Atitit.全文检索的解决方案概览overview

|  |  |
| --- | --- |
| Author撰写人： | ***艾龙 (Attilax Akbar***  ***)*** |
|  |
| ver当前版本： | **V1.0** |
| Pub date发布日期： | Q09 (2016-9-9) |
| seclev密级 | Top secret **机密文件内部使用** |
| id文档标识号 | Oa\_v1.0\_q09 |

版本历史

|  |  |  |  |
| --- | --- | --- | --- |
| 版本 | 修改内容 | uke日期 | 日期 |
| V1 | 初稿 | Q09 | 2016-9-9 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

作者信息

艾龙 (Attilax Akbar al Rapanui  )

Tel电话：18573344543

Qq ：1466519819

Wechat微信 ：attilax

[1. 一、功能模块概览 4](#_Toc29795)

[1.1. 业务功能 4](#_Toc635)

[1.2. 全文检索 4](#_Toc2095)

[1.3. 分词 4](#_Toc4939)

[1.4. 分词主要流程 4](#_Toc22719)

[1.5. 分词的属性如下表 4](#_Toc25020)

[1.6. 全文检索 4](#_Toc20907)

[1.7. 增量索引文件 5](#_Toc23645)

[2. Ui设计与功能文档 5](#_Toc31540)

[3. 项目的一些原则 5](#_Toc13975)

[3.1. Isv金字塔模型 5](#_Toc32424)

[4. dhv 5](#_Toc12795)

[4.1. Dsl 6](#_Toc18143)

[4.2. Hybrid化,集成化 6](#_Toc18247)

[4.3. Vm化，隔离化 6](#_Toc26054)

[5. 5c 6](#_Toc21584)

[5.1. Crossplatform跨平台， 6](#_Toc28431)

[5.2. 跨设备cross device， 6](#_Toc24173)

[5.3. 跨语言cross lang， 6](#_Toc14710)

[5.4. 跨架构bs cs 6](#_Toc13599)

[5.5. 跨api cross api 6](#_Toc17227)

[6. 3s 6](#_Toc24323)

[6.1. Scalable可扩展性 6](#_Toc7191)

[6.2. secury 6](#_Toc19286)

[6.3. OO五大原则solid 6](#_Toc1174)

[7. 2M 7](#_Toc21649)

[7.1. Meta元编程 7](#_Toc16223)

[7.2. MultiFun多功能 7](#_Toc17637)

[8. HIF 7](#_Toc8177)

[8.1. H5，ui h5化 7](#_Toc21411)

[8.2. Ioc容器化 7](#_Toc4861)

[8.3. FURPS模型 7](#_Toc2768)

[9. 项目的架构，框架，配置与环境envi 7](#_Toc29218)

[9.1. 运行环境 7](#_Toc15994)

[9.2. 开放架构体系java 7](#_Toc18044)

[9.3. 绿色版无需繁琐安装 8](#_Toc19331)

[9.4. 跨平台跨设备 8](#_Toc30900)

[9.5. 跨bs cs hybrid架构 8](#_Toc16898)

[9.6. 界面h5化dsl化 8](#_Toc8811)

[9.7. 分词引擎ik,mmseg等 8](#_Toc32615)

[9.8. 全文引擎Lucene 8](#_Toc5183)

[10. 开发环境 与 工具 以及技术框架dev 9](#_Toc7940)

[10.1. Ide Eclipse 9](#_Toc31931)

[10.2. Jdk8 9](#_Toc7090)

[10.3. 界面ui技术H5 9](#_Toc16361)

[10.4. 容器ioc使用guice 9](#_Toc5021)

[11. 存储文档strory doc 9](#_Toc19560)

[12. 安全性secury 9](#_Toc26499)

[13. 性能perf 9](#_Toc25133)

[14. 接口Interface 9](#_Toc30959)

[15. Other其他 10](#_Toc8284)

# 一、功能模块概览

## 业务功能

参见业务方面的文档

## 全文检索

## 分词

## 分词主要流程

字符切分，对比字典词典得到分词，并查阅词典，标注词性。完善其他属性，输出

## 分词的属性如下表

|  |  |  |
| --- | --- | --- |
| Col | 列数 |  |
| Filename | 所属文件名 |  |
| Line | 行数 |  |
| Text | 分词 |  |
| Type | 词性 |  |

## 全文检索

索引的类型有Btree索引（分块索引）  Hash 索引（编号索引）  倒排索引

建立全文检索需要使用倒排索引才可。。

## 增量索引文件

# Ui设计与功能文档

参见详细设计

# 项目的一些原则

## Isv金字塔模型

第一层，Implements 层，着重与功能的实现。。

第二次，spec层，理论层，设计规范，接口，等。流程、方法论

顶层，val层，价值观层，原则，法则，定律等。

# dhv

## Dsl

## Hybrid化,集成化

## Vm化，隔离化

# 5c

## Crossplatform跨平台，

## 跨设备cross device，

## 跨语言cross lang，

## 跨架构bs cs

## 跨api cross api

# 3s

## Scalable可扩展性

## secury

## OO五大原则solid

# 2M

## Meta元编程

## MultiFun多功能

# HIF

## H5，ui h5化

## Ioc容器化

## FURPS模型

是功能（function）、易用性（usability）、可靠度（reliability）、性能（performance）及可支持性（supportability）

**最早是由惠普公司的罗伯特·格雷迪（Robert Grady）及卡斯威尔（Caswell）提出**

# 项目的架构，框架，配置与环境envi

## 运行环境

目前主要实现电脑版，预留未来的尽可能的跨平台跨设备(win,linux,android安卓，ios，iphone，ipad，wechat微信等）

## 开放架构体系java

Java体系资料繁多。开放体系，立足长远。。

Java体系还有个很好的优点就是，语言与安卓相同。。将来二期如果需要开发安卓手机版则会方便很多，源码移植会绝大部分通用（只需要注意，避免使用仅能在pc使用的技术，尽可能的考虑移动设备的通用性）。。

## 绿色版无需繁琐安装

基于java体系可以很容易做到绿色版，直接可将所有需要的基础类库sdk集成进去，无需单独安装依赖的基础类库。。

## 跨平台跨设备

目前主要实现电脑版，预留未来的尽可能的跨越不同的设备与平台

## 跨bs cs hybrid架构

项目可以运行为bs或者cs模式 以及hybrid模式。。几种模式均可运行

## 界面h5化dsl化

H5技术是目前mis oa系统中界面ui技术可以说最好的。。可以跨cs bs架构，以及卓越的跨平台跨设备特性。。为我们的未来和不远的将来预留最好的前景。。

相比native（winform ,swing,android ui,ios ui )的方案要适用很多。。同wpf相比，同为dsl，主要优点是胜在开放性与标准化以及跨平台与不同设备特性。

## 分词引擎ik,mmseg等

可以使用此集中流行的分词引擎，进行分词即可。。**Lucene默认集成的分词引擎对中文支持比较不太好，尽可能使用独立的分词引擎，也方便性能提升。。可以同步进行。。**

## 全文引擎**Lucene**

**目前Lucene是个很流行的全文引擎，基于此引擎即可实现我们的核心**

# 开发环境 与 工具 以及技术框架dev

## Ide Eclipse

## Jdk8

目前最新的jdk版本

## 界面ui技术H5

优秀的跨平台跨设备的 界面ui解决方案

## 容器ioc使用guice

轻便的ioc技术，方便跨设备使用

# 存储文档strory doc

参见存储详细设计文档

# 安全性secury

略，基础安全性即可

# 性能perf

略

# 接口Interface

略。。尽可能的为未来的第三方系统 跨平台，多设备预留足够的接口

# Other其他

略。