一、

arduino 
studio build module 
apache 
web 
type 
github 
edit6il 
v Irm 
data 
template 
auto 
python 
raspberry 
embedded 
upwuser 
android 
typescript 
firebase 
directory 
visÜa1nnstudio 
intel 
dynamic 
wqnclow 
inte 1 i J 
plugin 
installation 
[cleaerru 
virtual 
vba 
commi 
emacs 
linport E 
lisp 
environment 
for tran 
Invc 
xcodeID 
o 
spring 
deployment 
mobile 
sq 
compiler 
static 
scrip - 
webs torm 
memory 
file 
lirpux 
project 
net core 
access 
structure 
gradle 
excel 
.pa th Coo 
bash 
ou]ao 
inst 
jupyter 
keyboard 
j sapi 
Q) google react 
-esource 
connentscn 
ava 
php ser-ver- 
builder 
tool 
browser 
shortcut 
xamarln 
header character 

这是我对stackoverflow上的3000条提问所属于的标签的爬取，可以看到最醒目的是几个编程语言的名称（如java, python）、编程IDE的名称（如visual studio，eclipse，pycharm，netbeans）以及编程平台的名称（Android，Linux）。另外还有一个醒目的单词"debugging"，程序员确确实实会遇到相当多的debug困难。此外，还有关于web开发关键词web。可以猜测出，IDE要支持几种编程语言，并且要有一定的debug工具。

次醒目的是几个次常用的编程语言，如js，net，php；几个常用的代码开发框架，如maven，gradle，spring；这里可以看出IDE的基本需求肯定是支持语言。并且，要能够支持提供代码框架节省开发时间。

再细心一些，我们会发现用关于版本控制的标签，比如git,github，可能还用cloud等等。这里就可以看出IDE需要提供云存储、版本控制的功能。

再仔细看，我们可以看到比如engine,designer(有可能是QT designer等）,interface,builder。有些是开发引擎，有些是构建器（比如Qt的builder），有些是界面、接口。这些可能体现出比较高级的需求。根据这些关键词查到对应的句子，我们可以发现：

1.有关于engine的句子

Engine for simple client-server two-player turn-based game with random match-making and online statistics

这是程序员有开发游戏引擎的需求，他们需要IDE提供一些现成的方便改进的引擎

WebStorm syntax highlighting for Swig template engine

这是js模板引擎

2.有关于designer的句子

VB6 designer doesn't display opened modules

这是VB6提供的设计器，问答显示他的软件出现了双击打不开模块的bug

这些能体现一些IDE比较smart的部分，他们需要提供比较丰富的组件

3. 我们还能从词云图当中看到mysql，Oracle,nosql等词汇。其中有一条十分简洁的问题是这样的：

How to connect mysql with eclipse?

说明IDE应该能够支持一些访问数据库的操作，提供访问数据库的接口。

is there a valid extension for connect **mysql** via ssh on visual code studio? I can't find nothing!; I've been tryed with **Mysql** managment tool but seem there are not options for ssh …

可以看出vscode对数据库的访问支持还不够，此需求可以作为smart需求的一部分

我们还可以发现vscode与数据库连接的操作问题比较多：

 i'm not able to connect new version of vscode and new version of mysql

Suggestions for NodeJS+MySQL

Connect to a MySQL datatbase using ssl certificates in using C# Dotnet Core 3

vs code error when i run some simple mysql connecting code with code runner

…

除去连接问题以外，我们还能看到一些直接的建议：

[Suggestions for NodeJS+MySQL](https://stackoverflow.com/questions/46902076/suggestions-for-nodejsmysql?r=SearchResults)

得出数据库对语言的支持还需要完善

4.在前几次课的讨论访谈中，我们也讨论到IDE对于插件的支持。在此图中出现了extension这个字眼，其中的需求有：如何导入插件？如何离线下载插件？等等

5.此外，还有有关于网络编程，web开发等问题。大量人工查看这些问题比较困难，于是我对标签进行了词云分析，并且对所有问题文本使用单层神经网络的word2vec工具进行了对关键词的聚类分析。（第二部分）

二、我们确定的开源IDE项目是VSCode（准确地说VSCode是一款编辑器，但其插件十分丰富），因此我更换了对Stackoverflow问题的爬取策略，关键词设置为[visual-studio-code]。爬取大量文本后，获得所有的标签词云图，通过标签词云图得到我们的分类标准，然后使用思路二中的聚类分析，对需求进行分类和总结：

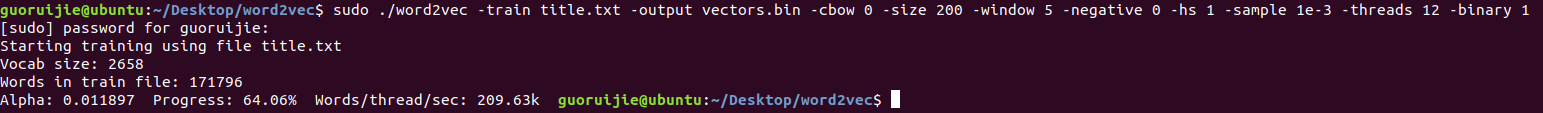
图片包含 报纸, 文字

描述已自动生成

根据观察这个词云图，去除一些对分类无用的关键词，并对这些关键词先进行观察性的分类，我最终确定对需求的分类准则有以下几条：

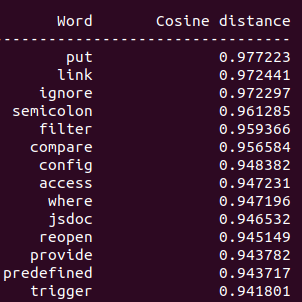
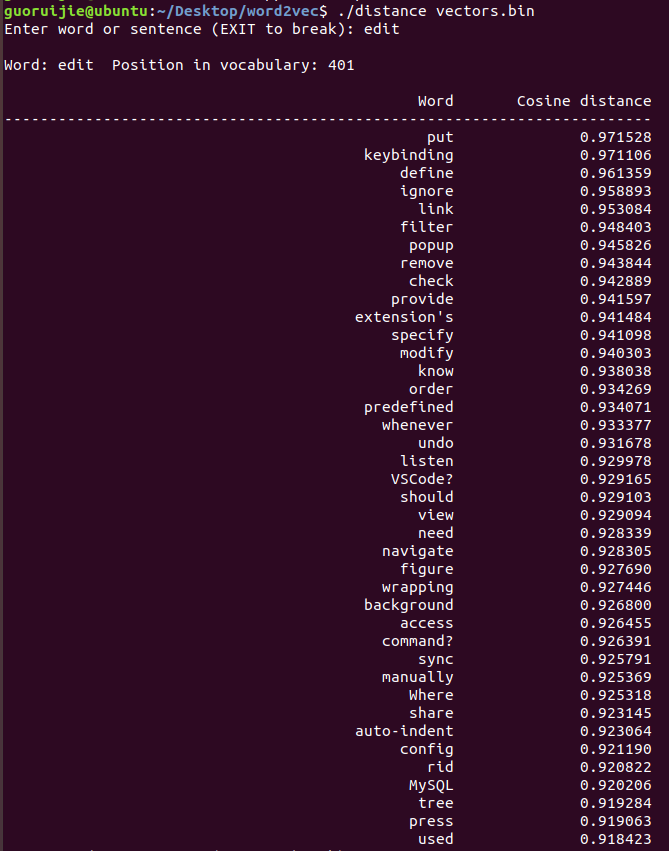
编辑器和语言 嵌入式 网络编程 debug工具 插件 版本控制 智能（自动）数据库支持

使用word2vec工具对文本所有词汇向量化，并进行训练：



训练完成，进行聚类分析。聚类分析的相关词汇输出结果中并非所有词汇都令人满意，因此我将其结果

1. 编辑器和语言。首先是编辑器(editor, input, edit)



总结关于编辑器的功能如下：

Auto-indent，要求自动缩进

Background，要求能够设置背景颜色

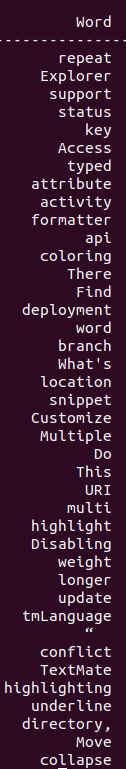
Semicolon，需要对缺少的分号进行提示

Tree，需要能够展示当前项目目录树

Navigate，需要有使用导航（帮助）

Undo，支持撤回，重做（附带的还有一系列的文本编辑功能，复制粘贴等等）

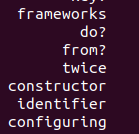
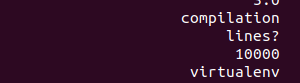
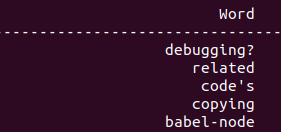
关于语言(language)



通过这一栏可以看出，相关性比较有价值的词有textmate highlight underline Find color等。因此还是属于编辑类范畴，要求编辑器能够支持：

高亮语法，编辑工具，下划线，文本查找，文字涂色

1. 嵌入式开发(embed)



嵌入式开发需要支持：

Debug工具(debugging)

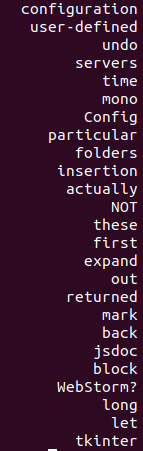
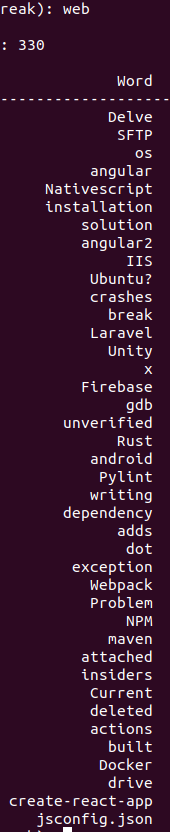
嵌入式开发编译器(compilation)

虚拟开发环境(virtual environment)

开发框架(frameworks)

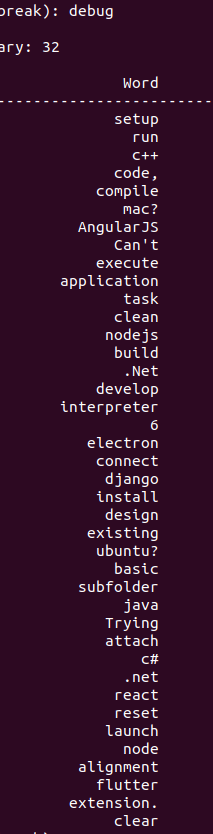
必要的配置工具(configuring)

1. 网络编程（web,net,.net,html）



支持js框架，支持调试器(如gdb)。还有支持不同平台(Android,ios)

1. Debug

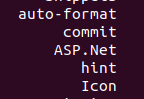
 对多种语言的支持（c++，.net,c#,nodejs）

有良好的扩展插件支持(extension)

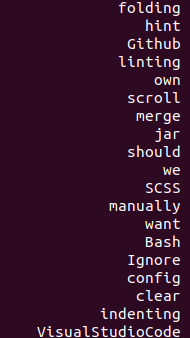
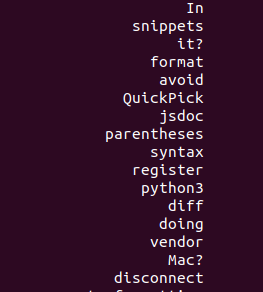
针对框架的debug能力(Django)

跨平台(ubuntu)

1. 插件(extension)（聚类分析结果不够理想）
2. 版本控制(version,git,github)

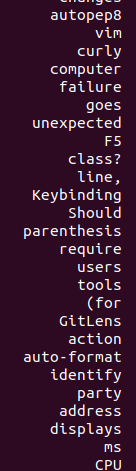
 自动格式化

1. 智能(smart,automatic)

auto-formatting（自动格式化），indenting（自动补全）.

Snippet是一个能够提高编码效率的插件，拥有极端强大的自动补全功能，可以考虑在插件列表中添加其获得相应的支持

1. 必要的工具

autopep,vim,keybinding,gitlen是一些呼声比较高的工具