

自动发帖机器人调研

戴立森

Nov 10, 2020

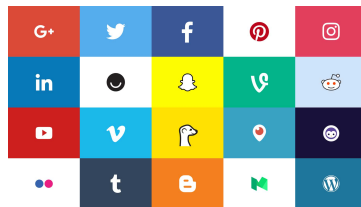
场景介绍

国内社交媒体

核心和衍生社会化媒体平台矩阵关系示例



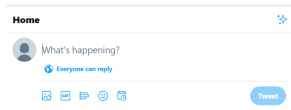
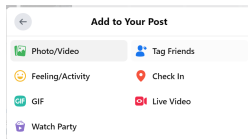
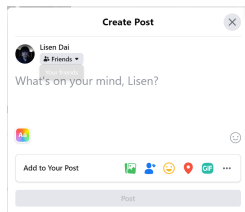
海外社交媒体



场景介绍

一个“帖子”的可能元素

- 文字
- 图片
- 短视频
- tag
- 超链接
-



任务结构

- API 接口设置

- ① 利用社交媒体给出的接口，或者直接在网页上模拟人的交互，进行发帖操作。
- ② 利用社交媒体给出的接口，或者利用爬虫等抓取方式，获取网络上的信息以供后端处理。

- 生成帖子

- ① 利用已知信息生成新的信息
- ② 自动识别并且直接借用他人素材、链接，和现有的其他信息搭配

API 接口简述

- 读取

- ① 调用开发者提供的接口读取
- ② 爬取网页内容

- 写入

- ① 调用开发者提供的接口读取
- ② 爬取网页前置接口，分析，并且记录特征，调用之

信息生成

- 生成独立信息 不依赖其他信息，仅仅依靠简单的信息（比如话题）创生出内容来。可以是一段文字，一个其他话题，一个图片，甚至是一个视频。
- 生成依赖信息 依赖其他信息，生成相关的“不抄袭”的信息，多见于模仿创作，或者文章、图片和视频“解读”式创作。
- 生成关联信息 关联其他帖子，生成相关的“不抄袭”的信息，多见于评论，或者图片视频二次创作。

推特 API(1.1)

推文

- 管理账户设置和信息
- 禁言、屏蔽、举报其他用户
- 跟随，搜索其他用户
- 生成和管理用户列表
- 用户信息图片和主页背景

用户

- 发布、接受、使用推文
- 推特时间轴
- 整理推文集
- 搜索推文：七天内的
- 按照需求过滤实时推文
- 样例实时推文

推特 API(1.1)

直接消息

- 发送和接受事件
- 开场消息
- 消息附件
- 快速回复
- 按键
- 键入提示符和读取回执
- 对话管理
- 自定义对话资料
- 顾客反馈卡

媒体、热点和地理

- 上传媒体文件
- 获取接近于某个地区的热点
- 获取按照地区整理的热点
- 获取关于某个地点的信息
- 获取接近于某个地区的地点

微博 API

微博

- 读取接口
 - 获取当前登录用户及其所关注用户的最新微博
 - 获取用户发布的微博
 - 返回一条原创微博的最新转发微博
 - 获取 @ 当前用户的最新微博
 - 根据 ID 获取单条微博信息
 - 批量获取指定微博的转发数评论数
 - 根据 ID 跳转到单条微博页
 - 获取官方表情
 - 第三方分享到微博
- 写入接口
 - ① 第三方分享到微博

微博 API

评论

- 读取接口
 - 获取某条微博的评论列表
 - 我发出的评论列表
 - 我收到的评论列表
 - 获取用户发送及收到的评论列表
 - 获取 @ 到我的评论
 - 批量获取评论内容
- 写入接口
 - ① 评论一条微博
 - ② 删除一条我的评论
 - ③ 批量删除我的评论
 - ④ 回复一条我收到的评论

微博 API

用户

- 读取接口
 - ① 获取用户信息
 - ② 通过个性域名获取用户信息

公共服务

- 读取接口
 - ① 通过地址编码获取地址名称
 - ② 获取城市列表
 - ③ 获取省份列表
 - ④ 获取国家列表
 - ⑤ 获取时区配置表

网页内容接口

source code.bb

The screenshot shows a web browser displaying the Weibo homepage. The browser's address bar shows the URL. The page content includes a header with the Weibo logo, navigation links (首页, 视频, 发现, 游戏, 注册, 登录), and a main content area with a '热门' (Popular) section and a '帐号登录' (Account Login) section. A developer tool is open at the bottom, showing the 'Sources' tab with the file 'login.php' selected. The source code of 'login.php' is displayed, showing HTML and JavaScript code. The code includes meta tags for charset and viewport, a description of the site, and various configuration variables for the login process.

```

1 <!doctype html>
2 <html>
3 <head>
4 <meta charset="utf-8">
5 <meta http-equiv="X-UA-Compatible" content="IE=edge,chrome=1">
6 <meta name="viewport" content="initial-scale=1,minimum-scale=1" />
7 <meta content="随时随地发现新鲜事！微博带你欣赏世界上每一个精彩瞬间，了解每一个幕后故事，分享你想表达的，让全世界都能听到你的心声！" name="description" />
8 <link rel="mask-icon" sizes="any" href="//img.t.sinajs.cn/t6/style/images/apple/wfbont.svg" color="black" />
9 <link rel="shortcut icon" type="image/x-icon" href="//favicon.ico" />
10 <script type="text/javascript">
11 try(document.execCommand("BackgroundImageCache", false, true));catch(e){}
12 </script>
13 <title>微博-随时随地发现新鲜事</title>
14 <link href="//img.t.sinajs.cn/t6/style/css/module/base/frame.css?version=91c79ed46b5606b9" type="text/css" rel="stylesheet" charset="utf-8" />
15 <link href="//img.t.sinajs.cn/t6/style/css/pages/growth/login_v5.css?version=91c79ed46b5606b9" type="text/css" rel="stylesheet" charset="utf-8">
16 <link href="//img.t.sinajs.cn/t6/skin/default/skin.css?version=91c79ed46b5606b9" type="text/css" rel="stylesheet" id="skin_style" />
17 <script type="text/javascript">
18 var $CONFIG = {};
19 $CONFIG['islogin'] = '0';
20 $CONFIG['version'] = '91c79ed46b5606b9';
21 $CONFIG['timeDiff'] = (new Date() - 1605532329000);
22 $CONFIG['lang'] = 'zh-cn';
23 $CONFIG['jsPath'] = '//js.t.sinajs.cn/t5/';
24 $CONFIG['cssPath'] = '//img.t.sinajs.cn/t5/';
25 $CONFIG['imgPath'] = '//img.t.sinajs.cn/t5/';
26 $CONFIG['serverTime'] = 1605532329;
27 $CONFIG['location'] = 'login';
28 $CONFIG['bigpipe'] = 'false';
29 $CONFIG['bpType'] = 'login';
30 $CONFIG['mjsPath'] = [ '//js(n).t.sinajs.cn/t5/', 1, 2];
31 $CONFIG['mcssPath'] = [ '//img(n).t.sinajs.cn/t5/', 1, 2];
32 $CONFIG['redirect'] = '';
33 $CONFIG['vid'] = '';
34 </script>
35 <style#js_style_css_module_global_WB_outframe(height:42px);</style>
36 </head>
37
38 <body class="WB_mac">
39 <div class="B_unlog">

```

信息生成

通过实际应用层面的需求，可以看出，大的方面来说，一个“自动发帖机器人”需要有如下功能：

- 建立“用户画像”，针对自己的“兴趣、职业”，进行发帖等操作；
- 建立“用户关注画像”，针对自己的“关注话题”，进行转帖等操作；

发帖信息生成

根据“用户的兴趣和特点”建立初始的“用户画像”，并且，我们可以建立一个初始数据库，由此：

- ① 文字（内容或是话题）
- ② 图片
- ③ 视频
- ④ 文字引用生成评论-NLP 特征提取、NLP 生成模型；
- ⑤ 图片配文-图片特征提取、NLP 生成模型；
- ⑥ 视频配文-图片、视频特征提取，NLP 生成模型；
- ⑦ 经由图片的图片生成-GAN 图像翻译；
- ⑧ 经由文字的文字生成-NLP 生成模型；
- ⑨ 为自己的关注用户、关注话题和关注内容生成分类标签
- ⑩ @ 符合内容特征的关注用户

用户画像完善

在初始“用户画像”上，根据“关注点”和“(地区)热点”而收到信息：

NLP 特征提取、NLP 生成模型、随机森林等等

接收到：

- ① 文字
- ② 图片
- ③ 视频
- ④ 整体帖子

- ① 根据文字、图片、视频的特征，是否加入数据库
- ② 根据整体帖子的所有特征（内容、关注人群、热度、地区）综合，是否转帖
- ③ 根据整体帖子的所有特征（内容、关注人群、热度、地区）综合，是否进入关注组
- ④ 根据整体帖子的反馈，扩大“关注点”
- ⑤ 根据其他用户和自己的“关注点”修改自己的用户画像
- ⑥ ...

数据池说明

- 内容库（文字（话题）、图片等）
- 关注点库
- 用户画像库（自己和其他）
- ...