**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

Id标识

类型名称多语言[{Chinese:””},{English:””},…]

关键字（用于采集更新）:[]

说明：

外部数据URL:

数据发布者：对象id

组合数据ID:（加入新数据类型时如果没有对于的组合数据需要生成一个）

数据采集来源（SrcURL）:url + md5

* 元数据MetaData\_数据类型标识

(每个数据类型对应一个元数据collection,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)

-区域，对象1(Target1)，对象2(Target2) ，周期，数值s[时间，数值，更新时间，备注s，是否为预测值]

* 组合数据CombinedData：数据选择器，用来筛选数据

-组合条件s：[数据类型ID，区域条件，对象1条件，对象2条件，周期条件]

-组合信息：id，名称，star，follow，comments，views，所属目录s，文字说明，最近更新时间,组合类型(单一数据，组合数据)

-可选显示：图表s，地图

* 数据类别：用来对数据进行分类Catalog

id

类别name

类别name\_locs

上级类别s[]

* 区域：用来描述数据所属地理区域范围Area

ID

区域名本地化:[{Chinese:中国}]

上级区域

区域类型

* Target对象,Owner对象：数据的描述对象，发布对象

-id

-本地化名称[{Chinese:””,English:””}]

-类型（industry，org，user，company，country，city，person，indicator）

-备注

-相关URL：[]

**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

Line Chart:

|  |  |  |
| --- | --- | --- |
| 对象1数量\对象2数量 | 1 | N |
| 1 | 1Chart，1Line | 1Chart，NLine |
| N | 1Chart,NLine | NChart,NLine |

Bar chart + Pie chart

|  |  |  |
| --- | --- | --- |
| 对象1数量\对象2数量 | 1 | N |
| 1 |  |  |
| N |  |  |

Map chart

|  |  |  |
| --- | --- | --- |
| 对象1数量\对象2数量 | 1 | N |
| 1 |  |  |
| N |  |  |

  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)

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

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

#### 未处理

<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