

软件需求工程Lab1实验报告

2020.11.1

小组成员

学号	姓名	成绩占比
171870559	季镇澜	20%
181860013	储备	20%
181860077	余帅杰	20%
181860094	王佳奕	20%
181860151	周清远	20%

实验目的

获取Smart IDE的潜在需求，并对需求进行分类和分析。

实验方法、过程分析及结果展示

数据获取

我们先在Stack Overflow上通过“IDE”的标签进行搜索，找到了大量关于IDE的问答，这些回答可以作为Smart IDE的潜在需求，因此，我们通过爬虫抓取了这些问答 (<https://api.stackexchange.com/2.2/questions?order=desc&sort=activity&tagged=ide&pagesize=100&site=stackoverflow&filter=!5IW-1CBPjKYw2xymfiaebgSti&key=602UeLTB17TQ0y7WP3t>)，爬虫抓取的部分数据展示如下，详细内容可在我们的仓库中查看（仓库地址：https://github.com/NJU-Software-Requirement/Software_Requirement_EXP1/blob/master/data/ide ques.txt）：

10001 lines (10001 sloc) | 532 KB

```
1 Finding definitions of modules which are parameters with LSP (pyright)
2 How to set relative search path for require/import in Visual Studio Code
3 PhpStorm IDE: Cursor gets stuck in text-select (highlight) mode in Ubuntu 18 LTS
4 Netbeans 8.2 Locking Up Frequently
5 How can I make mousewheel work in VB6 IDE?
6 Eclipse - Create new web project from existing one without any dependencies
7 Having trouble adding header files on KIEL
8 How to make visual studio break only on unhandled exceptions?
9 IPython IDE for Windows
10 Where to program with Nim (linux)?
11 How to import javax.comm API to Netbeans IDE?
12 IDE like Excel--Put your lines of Code Wherever?
13 For unity C#, Python or java, What is better, an IDE or a Code Editor?
14 Integrate Spring with IntelliJ
15 IDE adding view in the wrong folder
16 How to type hint Google Types in Google Scripts?
17 Developing and deploy jobs in Apache Flink
18 Can't install Android Studio (IDE cannot create directory Mac)
19 How to make ide in Android like AIDE to compile android project in android phone itself
20 QtCreator: No valid kits found
21 Is there a NetBeans plugin for Google Cloud Platform development?
22 emIDE download seems to be unavailable
23 Android Studio with custom framework - spurious unresolved reference warning: how to suppress it?
24 How to rebind a key to 'Continue' command other than F5 in Visual Studio?
25 How to jump back to NERDTree from file in tab?
26 How to format too long lines in Webstorm?
27 Recover deleted files in spyder 4
```

另外，很多文字没有很强的结构特征，对于这类的数据，我们需要一个分类器，对这样的数据判断其类型，进而提取Smart IDE的潜在需求。我们还爬取了10000条左右IDE相关的问答数据以此为按照9: 1的比例划分训练集和测试机训练了一个分类器，分类器会根据问答的内容对没有打上标签的问答进行分类，将它们归类到某一标签，准确率可达80%。分类器的效果展示如下：

```
epoch 7 loss: 0.25687330961227417 right 1884 total 2128 Acc: 0.8853383458646616
Eval Result: right 195 total 238 Acc: 0.819327731092437
TOP ACC: 0.819327731092437 Save Model!
epoch 8 loss: 0.572472095489502 right 1927 total 2128 Acc: 0.9055451127819549
Eval Result: right 197 total 238 Acc: 0.8277310924369747
TOP ACC: 0.8277310924369747 Save Model!
epoch 9 loss: 0.32057467103004456 right 1969 total 2128 Acc: 0.9252819548872181
Eval Result: right 200 total 238 Acc: 0.8403361344537815
TOP ACC: 0.8403361344537815 Save Model!
epoch 10 loss: 0.16013748943805695 right 2014 total 2128 Acc: 0.9464285714285714
Eval Result: right 199 total 238 Acc: 0.8361344537815126
epoch 11 loss: 0.0810876414179802 right 2028 total 2128 Acc: 0.9530075187969925
Eval Result: right 197 total 238 Acc: 0.8277310924369747
epoch 12 loss: 0.31996485590934753 right 2070 total 2128 Acc: 0.9727443609022557
Eval Result: right 200 total 238 Acc: 0.8403361344537815
epoch 13 loss: 0.2790198028087616 right 2055 total 2128 Acc: 0.9656954887218046
Eval Result: right 204 total 238 Acc: 0.8571428571428571
TOP ACC: 0.8571428571428571 Save Model!
```

最后一轮训练结果准确率高达85.7%。

Stack Overflow只是我们本次实验中众多数据来源中的其中一个，这个开发者问答平台内容丰富，种类繁多。要想获取Smart IDE更多、更详尽的潜在需求，我们还需要综合考虑当前众多知名IDE项目各自的特性和缺陷。于是，我们还爬取了github上JetBrains IntelliJ IDEA的Pull Request(PR)约1400条内容 (https://api.github.com/repos/JetBrains/intellij-community/pulls?state=all&per_page=100)，以及Visual Studio Code的Issue约10000条 (<https://api.github.com/repos/microsoft/vscode/issues?stat>)

[e=all&per_page=100](https://api.github.com/repos/microsoft/vscode/pulls?state=all&per_page=100)) 和Pull Request约8000条 (https://api.github.com/repos/microsoft/vscode/pulls?state=all&per_page=100) , 爬取的部分内容展示如下:

VScode中issue的title (仓库地址: https://github.com/NJU-Software-Requirement/Software Requirement EXP1/blob/master/data/vscode_issues_title.txt) :

```
10000 lines (10000 sloc) | 504 KB

1 Can u add a built in function to open a server to test and see our html code? Thanks!
2 Why are the floating prompts of variables and functions in JS inconsistent with those in HTML?
3 Git Push (Follow Tags) not working in github
4 wrong color
5 New branch not visible
6 Bad Auto-Format on EJS
7 VS Code scroll lags on macos
8 Provide a way for a CompletionItem to have both a color and documentation
9 VS Code setup removes js files association with Windows Scripting Host (WSH)
10 linter toggle
11 Regex Replace not working for look-behind expression matching a newline
12 TS Server fatal error: Cannot destructure property `name` of 'undefined' or 'null'.
13 Provision to get a `pickString` dynamically populated in a VSCode task
14 Provision to select multiple values from the list in task
15 The cursor lags when typing/selecting/deleted text quickly if the editor.cursorSmoothCaretAnimation option is set to true
16 I would like dock toolbar to top of the sidebar
17 ${fileDirname} does not resolve correctly
18 natvis does not support enum class template argument types
19 Unwanted HTML tag highlight since recent updates
20 [1031/181118.450:ERROR:registration_protocol_win.cc(103)] CreateFile: 系统找不到指定的文件。 (0x2)
21 Should "Report Issue" work for built-in extensions in Running Extensions view?
22 OPEN_TO_SIDE_COMMAND_ID should not use fileservice to detect folders
23 Editors: revisit "preview" editors default behaviour
24 Not Getting Auto Information
25 Readonly editor hover appears weird for first line
26 Semantic Highlighting can't be disabled
27 Allow filter in quick navigation (Ctrl-Q)
```

VScode中PR的title (仓库地址: https://github.com/NJU-Software-Requirement/Software Requirement EXP1/blob/master/data/vscode_pr_title.txt) :

8125 lines (8125 sloc) | 344 KB

```
1 Create FUNDING.yml
2 revert fix#47400
3 cmd+e on empty selection should feed the word under cursor to find input
4 Fixes non-themed comment body border
5 Merge configuration defaults of language overrides
6 Joh/extbisect
7 fix #109733 prevent double '(read-only)' suffix in titlebar
8 Improve spelling of items relating to nls.localize()
9 Support clipboard write method to support other formats
10 Fix scope checking mismatch
11 typo: tanslations -> translations
12 Fixed typo 'sequece' to 'sequence'
13 Reconnect terminals at startup instead of waiting for the terminal view to become visible
14 Support git worktree directory detection
15 Linux Snap: Prevent GDK_PIXBUF env from leaking out
16 #109255 - Add encoding description to settings "File: Encoding" drop-down
17 Link tsconfig compiler option names to reference pages
18 add 'deleteWordEntire' command
19 New accessibility feature: Dark+ (color blind) theme. (related to issue: #109132)
20 UI items are incorrect order when applying the language pack
21 Use find-yarn-workspace-root over find-yarn-workspace-root2
22 notebooks: add generic colors for renderers
23 use window as background for -ms-high-contrast
24 debt - remove updateBackupPath method from environment service
25 Firefox: avoid menu actions firing twice
26 Changed shebang same as code.sh
27 Use a custom dialog for proxy authentication
28 Fixes middle-clicking on a notification sometimes pastes the contents of the clipboard
29 Better handling of configuration keys with objects in context keys
30 Reduce host service dependencies
31 Send `vscode.workspace.onDidCreateFiles` events for folders #109088
--
```

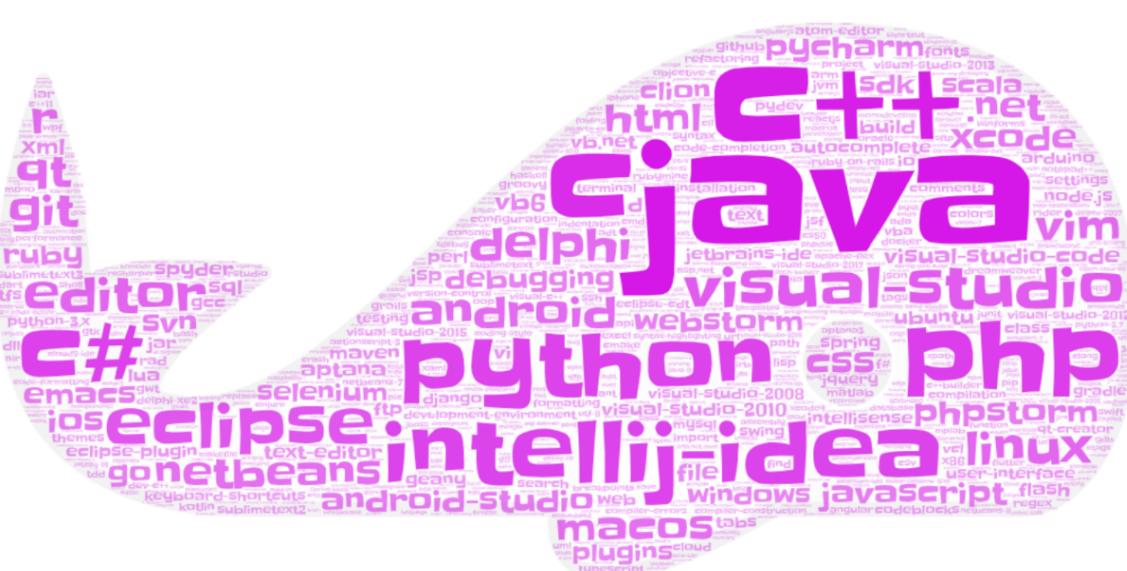
IDEA中PR的title (仓库地址: https://github.com/NJU-Software-Requirement/Software_Requirement_EXP1/blob/master/data/idea_pr_title.txt) :

1384 lines (1384 sloc) | 73 KB

```
1 Add support for parametrized model builders injected from the IDE
2 IDEA-229823: Add more YAML settings to editorconfig
3 IDEA-252328: Consider Java installations from Gradle
4 IDEA-253543: Include annotations in ChangeParametersRequest
5 Replace size zero check with 'isEmpty'
6 Fix typo
7 IDEA-193572 Detect hooksPath from git config
8 Add Invoker#forBackgroundPoolWithoutReadAction to allow parallel exec...
9 reuse Gradle connections slightly less
10 Remove redundant .toString()
11 IDEA-252935 Lookup class from MethodLocation if available
12 Bump junit from 4.12 to 4.13.1 in /plugins/lombok/test-manual
13 Merge roots changed notifications from SourceFolderManager
14 IDEA-252593: Introduce TempDir macro
15 PY-44714: TypedDict: don't register problem with None
16 Add Slack shield
17 [IDEA-252000] Update firefox icon
18 Include libraries in use-scope
19 PY-40863 Highlight "{{", "}}" as escape sequence in fstrings
20 Fix order of the Trinity constructor assignment
21 IDEA-251257 Use selected classpath of run configuration for Gradle delegated application runs
22 Fix IDEA-251181 by upgrading gradle-idea-ext to 0.9
23 Revert usage of keymap.current.os.only key but set it to true by default
24 Dependency resolution fixes in Gradle composite builds.
25 Fix NPE in UsageViewImpl
26 IDEA-124383 save dialog should edited file name
27 File Colors Settings: Make new scopes available for colors configuration
28 Scopes Settings: Align files colors when added to scope
29 i18n: fix message key typo
30 Add trusted host system property to InstallPluginService
31 Fix IDEA-92793
```

对获取数据的初步分析

首先，我们对从Stack Overflow上爬取的含有“IDE”标签的问答进行了各个tag的统计，并做了一个词云图，展示如下：



可以看到，一些知名的开发工具（如Eclipse、Visual-studio、Pycharm、Visual-studio-code、Android-studio等）和一些编程语言（如Java、C、C++、C#、Python、PHP等）在Stack Overflow上伴随“IDE”问答出现的频率极高。

这是在我们预期之内的，一方面，各知名开发工具，即用户常用的IDE，有各自的特点和功能，另一方面，各常用编程语言由于各自的特性，各有自适配的开发环境和开发工具。虽然这些关键词不能给我们展示很直观的Smart IDE的需求，但是它给我们透露了以下信息：

1. Smart IDE需要能够配置各种编程语言的开发环境（包含针对专一语言的IDE的环境配置，如IntelliJ-IDEA，和针对多种语言的IDE的不同环境配置，如VScode）。
 2. 出现频率越高的词显然越值得我们去关注，这即代表该相关语言或开发工具的使用者更多，也代表其存在可改进可完善的内容也更多，因此在后续需求的进一步获取和分类、分析时，我们可以围绕这些关键词方向进行内容的展开。

在筛选掉关于开发工具和编程语言的标签后，对剩下的词生成词云图如下：



可以观察到，系统的名称成了高频词（像Linux、MacOS、Windows等），这直观反映了IDE需要对不同系统进行适配的需求。程序员会使用不同的计算机操作系统进行软件开发，因此Smart IDE需要对各种常用于软件开发的操作系统推出不同的版本，以满足各操作系统用户的使用需求。

另外，除去这些操作系统的名称后，剩下的词语几乎都能体现出IDE的一些具体需求了。引入眼帘的editor反映出IDE的文本编辑相关需求；debugging也是程序员常见的痛处，如果没有易于使用的调试工具、调试器，或许就不会有程序员使用这款(Stupid) IDE了；还有关于git的词语也是频频出现，版本控制已经是目前程序员开发项目工作中不可或缺的一角，Smart IDE确实应该有关于使用git这一分布式版本控制系统的需求。

实验中，我们还利用了wordToVec，对爬取的带有“IDE”标签的问答内容进行词汇向量化，训练了一个模型。根据上面的词云图，我们选出了一些可能有用的关键词，对文本聚类分析后，得到以下结果，部分展示如下。

```

editor
[('like', '0.9999887452888489), ('support', '0.99998823079872131), ('or', '0.9999881220817566), ('PHP', '0.9998782277107239), ('into', '0.9998772749364075), ('different', '0.9998723864555359), ('has', '0.9998085717582183), ('id', '0.999868353164673), ('html', '0.999866386781688), ('which', '0.9998562339816128), ('function', '0.999849954799387), ('does', '0.9998618968388554), ('key', '0.9998601675933569), ('Google', '0.99985333838331), ('windows', '0.99985856201179), ('running', '0.9998583793640137), ('other', '0.999858021736145), ('web', '0.9998581117582183), ('JavaScript', '0.999847769372437), ('configuration', '0.99985856225738525), ('as', '0.9998528957366943), ('feature', '0.99984736965212), ('What'), '0.999847769372437), ('closing', '0.9998474717148198), ('vim', '0.9998469325722168), ('Delphi', '0.999846875667572), ('Best', '0.9998463988384138), ('Javascript', '0.9998462280164795), ('IDEs', '0.9998459219932565), ('tabs', '0.9998459219932565), ('language', '0.999845882783961), ('over', '0.9998454561568974), ('form', '0.9998382994052673), ('IDE?', '0.999837809918184), ('PyCharm', '0.99983186994289), ('through', '0.9998326876721155), ('GUI', '0.999831949017334), ('external', '0.9998340818643865), ('while', '0.99983322258362), ('text', '0.999833226203915), ('PyCharm', '0.999833186994289), ('sime', '0.9998357892836436), ('internal', '0.9998340818643865), ('while', '0.99983322258362), ('text', '0.999833226203915), ('PyCharm', '0.999833186994289), ('through', '0.9998326876721155), ('GUI', '0.999831949017334), ('side', '0.999838482181274), ('tools', '0.999838422888379), ('source', '0.99983842459161797), ('Source', '0.999829415665283), ('equivalent', '0.9998283982276917), ('Not', '0.9998264987878588), ('editing', '0.999826120651245), ('debugging', '0.99982619285835), ('library', '0.9998258352279663), ('remote', '0.999825168168768), ('than', '0.9998254179954529), ('functions', '0.9998252987861633), ('inside', '0.999815185884), ('Find', '0.999825193972630919), ('opening', '0.999825193972630919), ('Change', '0.9998397231179876), ('Custom', '0.9998397231179876), ('ws', '0.9998397231179876), ('app', '0.99982331828857), ('solution', '0.9998219813214111), ('both', '0.999821536935425), ('lines', '0.9998205908192261), ('single', '0.999820351608647), ('inside', '0.99981891515884), ('Find', '0.9998193972630919), ('opening', '0.9998193972630919), ('Change', '0.9998193972630919), ('Custom', '0.9998193972630919), ('ws', '0.9998193972630919), ('app', '0.99981784285564), ('Mac', '0.999816945125), ('Cannot', '0.99981634926652), ('netbeans', '0.99981447297366), ('two', '0.9998157024383545), ('Keyboard', '0.9998157024383545), ('more', '0.9998152256611963), ('IDEs', '0.999814978126172), ('system', '0.9998147487640381), ('Builder', '0.999813566711426), ('PHPStorm', '0.9998123849738623), ('creating', '0.999812851643127), ('supports', '0.999812866555623), ('me', '0.999818576438938), ('Web', '0.999818576438938), ('++', '0.999818576438938), ('based', '0.999818576438938)] git
[('window', '0.99987902526855), ('application', '0.9998530149459839), (':', '0.999852418899536), ('server', '0.999847948551178), ('menu', '0.999847948551178), ('Error', '0.999847948551178), ('after', '0.9998465776443481), ('while', '0.9998465510864285), ('class', '0.9998445510864285), ('errors', '0.9998437166213998), ('not', '0.999841166213998), ('source', '0.99983842459161797), ('Source', '0.999829415665283), ('equivalent', '0.9998283982276917), ('Not', '0.9998264987878588), ('editing', '0.999826120651245), ('debugging', '0.99982619285835), ('library', '0.9998258352279663), ('remote', '0.999825168168768), ('than', '0.9998254179954529), ('functions', '0.9998252987861633), ('inside', '0.999815185884), ('Find', '0.9998193972630919), ('opening', '0.9998193972630919), ('Change', '0.9998193972630919), ('Custom', '0.9998193972630919), ('ws', '0.9998193972630919), ('app', '0.99981784285564), ('Mac', '0.999816945125), ('Cannot', '0.99981634926652), ('netbeans', '0.99981447297366), ('two', '0.9998157024383545), ('Keyboard', '0.9998157024383545), ('more', '0.9998152256611963), ('IDEs', '0.999814978126172), ('system', '0.9998147487640381), ('Builder', '0.999813566711426), ('PHPStorm', '0.9998123849738623), ('creating', '0.999812851643127), ('supports', '0.999812866555623), ('me', '0.999818576438938), ('Web', '0.999818576438938), ('++', '0.999818576438938), ('based', '0.999818576438938)] git
[('window', '0.99987902526855), ('application', '0.9998530149459839), (':', '0.999852418899536), ('server', '0.999847948551178), ('menu', '0.999847948551178), ('Error', '0.999847948551178), ('after', '0.9998465776443481), ('while', '0.9998465510864285), ('class', '0.9998445510864285), ('errors', '0.9998437166213998), ('not', '0.999841166213998), ('source', '0.99983842459161797), ('Source', '0.999829415665283), ('equivalent', '0.9998283982276917), ('Not', '0.9998264987878588), ('editing', '0.999826120651245), ('debugging', '0.99982619285835), ('library', '0.9998258352279663), ('remote', '0.999825168168768), ('than', '0.9998254179954529), ('functions', '0.9998252987861633), ('inside', '0.999815185884), ('Find', '0.9998193972630919), ('opening', '0.9998193972630919), ('Change', '0.9998193972630919), ('Custom', '0.9998193972630919), ('ws', '0.9998193972630919), ('app', '0.99981784285564), ('Mac', '0.999816945125), ('Cannot', '0.99981634926652), ('netbeans', '0.99981447297366), ('two', '0.9998157024383545), ('Keyboard', '0.9998157024383545), ('more', '0.9998152256611963), ('IDEs', '0.999814978126172), ('system', '0.9998147487640381), ('Builder', '0.999813566711426), ('PHPStorm', '0.9998123849738623), ('creating', '0.999812851643127), ('supports', '0.999812866555623), ('me', '0.999818576438938), ('Web', '0.999818576438938), ('++', '0.999818576438938), ('based', '0.999818576438938)] debug
[('build', '0.9999116791915893), ('automatically', '0.999911467334743), ('can', '0.9999135732653787), ('output', '0.9999115467815133), ('Xcode', '0.9999114274978638), ('highlight', '0.999911659328369), ('multiple', '0.999989520149231), ('about', '0.999989409039414), ('me', '0.9999889241207832), ('view', '0.9999877320898877), ('Eclipse', '0.999971359634399), ('cursor', '0.999986718780265), ('be', '0.9999840113822937), ('type', '0.9999836192893982), ('if', '0.999981158943176), ('this', '0.99998099686850281), ('property', '0.999980996868618842), ('property', '0.9999827549744), ('PyCharm', '0.9999879105224689), ('what', '0.99998791313174), ('custom', '0.9999825451736848), ('import', '0.99998981356620879), ('function', '0.999971223831177), ('Project', '0.9999870622784729), ('all', '0.999982331828857), ('PhPStorm', '0.9999822473678587), ('folder', '0.9999822162692832), ('missing', '0.9999821364879882), ('variable', '0.99998213852749634), ('Cannot', '0.999821184656738), ('error', '0.99982064645368), ('SDK', '0.999818384645368), ('NetBeans', '0.999817289962769), ('Server', '0.999817289962769), ('control', '0.999817499576978), ('about', '0.9998168349266852), ('Javascript', '0.999816753219684), ('methods', '0.999816298484822), ('Auto', '0.999815832298649), ('), '0.9998149871826172), ('vs', '0.9998148679733276), ('app', '0.9998125433927814), ('editing', '0.9998124838757566), ('Using', '0.9998114756885754), ('web', '0.999819993972772773), ('The', '0.9998192191123962), ('won', '0.999818623065942432), ('Android', '0.999818623065942432), ('SDK', '0.999818623065942432), ('format', '0.999888134391785), ('version', '0.999887842298899), ('has', '0.999887357888895), ('method', '0.999886425323468), (')', '0.999886425323468), ('Debugging', '0.999886425323468), ('7368229494), ('same', '0.999886425323468), ('classes', '0.9998844967651367), ('options', '0.9998842583465576), (')', '0.9998842583465576), ('help', '0.999884019279785), ('up', '0.999883908718698), ('over', '0.999883721947546), ('multiple', '0.999883238815386), ('Debug', '0.9998830662536261), ('Running', '0.99988282735083)], ('format', '0.99988282735083)]
[('build', '0.9999116791915893), ('automatically', '0.999911467334743), ('can', '0.9999135732653787), ('output', '0.9999115467815133), ('Xcode', '0.9999114274978638), ('highlight', '0.999911659328369), ('multiple', '0.999989520149231), ('about', '0.999989409039414), ('me', '0.9999889241207832), ('view', '0.9999877320898877), ('Eclipse', '0.999971359634399), ('cursor', '0.999986718780265), ('be', '0.9999840113822937), ('type', '0.9999836192893982), ('if', '0.999981158943176), ('this', '0.99998099686850281), ('property', '0.999980996868618842), ('property', '0.9999827549744), ('PyCharm', '0.9999879105224689), ('what', '0.99998791313174), ('custom', '0.9999825451736848), ('import', '0.99998981356620879), ('function', '0.999971223831177), ('Project', '0.9999870622784729), ('all', '0.999982331828857), ('PhPStorm', '0.9999822473678587), ('folder', '0.9999822162692832), ('missing', '0.9999821364879882), ('variable', '0.99998213852749634), ('Cannot', '0.9999821184656738), ('error', '0.999982064645368), ('SDK', '0.9999818384645368), ('NetBeans', '0.9999817289962769), ('Server', '0.9999817289962769), ('control', '0.9999817499576978), ('about', '0.99998168349266852), ('Javascript', '0.9999816753219684), ('methods', '0.9999816298484822), ('Auto', '0.9999815832298649), ('), '0.99998149871826172), ('vs', '0.99998148679733276), ('app', '0.99998125433927814), ('editing', '0.99998124838757566), ('Using', '0.99998114756885754), ('web', '0.9999819993972772773), ('The', '0.99998192191123962), ('won', '0.9999818623065942432), ('Android', '0.9999818623065942432), ('format', '0.999988134391785), ('version', '0.999987842298899), ('has', '0.999987357888895), ('method', '0.999986425323468), (')', '0.999986425323468), ('Debugging', '0.999986425323468), ('7368229494), ('same', '0.999986425323468), ('classes', '0.9999844967651367), ('options', '0.9999842583465576), (')', '0.9999842583465576), ('help', '0.999984019279785), ('up', '0.999983908718698), ('over', '0.999983721947546), ('multiple', '0.999983238815386), ('Debug', '0.9999830662536261), ('Running', '0.99998282735083')]]
```

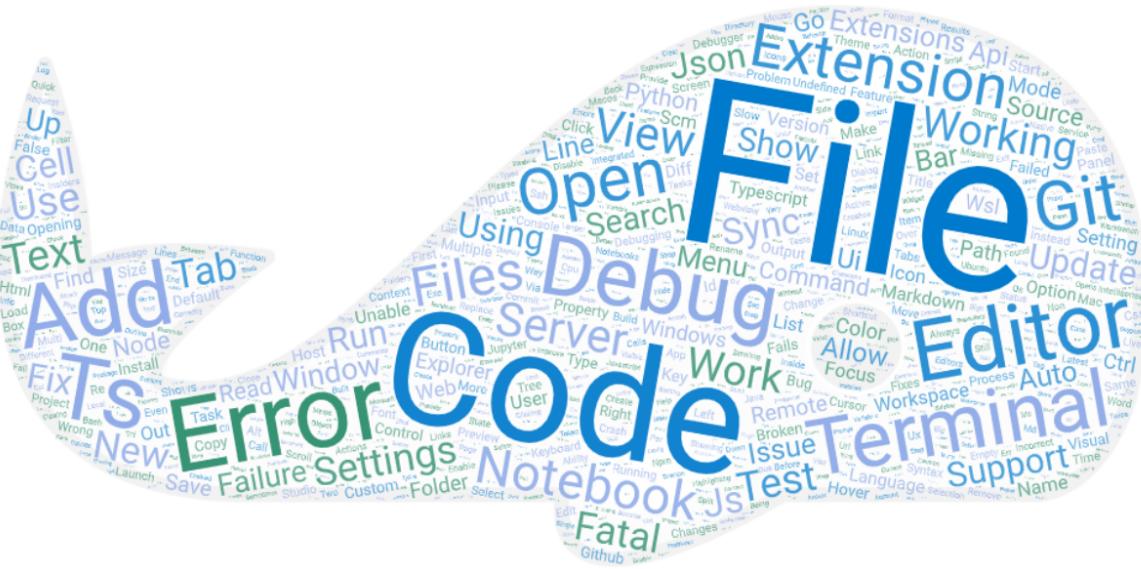
词向量聚类结果可以在我们的仓库中查看，由于内容过多，暂不一一截图展示（仓库地址：https://github.com/NJU-Software-Requirement/Software Requirement EXP1/blob/master/data/ide_ques vec.txt）

遗憾的是，结果中的相关词汇很大部分都不能体现出IDE的相关需求，但我们做了一部分筛选后，也能发现一些反映需求的词汇，比如：

在editor类的向量相关词语中，可以发现remote（IDE需要支持远程编辑）、slow/quick（IDE需要快速打开编辑器）、auto-save（IDE的编辑器需要有自动保存的功能）、copy/selected（IDE编辑器支持复制、粘贴、选中等基本文本编辑操作）等。

接着，我们把爬取的VScode的issue内容也做了类似的处理。

我们发现爬取的数据不是所有内容都带有label，所以我们将爬取的数据做了分类，对于没有label的issue，直接采用title生成了词云图，展示如下：



对于带有label的issue，我们用label生成词云图，展示如下：



从issue生成的第一个词云图中可以看出几个比较重要的关键词：File, Code, Terminal, Editor。这些都是有关软件功能的词汇，使用过vscode的同学应该都有体会，vscode的文件交互系统做的非常合理易用且功能强大。说明我们在开发一款新的IDE时，也应该充分考虑到文件的管理。而Code和Editor是最主要的文本编辑功能，这也是我们开发IDE软件的核心业务需求方向，在本项目之后的工作中，应该重点关注其他成功IDE有关需求和用户需求。最后对于Terminal，尽管现在多数IDE都是GUI，但是有些时候Terminal会比GUI更加有用，所以对于Terminal的功能设计分布也是本项目的一个侧重点。

第一张图是从IDE功能本身出发进行总结的，下面我们从用户需求角度出发进行讨论。从issue生成的第二个词云图中可以看出用户对于IDE的需求的一些关键词debug ,UX, verified也就是调试程序，程序员最需求的功能，UX也就是用户体验，当然这个词太过于宽泛，接下来的工作我们会进行更加细致的讨论。最后是verified，说明程序员对于安全性也有比较多的需求。

根据词云图的显示，我们选取了一些可能有用的关键词对爬取的issue内容进行词向量聚类分析，得到以下结果：

(同上，聚类结果不便于展示，可在仓库中查看保存结果的文本文件：https://github.com/NJU-Software-Requirement/Software_Requirement_EXP1/blob/master/data/vscode_issues_vec.txt）从词向量的聚类结果中，我们观察得到的相关词汇，提取、分类了一些可能的需求，将在下文的结果中列举。

另外，我们把爬取的VScode的Pull Request和IDEA的Pull Request也做了类似的处理。展示如下





对于以上两种主流的IDE的PR内容生成的词云图，可以发现大部分高频词都是关于修复、添加、更新等软件版本更新相关的动词，具体体现为在版本更新中“修复了xxxx的bug”、“添加了xxxx的功能”、“更新了xxxx的内容”等，非常遗憾的是，我们小组现在掌握的NLP技术和机器学习技术并不能很好地帮助我们从中获取出更详细的需求，因此我们也使用了人肉分类器从众多的PR内容中提炼了分类了一些Smart IDE的潜在需求，具体内容会在下面的结论中展示。

结论

最终，我们对所有的实验得到的材料进行更详细的分类和分析，形成以下需求。

文本编辑类

- 支持文本编辑器的常用操作，包括选中、全选、复制、粘贴、撤销、重做。
 - 具备代码自动补全的功能。
 - 具备代码自动缩进的功能。
 - 具备文本内容查找的功能，在查找时，要给文本中出现的查询关键词刷上醒目的颜色。
 - 具备文本关键词替换的功能，此功能可基于查找功能实现，支持把查找到的关键词替换成新输入的关键词、支持批量替换、一键全部替换。
 - 支持语法高亮。
 - 可以同时开启多个窗口/多个文件进行编辑。
 - 多窗口同时打开同一个文件时进行文件内容的同步控制，具体做法可以有：

1. 禁止同一个文件同时被多窗口打开，如果同一个文件被多窗口二次打开，唤醒之前打开此文件的窗口；
 2. 允许多窗口同时打开同一个文件，当文件内容被某一窗口更改时，其他窗口也自动同步更改内容；
 3. 允许多窗口同时打开同一个文件，当文件内容被某一窗口更改并保存时，提醒用户当前有多个窗口已经打开此文件，让用户自己选择是否同步更改内容。

以上不同的控制可以由用户在Smart IDE的设置中自行设定。

- 对于不同的编程语言，自动补全易错语法，如C++、Java的开发自动补全遗漏的分号，Python自动补全遗漏的冒号，条件语句自动补全判断相等的双等号等。
- 支持编辑操作的自动保存，自动保存的频率由使用者自行设定。
- 复制粘贴大段代码的时候要保持代码的缩进状态。
- 鼠标指针停留在标识符上时，显示标识符的信息，如变量的类型，常量或宏的值（如果有的话），函数的参数列表及返回值。
- 对代码中明显的语法错误、待完善的函数或方法、未初始化的变量用特殊标记在文本中标出，如语法错误可以用红色波浪线在出错处下方标出，只声明、未定义的函数可以用绿色波浪在函数声明处标出。
- 检查英语单词的拼写，对出错词提供更正选择。
- 支持代码块的折叠。

调试类

- 支持常见的调试操作
 1. 支持断点、条件断点等
 2. 逐行运行
 3. 调试中显示当前变量
 4. 监视系统资源使用情况（如CPU）
- 提供独立的调试界面，包含以下内容
 1. 调试器中当前文件的大纲
 2. 函数、类列表
 3. 控制台
 4. 任务(task)
 5. 调试输出
 6. 变量视图，显示选定变量的运行时值。
 7. 断点列表
 8. 显示最新的运行时调用堆栈

用户界面类

- 用户界面主要包含以下多个视图
 1. 文本编辑器（细节见上文）
 2. 项目结构
 1. 包含该项目所有相关文件
 2. 能在该视图中对所列出文件进行基本的文件操作，如：复制、粘贴、剪切、重命名、删除、选定两个文件进行比较等
 3. 提供对文件排列的多种排序方式
 4. 能够搜索文件
 5. 提供过滤筛选操作
 3. 当前文档结构
 1. 展示类、函数、全局变量等元素
 2. 点击之后能够在文档编辑器中跳转到被选定元素所在位置
 3. 对元素进行排序
 4. 搜索
 5. 过滤筛选

6. 收起和展开（适用于一个元素包含多个元素的情况，如一个类中含多个函数）

4. 错误和警告

1. 显示错误和警告

2. 定位：点击条目后能够跳转到位置

3. 复制报错内容

4. 跳转到浏览器查询错误号或错误描述

5. 相应程序语言文档浏览界面

1. 显示当前程序语言对应的手册文档内容

2. 搜索功能

3. 点击文档中的链接能够跳转到其他条目或在浏览器中打开

· 支持个性化设置，用户可以对IDE的界面进行设置，选择自己喜爱的背景和风格；

· 支持用户自定义字体颜色、大小、样式，对代码的语法高亮也可以自定义颜色；

· 编辑器界面对文本的已保存内容和新增加内容用不同的颜色或字体背景色区分；

· 用户支持页面缩放；

· 支持多国语言。

编码支持类

· 设置每个项目的默认编码（用于从资源管理器视图中打开文件的编码方式）；

· 提供可用编码选项打开文件；

· 使用特定的编码或默认编码保存或重新打开当前文件；

· 显示当前文件编码方式；

· 当用户切换编码或者使用不正确的编码打开文件后给出警告。

插件类

· 支持加载官方或者是开发者开发的相应插件。

· 提供插件管理界面，能够选择删除、禁用。

· 自动检测插件版本，进行更新。

开发效率类

· 能够快速打开上次开发的（最近访问的）项目文件目录，定位到上次修改代码的光标处。

· 能够自动检测文件的编程语言，自动配置不同语言的编程环境。

· 支持多种便捷的快捷键、组合键，且可以由用户自定义设置。

· 能够关联git账户，便于用户进行版本控制，需要保证用户账户的安全。

可移植性

- 支持多语言环境下的文本分析。
- 具有各主流操作系统对应的使用版本，包括移动终端操作系统。