolder layout puts top right item one pixel to far to the right	<a href="#">查看</a>
80276	CBannerLayout calls Control.update too often	<a href="#">查看</a>
78854	Memory leak in ClipboardProxy.getFunc()	<a href="#">查看</a>
81265	background of tree is incorrect when disabled	<a href="#">查看</a>
80506	[consistency] forceActive and setActive are inconsistent across platforms	<a href="#">查看</a>
58185	Eclipse fails to load oldstyle BMP files	<a href="#">查看</a>

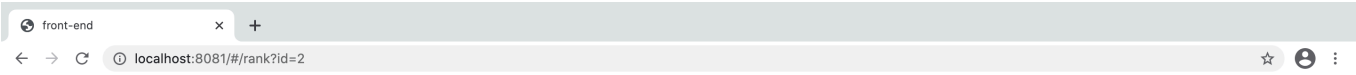
Q 评估

显示评估指标

Top@1 : 53.07%  
Top@5 : 74.49%  
Top@10 : 84.7%  
MRR : 0.64  
MAP : 0.54

取消

展示指定缺陷报告下按相似度降序排列的文件列表



IRBL

文件排序	相似度	文件序号	文件名	操作
1	0.5731108186820884	40	CLabel.java	<a href="#">查看</a>
2	0.2763945673610844	1	SWT.java	<a href="#">查看</a>
3	0.27298851475295305	86	ViewForm.java	<a href="#">查看</a>
4	0.26601050113711566	43	CTabFolder.java	<a href="#">查看</a>
5	0.25695926185402485	226	Label.java	<a href="#">查看</a>
6	0.24806188446104485	212	Control.java	<a href="#">查看</a>
7	0.24575416927834276	206	Button.java	<a href="#">查看</a>
8	0.24345895834621634	223	Group.java	<a href="#">查看</a>
9	0.2331748976894578	218	Display.java	<a href="#">查看</a>
10	0.22477206498164176	260	Widget.java	<a href="#">查看</a>
11	0.21361937280850926	228	Link.java	<a href="#">查看</a>
12	0.20849547360239948	3	SWTException.java	<a href="#">查看</a>
13	0.20842461367427715	49	StyledText.java	<a href="#">查看</a>

展示指定文件的详细内容

IRBL

Program.java

swt-3.1/src/org/eclipse/swt/program/Program.java

```

/*****
 * Copyright (c) 2000, 2005 IBM Corporation and others.
 * All rights reserved. This program and the accompanying materials
 * are made available under the terms of the Eclipse Public License v1.0
 * which accompanies this distribution, and is available at
 * http://www.eclipse.org/legal/epl-v10.html
 *
 * Contributors:
 *   IBM Corporation - initial API and implementation
 *****/
package org.eclipse.swt.program;

import org.eclipse.swt.internal.*;
import org.eclipse.swt.internal.win32.*;
import org.eclipse.swt.*;
import org.eclipse.swt.graphics.*;

import java.io.IOException;

/**
 * Instances of this class represent programs and
 * their associated file extensions in the operating
 * system.
 */
public final class Program {
    String name;
    String command;
    String iconName;
}
```

3.1.2 通信接口

用户浏览器与服务器使用 HTTP 协议进行通信。

3.2 功能需求

3.2.1 文件预处理

3.2.1.1 特性描述

对源文件进行预处理，并将预处理好后的结果保存为json文件。

优先级 = 最高

3.2.1.2 刺激/响应序列

刺激：用户指定项目进行预处理

响应：系统对项目内的文件进行预处理，返回预处理好的json文件。

3.2.1.3 相关功能需求

编号	描述
JDT.ASTVisitor	用于构建AST树
NLP.stanfordNLP	对传入的字符串进行预处理
FileTranslator.write	将预处理好后的结果保存为json文件

3.2.2 缺陷报告的文件相似度排序

3.2.2.1 特性描述

根据预处理好的缺陷报告，获取源代码文件的相似度排序。

优先级 = 最高

3.2.2.2 刺激/响应序列

刺激：用户选择指定的缺陷报告

响应：系统返回按照与该缺陷报告的相似度得分对这些源代码文件的降序排列的列表

3.2.2.3 相关功能需求

编号	描述
Report.getAllReport	得到所有缺陷报告的列表
VSM.sort	按照与指定缺陷报告的相似度得分对这些源代码文件进行降序排列
VSM.calculateSimilarity	使用VSM为选定的缺陷报告与选定源代码文件的相似度计算
File.calculateTF	计算指定文件中指定术语的出现频率（TF）
Project.calculateIDF	计算指定术语在指定Project中的珍稀度（IDF）

### 3.2.3 查看相似信息

#### 3.2.3.1 特性描述

用户点击相似度排序中的单个源代码文件，可以查看具体的相似度与相似信息。

优先级 = 高

#### 3.2.3.2 刺激/响应序列

刺激：用户点击相似度排序中的单个源代码文件

响应：系统返回该源代码文件相似度与具体相似信息（相似度与相似的代码行标注）。

#### 3.2.3.3 相关功能需求

编号	描述
File.readFile	读取指定文件
Report.getQueryWords	得到指定缺陷报告作为查询处理后对应的一组词汇（在前端再对读取的文件做相似代码行的标注）
VSM.calculateSimilarity	使用VSM为选定的缺陷报告与选定源代码文件的相似度计算

### 3.3 性能需求

类别	ID	需求内容
速度	PR1	所有用户的查询都必须在3秒内完成
速度	PR2	查询后1秒内显示排序列表
容量	PR3	系统至少能存10万条用户数据
吞吐量	PR4	系统每分钟应该处理超过10万条请求
负载	PR5	系统允许10万个用户同时使用，还能正常工作
实时性	PR6	点击后3秒内显示详细相似信息

3.4 约束

ID	需求内容
C1	系统每半年出现的服务器中断次数不得超过2次
C2	系统的最大响应时间不得超过5s
C3	系统要保证较高的安全性需求，保证用户的信息的安全
C4	系统需要允许资源可以永久扩展

3.5 质量属性

ID	特性	需求内容
QR1	功能性	系统需满足所用的用户需求
QR2	可移植性	系统为程序员定制，不需考虑可移植性
QR3	可维护性	系统易于维护，易于进行修改和缺陷剔除
QR4	效率	系统要及时响应用户的请求，保证系统的效率
QR5	可靠性	如果在同客户交互时，网络出现故障，系统不能出现故障，数据库的数据实时备份，在丢失或破损后可自动修复
QR6	易用性	无需使用手册或培训即可直接使用系统，使用系统的过程中，用户无需复杂的操作，页面转换不超过5次



3.6 其他需求

ID	需求内容
1	系统对数据进行备份，定时保存
2	系统安装部署时，需同时导入数据与配置文件，保证系统安装后即可供用户使用