

用户手册：

基于主页标题文字的
热点新闻跟踪与分析、可视化展示

开发小组

2014 年 6 月

用户手册

1 功能概述	3
2 使用方法	3

1 功能概述

目前新闻网站内部的新闻有比较好的分类，可以比较容易的收集关于同一个事件的相关新闻。但是收集不同门户网站的相关新闻就显得比较困难。而用户希望得到多个网站相关的新闻信息。所以将门户网站的新闻进行聚类整理就显得十分重要。

同时，如果不同门户网站同时出现了相关新闻，也能反映出当前社会的热点。对门户网站新闻进行聚类也有利于社会热点的研究，如果有大量历史数据的数据库，则在社会研究上也有较好的附加效益。

为此，本项目提供了对于几大新闻网站首页和滚动新闻的标题自动抓取，并能将结果存储至数据库中。需要时，用户可以通过 web 界面调取一段时间内的新闻标题，并获得聚类后的结果与关键词等信息。

2 使用方法

注意：程序仅提供了 Mac OS X 下的二进制程序文件

打开终端，进入应用程序 bin 目录，运行 focus 程序。

程序会加载模型文件，加载完成后会提示用户打开 web 界面，并在后台定期抓取最新的新闻条目。用户可以访问 <http://localhost:8888/> 查看 web 界面，在右上角设定开始结束日期后，点击 GO 按钮获取聚类结果。点击标题可以打开原始新闻网站。对于每一个类别，还可以点击 ViewTags 按钮查看关键词云。