**设计文档**

**前端技术**

跨平台，包括web site, web app, android, ios, phoneGap做成hybrid app

**框架**

**Jquery Mobile**

**d3.js**

支持大多数浏览器，不支持的如IE6可通过 Node.js后端生成SVG文件，使用svgweb显示

**后端技术**

#### 数据抓取

Scrapy(Python)+MongoDB+raids

Crawler设计

1. 分两类，数据集中式网站和数据分散式网站

#### 数据结构

采用MongoDB存储，无模式，便于扩展和程序重构

* 抽象关键字 Abstract Keyword

主要指数据信息中会出现的高频词，如同比，这些关键字是分散式数据网站的数据抽取的训练集

* 数据类型 IndicatorType

基本数据类型信息

* 元数据MetaData

(最基本的数据单位，往往需要组合数据条件来取其中满足条件的数据显示，由于数量巨大，后期需要根据分类来创建不同的collection)

-数据结构[  id（自动生成），数据名称DataKeywords ]

-数据值[  区域，具体周期，主体1(Target1)，主体2(Target2)，数据时间，来源，数值s，更新时间，备注s，是否为预测值 ]

* 组合数据CombinedData：数据选择器，用来筛选系列数据

-组合条件s：区域条件，周期条件，主体1条件，主体2条件，数值条件，外部数据

-组合数据s:

-组合信息：id，名称，star，follow，comment，点击率，所属目录s，文字说明

-可选显示：图表s，地图，说明

* 数据类别：用来对数据进行分类Catalog

类别标识

类别名称

上级类别s

* 区域：用来描述数据所属地理区域范围Area

区域ID

区域识别号

区域名称

上级区域

* Target主体,Owner主体：数据的描述对象

-名称（）

-类型（行业，机构，用户，公司，国家，城市，私人等）

-类型值

-标识ID

-备注

**需求驱动开发**

**第一阶段，数据显示，基本的SN**

V1.0.0：数据显示功能

  分类显示数据

  数据使用多种图表

  指定网站的数据采集

V1.0.1: SN功能

  基本的用户注册，评论

  Star/Follow

  更多方式的排序，包括按时间，按人气

v1.02:数据比较功能

  多种数据比较

**第二阶段：数据的采集，显示，管理，完善SN**

v1.1.0

  互联网数据检索通用功能，人工干预

  用户自定义数据发布

V1.1.1

  数据分享的SN功能，如微博

  用户纠错介入数据纠错

v1.1.2

  大数据的融合

**第三阶段，框架化，外部资源整合，SN细化**

  提供API给其他网站使用

## 数据来源

### 第一阶段

数据抓取来源：

#### 已处理

世界银行  
[http://data.worldbank.org.cn/](http://data.worldbank.org.cn/" \t "_blank)

#### 未处理

[http://finance.sina.com.cn/mac/#](http://finance.sina.com.cn/mac/)

<http://calendar.hexun.com/global433_1.shtml>

<http://finance.stockstar.com/finance/macrodata/jijinkaihu.htm>

<http://www.cnnic.net.cn/>

<http://visualizing.org/data/browse>

http://news.qq.com/bigdata/

http://www.qianzhan.com/qzdata/

http://www.stats.gov.cn/tjsj/

http://data.baidu.com/index.html

http://datacenter.cngrain.com/  
  
http://www.askci.com/data/  
各国数据  
http://www.gapminder.org/data/  
搜房数据  
http://fdc.soufun.com/data/

<http://www.ebrun.com/data/>

# 其他

## 未被注册域名

seease.com

viszee.com