

# 「基于微博的情绪可视化系统」

## ——项目概要介绍

### 1. 前言

目前，社交网络平台（SNS）的使用已经越来越广泛，其中新浪微博拥有 1.29 亿的月活跃用户，为此对微博信息的研究逐渐也成为热点。可以说微博是一个信息金库，可以从中挖掘出许多有价值的信息。

那么，微博用户们每天的心情怎么样？他们的心情分布有规律吗？和地域之间存在什么样的关系？黑色星期一之说真的存在吗？女性真的比男性更容易获得幸福感吗？针对这些问题，本项目以新浪微博为信息来源，探究社交网络平台用户的情绪状态及其背后的规律。

### 2. 创意描述

我们可以由微博得到什么？

众所周知，在互联网日益发展的今天，人们越来越习惯在社交平台上发表自己的感受，微博最直接的作用就是表达用户的心情。由此，我们想到从微博的内容得到用户最真实的心情，对其进行分析，以一种有趣且直观的形式得到“我们”真正的感受。

图 1 是本团队由微博展开，在讨论中得到的作品思维导图：

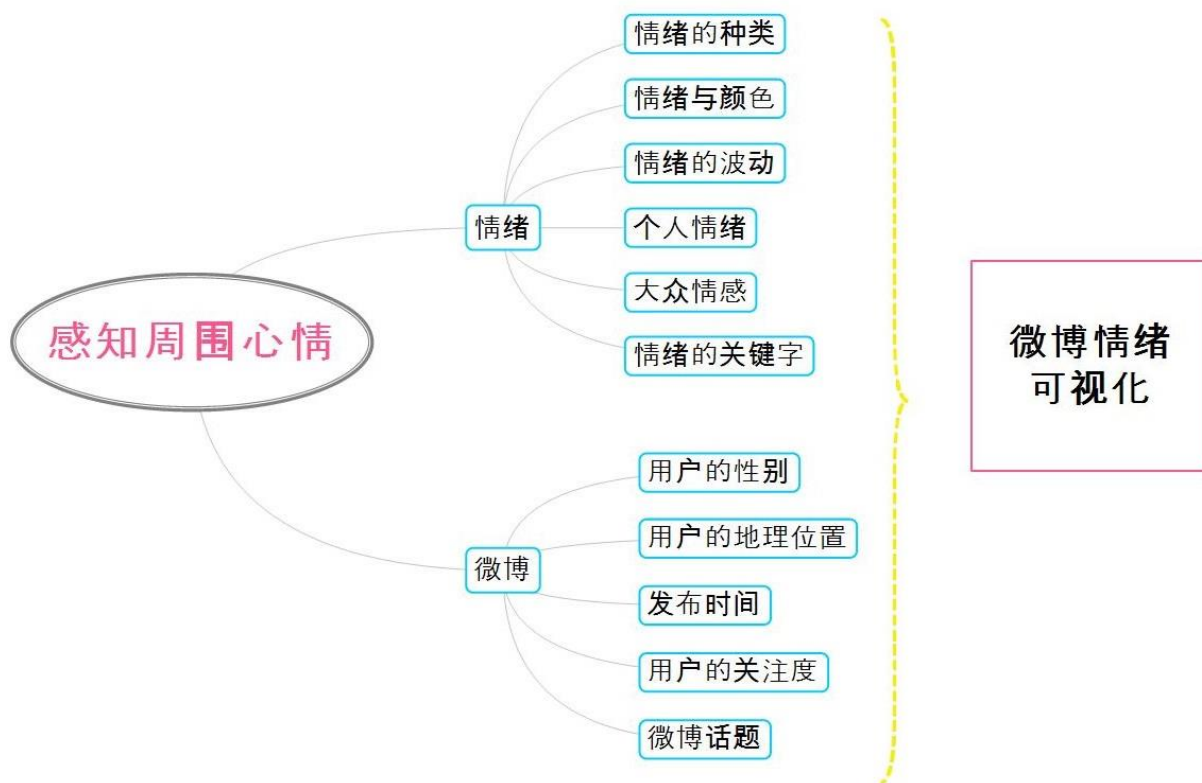


图 1：参赛作品构思思维导图

3. 功能简介

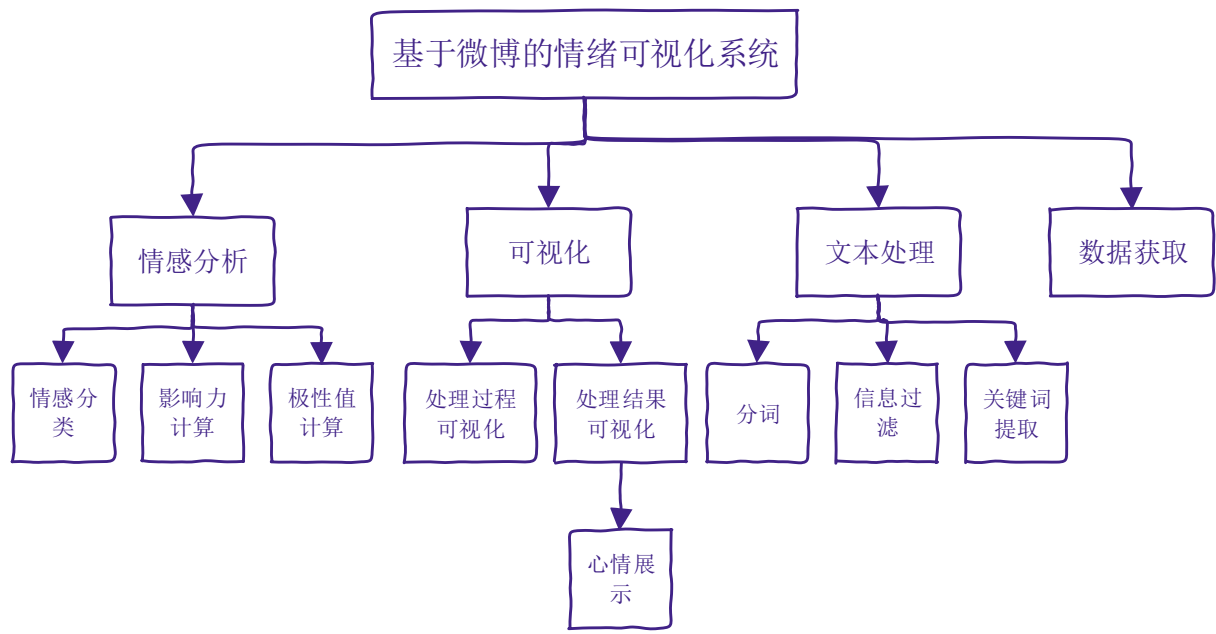


图 2：参赛作品主要功能结构

表 1 描述了本项目的功能模块划分和各模块功能设计：

功能模块	功能描述
数据获取模块	获取微博、微博话题、微博用户信息
噪声过滤模块	获取垃圾微博：广告和营销微博 提取垃圾词：对垃圾微博进行分词处理，提取关键词，统计词频 构造垃圾微博过滤器
情感分析模块	建立情感词典并标注词语的强度、极性、情感分类 建立停用词、程度副词、否定词、关联词词表 句法分析：根据依存关系构造句法树，提取情感词的主题词 建立情感模型：获得文本的情感分类、情感倾向 计算微博话题影响力：根据微博的赞数、评论数、转发数，计算该话题的影响力
文本处理模块	中文分词 提取关键词 构造信息过滤器：过滤表情符、用户名、转发符、网页链接、话题符

I/O 模块	文件读写 数据库读写
数据分析模块	按分类（性别、地理位置、时间、话题）统计不同类型心情（或不同情感极性）的用户数
可视化模块	处理结果可视化：浏览微博心情散点图，排序/打乱散点图；浏览心情时间波动；切换至心情/极性值时间波动；浏览心情性别对比；浏览心情地理分布；浏览心情话题波动；浏览热门心情关键词的微博；浏览名人微博情感走势  处理过程可视化：输入文本展现情感分析结果

表 1：参赛作品主要功能描述

## 4. 特色综述

### ♥ 设计创新：

- ✧ 将人类感性的情绪与理性的计算机数据分析结合，融入社会心理学与统计学知识，将心情以可视化的形式展现出来。

### ♥ 技术创新：

- ✧ 根据微博文本特征和网络用语特征，参考多个情感词汇本体及情感词典，自建噪声词库和情感词库。
- ✧ 设计适用于中文网络环境的情感分析模型及算法。
- ✧ 使用 HTML5，设计基于 Canvas 的图表绘制渲染模块。将图表绘制划分为数据解析，特征参数读入，数据模型填充，绘制等四部分。

### ♥ 应用前景：

- ✧ 目前项目的数据来源是微博，将来可以扩大数据，研究不同社会化媒体用户的情绪状态。
- ✧ 可以加入更多统计因素，获取心情与各种因素间的关联关系。
- ✧ 可以进行个性化定制，获取专人的心情“晴雨表”。

## 5. 开发工具与技术

开发工具	Eclipse Kepler for J2EE
技术	服务器端 采用 B/S 架构，服务器端使用 JAVA 的 SSH 开发框架，进行数据处理，使用 Tomcat 服务器，处理项目中的数据传输

路线	客户端	使用 jQuery 框架，利用 HTML5 中的 Canvas 标签特性，结合可视化的 JavaScript 库，将后台传送过来的以 JSON 为形式的数据，显示在页面上，根据统计项划分，结构清晰，易于功能扩展。
----	-----	--

## 6. 应用对象与环境

个人	大众可通过本项目获得一种有趣、直观的视觉享受，并了解周围人的心情。
学术	本项目提供了一种基于社交媒体的情感分析方法，可作为基础工具，为社会化媒体数据研究提供帮助。
商业	本项目中情感分析数据可作为商业决策的参考依据，分析方法可嵌入电影票房预测系统、商品推荐系统、用户个性化服务定制等，以提升结果的准确度。

## 7. 结语

在项目的进行中，团队成员齐心协力，遇到困难一起解决，积攒了项目经验，提升了实战能力和团队协作能力，是一次难忘的学习经历。

我们团队希望大家通过本应用更好地了解身边的人和事，在真实的世界里与大家有更多更好的交流。