

作者：ty10086

Drupal 8/9+火车采集器的原理与技术优势：

N 个渠道媒介（WWW、RSS、Feeds、XML 等任何可见内容）来源的原始内容

↓

经过特定规则清洗处理（火车采集器 V9.X+各种自带预设处理插件（分类/分词/翻译/伪原创/SEO/文件操作等一切数据的处理）+python/php/java 高度个性化内容处理）的标准化、规模化、二次处理内容

↓

发布到展示媒介工具（Drupal 8.8.8 以后的版本），以支持媒介工具对编排的内容的高度自由加工

↓

大规模自动化发布、修改、删除、更新

拟解决问题：

模拟手动发布任意已采集内容自动发布到 Drupal 8/9 的长期无接口问题、自动登录 Drupal8/9，以解决 Drupal 安全机制中的多个随机值问题导致无法发布。

Drupal 一般通用的预设基本情况：

本案例设置

预原装 Drupal 8.8.8 以后版本

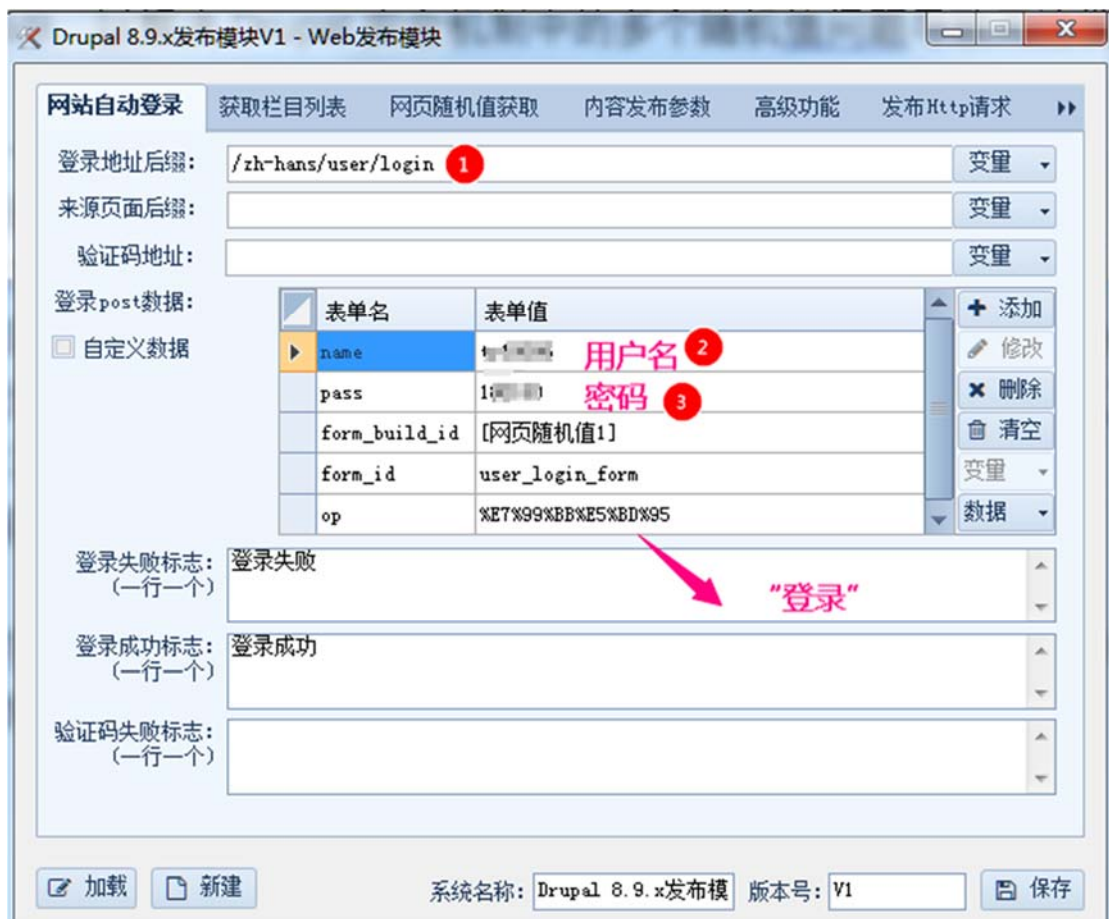
预设自定义 contentype：增加需要发布的内容类型，或者 entity clone

预配置为您的内容类型设置 Pathauto 模式：自动生产 URL（本工具以火车采集器的自动分词的 3 个标签生产 URL），非必须，可留空

预设当前用户的文本编辑器:Ckeditor 可使用 Full HTML 为了适配任意来源文章的奇怪 html 标签,此处必须

额外的字段设置 根据需求添加,将修改后的字段表单名/表单值录入到 Web 发布模块的模块内容发布参数中即可。

火车采集器 V9.X 一般通用的预设基本情况：



Drupal 8.9.x发布模块V1 - Web发布模块

网站自动登录 获取栏目列表 **网页随机值获取** 内容发布参数 高级功能 发布Http请求

标签名	获取地址	来源地址	前字符串	后字符串	是否每次获取
1 [网页随...]	/zh-hans/user...		name="form_b..."	" />	False
2 [网页随...]	/zh-hans/node...		name="form_b..."	" />	False
3 [网页随...]	/zh-hans/node...		name="form_t..."	" />	False

1为自动登录需要的随机值
2、3为发布内容需要的随机值

相关操作: + ✎ ✕ 常用标签: [标签:XXX] [分类ID] [全局变量] [分类名称] (*)

强制编码: UTF-8 使用强制编码后, 整个模块中的发布都使用该种编码

Drupal 8.9.x发布模块V1 - Web发布模块

网站自动登录 获取栏目列表 网页随机值获取 **内容发布参数** 高级功能 发布Http请求

发表地址后缀: /zh-hans/node/add/article_for_alerts 标签

来源页面后缀: 标签

发布post数据:

自定义数据

表单名	表单值
▶ title[0][va...]	[标签:标题]
form_build_id	[网页随机值2]
form_token	[网页随机值3]
form_id	node_article_for_alerts_form

发表错误标志:
(一行一个)

发表成功标志:
(一行一个) 已创建

加载 新建 系统名称: Drupal 8.9.x发布模 版本号: V1 保存



内容发布参数的完整表单名-表单值可用数据-自动抓取登录数据包自由修改/删除/增加, 表单值以采集到的字段标[标签:新字段]签替换即可



编辑任务：[任务名称] 全文提取

1. 网址采集规则 2. 内容采集规则 **3. 内容发布规则** 4. 其他设置

Web在线发布

Web发布设置列表

Web发布设置列表

Web发布设置

1、选择发布模块

- DiscuzX_3.x_门户
- Drupal 8.9.x发布模块V1**
- Drupal 8.X内容发布模块1.0
- DvbbsVersion 8.3.0 文章

2、网页编码: UTF-8 3、全局变量:

4、登录操作

网站地址: **1**

登录方式: 内置浏览器登录 数据包登录 不登录

UserAgent: Mozilla/5.0 (Windows NT 6.1; Win64; x64; Trident/...

Cookie: HtZhbEz0q4b49BKXL5DqCABiD3efY_7upLhA...zjw

5、获取分类/栏目列表 (分类ID及名称可手工填写)

获取列表:

分类ID: 分类名称:

测试当前发布

配置名: 保存

Drupal 8.9.x发布模块V1 - Web发布模块

网站自动登录

获取栏目列表 网页随机值获取 内容发布参数

登录地址后缀: **2**

来源页面后缀:

验证码地址:

登录post数据:

自定义数据

表单名	表单值
name	
pass	
form_build_id	[网页随机值1]
form_id	user_login_form
op	XE7%99%8B%25%BD%95

登录失败标志: 登录失败 (一行一个)

登录成功标志: 登录成功 (一行一个)

验证码失败标志: (一行一个)

重要：1处为根域名

2、3、5处为相对路径

1+2组成完整的登录域名

1+5组成完整的发布内容页域名

1+3同理 系统名称: Drupal 8.9.x发布模块 否则错误

加载 新建