

实验一：需求获取

实验目的

■ 开源ide的软件需求分析

实验思路

- 选择开源ide: **NetBeans**
- 信息源: **StackOverflow**
- 获取信息: 从StackOverflow中爬取以ide和NetBeans为标签的数据
- 处理数据: 对爬取的数据进行分析
- 需求提取和分类: 从处理过的数据中提取客户对软件的需求并进行分类

实验步骤

目标ide与信息源

- **NetBeans** IDE最新版下载量已经超过18万次，参与开发人员超过80万，用户量大，可以爬取的信息充足，可以较为准确的进行需求分析
- **StackOverflow**是一个与程序相关的IT技术问答网站，平台上面有很多有关ide及NetBeans相关的问答，有很多专业人士提出的使用过程中的问题，能较为真实的反应ide中存在的缺陷或未完善的需求

数据获取

- 文件

■ creeper.py 爬虫，爬取StackOverflow关于IDE和NetBeans的数据

- 获取思路

通过爬虫从StackOverflow抓取相关问答，并通过数据清洗筛选所需信息，从而获取有用数据

数据

- 文件

■ ide.txt 爬取Stack Overflow的以IDE为标签的数据

netbeans.txt 爬取Stack Overflow的以NetBeans为标签的数据

数据分析

- 文件

■ frequentAnalyse.py 数据分析程序，用于分析爬取的数据，主要是通过词频分析来确定需求

wc.py 词云生成工具，将获得的数据制作为图片，能够较为直观地显示出数据的出现频率

- 分析思路
 - 将爬取的数据文件通过编写的词频分析程序进行词频获取，提取出词频较高的50个关键字，并进行筛选，从关键字中提取出用户的潜在需求然后进行分类

分析结果

- 文件

`ide_frequence.txt` 保存对`ide.txt`中的数据进行词频分析后得出的数据

`netbeans_frequence.txt` 保存对`netbeans.txt`中的数据进行词频分析后得出的数据

需求总结

1. 对标签为IDE的数据进行分析，代表了对IDE的普遍需求，关键词主要有以下几类

- 代码、文本编辑功能：

关键词	<code>code</code>	<code>file/files</code>	<code>editor</code>	<code>line</code>	<code>text</code>
出现次数	829	438/338	365	301	230

关键词	<code>change</code>	<code>duplicate</code>	<code>source</code>	<code>color</code>	
出现次数	205	183	182	106	

- 代码分析、编译、调试功能

关键词	<code>error</code>	<code>run</code>	<code>class</code>	<code>new</code>	<code>build</code>	<code>debug/debugging</code>
出现次数	328	231	176	145	145	130/136

- 多语言编辑

关键词	<code>python</code>	<code>php</code>	<code>javascript</code>	<code>html</code>
出现次数	358	296	192	129

- 多环境编辑

关键词	<code>android</code>	<code>windows</code>	<code>linux</code>	<code>web</code>
出现次数	465	277	149	127

- 文件管理

关键词	<code>closed</code>	<code>open</code>	<code>add</code>
出现次数	818	197	154

- IDE平台

关键词	<code>phpstorm</code>	<code>pycharm</code>	<code>delphi</code>	<code>vs</code>	<code>xcode</code>	<code>webstorm</code>
-----	-----------------------	----------------------	---------------------	-----------------	--------------------	-----------------------

关键词	phpstorm	pycharm	delphi	vs	xcode	webstorm
出现次数	278	219	211	181	154	135

- 功能扩展

关键词	development	auto	find	plugin
出现次数	259	172	160	152

关键词	selenium	support	method	
出现次数	139	137	126	

2. 对标签为NetBeans的数据进行分析，代表了对NetBeans需求，关键词主要有以下几类

- 代码、文本编辑功能：

关键词	file/files	code	find	data	text	line
出现次数	1677/622	996	458	386	366	304

- 代码分析、编译、调试功能：

关键词	error	class	run/running	duplicate	build
出现次数	1494	848	736/409	589	528

- 多语言编辑

关键词	php	web	mysql
出现次数	663	662	441

- 多环境编辑

关键词	windows
出现次数	302

- 项目管理工具

关键词	jar	maven	glassfish	tomcat
出现次数	838	713	486	371

- 文件管理

关键词	add	closed	library
出现次数	568	470	336

- 性能需求

1. 代码、文本编辑功能：主要体现在外观（color）、以及行列（line、text）、文本复制（duplicate）等基础的编辑功能
 2. 代码分析、编译、调试功能：主要体现在对代码编译（run）、以及调试过程中出现错误的（debug/debugging、error）等的提问
 3. 多语言编辑：出现频率最高的为python、php、javascript、html
 4. 各种环境下的编译：如windows、linux、android、web等
 5. 文件管理功能：open、closed、add
 6. 功能的扩展：出现了自动（auto）、插件（plugin）、测试工具（selenium）以及支持（support）等词中，大致可以分析出一下几个潜在地需求：
 - 自动化：包括代码的自动补充，框架的自动填充等等
 - 插件：更好地实现IDE的个性化生成，使用插件可以让用户具有更好地体验
 - 测试工具：实现测试的自动化，可以高效、科学地测试代码，减少错误
 - 支持：可能包含支持多语言开发、支持多环境运行等等
- 对于NetBeans的情况来看，需求不仅包括上述IDE的普遍需求，也包含了一些具有针对性的需求
 1. 对于IDE的基本需求：包含在对IDE的整体情况分析中
 2. 项目管理工具：包含jar、maven、glassfish、tomcat这些高频词
 - jar用于聚合大量的Java类，以便开发Java平台软件或库
 - maven是一个采用纯java编写的开源项目管理工具，用于统一开发规范与工具以及统一管理jar包
 - glassfish、tomcat都是Servlet容器，可以用来部署多个web应用
- 由上述情况可以看出，NetBeans的用户还是较为关注对于开发项目的管理，因此是一个较为重要的需求
3. 性能需求：出现了高频词server，可能表示用户对性能也有一定的期望值。
 4. 功能需求：会出现数据库的使用，以及图形化界面的使用（gui、jframe、javafx、swing）在对NetBeans的数据分析中，有关图形化界面的词出现的频率很高，而且出现了不同的应用，说明对图形化界面的需求还是很大。
 5. 功能扩展：插件（plugin）的需求，以及高频词（method）的出现说明还是缺少许多需要实现的内容。

成绩分配

姓名	比例
谢宇轩	34%
刘馨蓬	33%
刘瑞芳	33%