

# 问题提交模块开发

作者：少林之巅

# 目录

1. 问题提交模块整体流程

2. 表结构设计

3. 敏感词过滤功能开发

4. 提交模块开发

# 问题提交模块简介

## 1. 问答提交模块前端页面

localhost:9090/#/ask

提问

标题

请输入标题

类别

生活

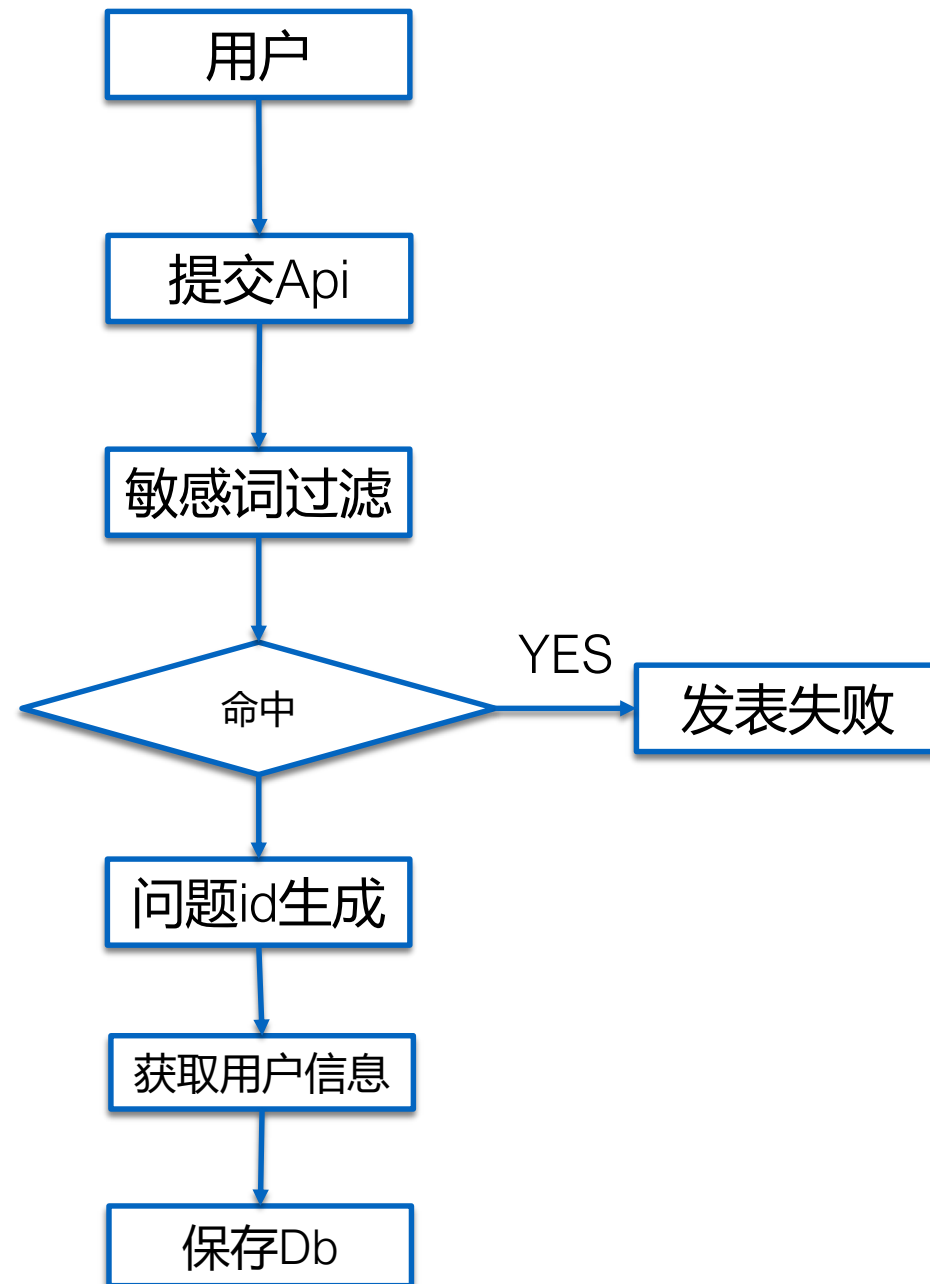
内容

请输入内容

提交

## 问题提交模块流程

### 2. 问题提交基本流程



## 表结构设计

### 3. 问题提问表结构设计

```
CREATE TABLE `question` (  
  `id` bigint(20) NOT NULL AUTO_INCREMENT,  
  `question_id` bigint(20) NOT NULL COMMENT '问题id',  
  `caption` varchar(128) COLLATE utf8mb4_general_ci NOT NULL COMMENT '问题标题',  
  `content` varchar(8192) COLLATE utf8mb4_general_ci NOT NULL COMMENT '问题内容',  
  `author_id` bigint(20) NOT NULL COMMENT '作者的用户id',  
  `category_id` bigint(20) NOT NULL COMMENT '所属栏目',  
  `status` tinyint(4) NOT NULL DEFAULT '1' COMMENT '问题状态',  
  `create_time` timestamp NULL DEFAULT CURRENT_TIMESTAMP COMMENT '创建时间',  
  `update_time` timestamp NULL DEFAULT CURRENT_TIMESTAMP ON UPDATE CURRENT_TIMESTAMP COMMENT '更新时间',  
  PRIMARY KEY (`id`),  
  KEY `idx_author_id` (`author_id`),  
  KEY `idx_question_id` (`question_id`),  
  KEY `idx_category_id` (`category_id`)  
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_general_ci;
```

## 表结构设计

### 3. 问题分类表结构设计

```
-----  
CREATE TABLE `category` (  
  `id` int(11) NOT NULL AUTO_INCREMENT,  
  `category_id` int(10) unsigned NOT NULL,  
  `category_name` varchar(128) COLLATE utf8mb4_general_ci NOT NULL,  
  `create_time` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP,  
  `update_time` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP ON UPDATE CURRENT_TIMESTAMP,  
  PRIMARY KEY (`id`),  
  UNIQUE KEY `idx_category_id` (`category_id`),  
  UNIQUE KEY `idx_category_name` (`category_name`)  
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_general_ci;
```

## 敏感词过滤模块开发

### 3. 敏感词过滤模块

- a. 采用正则表达式进行过滤和替换
- b. 采用trie树进行过滤和替换

## 敏感词过滤模块开发

4. 性能分析，假设敏感词库100w个，检测目标串，为：“我操你妈，多读书”

a. 正则表达式性能分析

i. 对应单次替换，时间复杂度为 $O(n*m)$

ii. 整体复杂度为， $100w * O(n*m)$

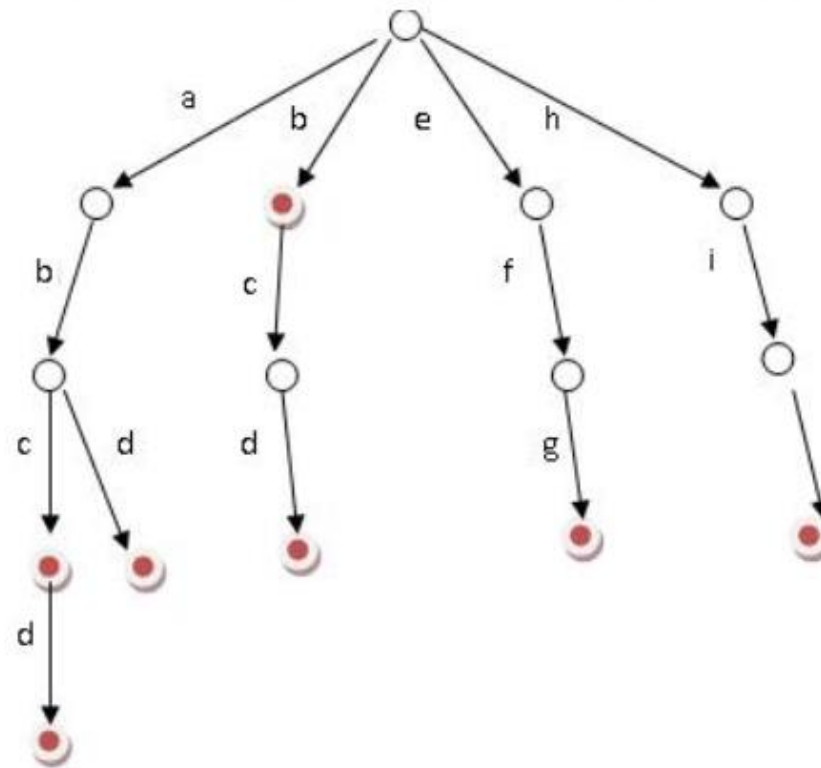


# 敏感词过滤模块开发

## 5. Trie工作原理

### 1、基本概念

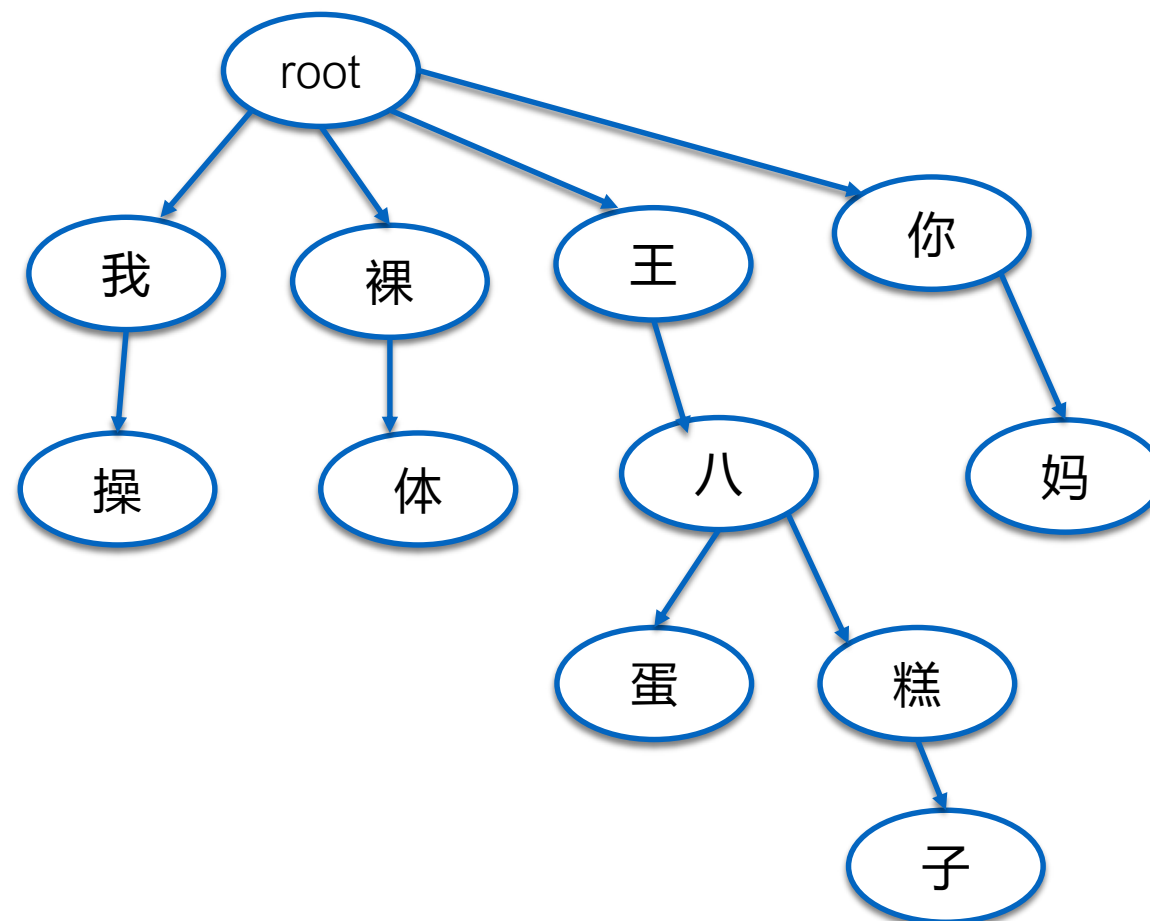
字典树，又称为单词查找树，Trie数，是一种树形结构，它是一种哈希树的变种。



## 敏感词过滤模块开发

### 6. Trie性能分析 “我操你妈，多读书”

- i. 对于每个字符为起点，需要判断该字符开始的字串，是否命中敏感词
- ii. 整体复杂度为， $O(n*m)$



## 敏感词过滤模块开发

### 7. 基于trie树敏感词过滤模块开发

a. Trie树构造

b. 词库加载

c. 敏感词过滤和替换

## 敏感词过滤模块开发

### 8. 基于正则表达式的敏感词过滤模块开发

a. 词库加载

b.敏感词过滤和替换

## 敏感词过滤模块开发

9. 提交api开发，集成敏感词过滤模块

a. 集成敏感词过滤模块

b. 集成账号系统，保存数据库