



项目编号：
2016070

CALIS 全国农学文献信息中心研究项目 结题报告

项目名称： 基于 Chrome 浏览器扩展的图书采访插件开发

项目关键词： 浏览器扩展 图书采访

项目单位（盖章）： 仲恺农业工程学院图书馆

通信地址： 广州市海珠区东沙街 24 号

项目主持人： 杨雁

联系电话： 13570394453

电子邮件： 280453778@qq.com

提交日期： 2017 年 4 月 24 日

项目结题验收单

1 专家验收表（主持人所在单位组织 3-5 名专家对项目进行验收、自评。）

项目名称	基于 Chrome 浏览器扩展的图书采访插件开发				
主持人	杨雁	职务/职称	馆员		
所在单位	仲恺农业工程学院图书馆				
专 家 意 见	<p>课题研究围绕 Chrome 浏览器的图书采访扩展进行研究与开发，前期准备充分，开过程度详情、成果具有多项功能，实现了从浏览器的数据获取，后台的信息整合与处理，及与汇文系统采访模块对接的一整套完整流程。项目成果极大的减少采访人员在数据获取、查重、及系统订单处理的时间，拥有较强的实践价值，可以为实际工作提供良好的支撑。同时项目成果具有一定的推广价值，能够为其它使用汇文系统的采访人员所用。</p> <p>经专家验收组研究认为该项目研究成果完成 CALIS 全国农学文献中心对该项目的任务与要求，实现了预期目标，同意通过评审，准予结题。</p>				
专家签字					
职务/职称					

1 研究背景、目的及意义

1.1 研究的背景

时代在前进，社会在发展，科技在进步，特别是当前正值知识爆炸的信息时代，图书馆更好的生存与发展和先进技术的运用是密不可分的。信息技术极大地推动了图书馆的现代化进程，同时也带来了信息的爆炸式增长。在当今知识经济信息时代的新形势下，不断加强和改进藏书建设是强化图书馆功能的重要前提，是否能从海量信息中选取有用、实用、新颖的知识信息，是关系到图书馆未来生存与发展的重大问题。

图书采访工作是图书馆文献资源建设的基础性工作。由于书价上涨、经费短缺、文献载体多样化、馆藏结构的改变、图书市场多元化发展等原因，图书馆图书采访工作面临着新的挑战。为保证馆藏结构的合理，各图书馆都在探索适合自己的采访模式，以适应当代图书馆的发展。随着计算机技术和网络技术的应用，在网络环境下，图书馆采用何种有效的图书采访模式来满足读者需求，是图书馆文献采访人员值得探讨的问题。

目前大部分图书馆都是通过招标确定图书供应商，以现场采访和书目采访为主要的采访方式，现场采访实物选书，直观、准确、保证质量，避免了圈选书目的盲目性和不确定性，但现书选购的品种依赖于书商的进货，无论书商的实力有多强，都不可能给图书馆提供所谓的“全品种”难免有缺口。而且，每家书商有它自己的经营倾向和对出版社的偏爱，选择什么样的图书进货不是图书馆能控制的。通过书目选书，选书范围广，易于满足馆藏需求。特别是对多卷书、丛书、上下册等不同时间出版发行的图书，较易收集齐全。但由于书商为了营利往往会把书目数据经过一定程度的过滤后再发给图书馆采访人员，而有价值的图书由于学科方向局限、出版数量少，书商或是拿不到书，或因为折扣率问题没有利润，往往被剔除，进而导致有价值的图书不能进入图书馆。同时由于书目选书时不能直接看到图书，对图书的内容、装帧、开本、页码、读者对象等信息只能通过阅读书目信息来了解。如果对这些信息阅读不仔细，就会将小开本、小册子及不适合本馆读者需求的图书选进馆，极大的影响馆藏图书的质量。

1.2 研究的目的及意义

本研究所设计的是一种书目采访的改进方式即书目数据来源直接从各大网上书店、出版社、读书分享站点等获取，目前主流的图书类电商如当当、京东、亚马逊、豆瓣等，在书目信息的提供上非常详细，除了纸质书目所有特征，如题名、责任者、版年、开本、页码、简介等，还在网上将图书的封面、目录挂上，有的甚至能够提供样章的试读，这些数据不仅提供了最新、最广泛的书目来源，而且通过其网站上的图书销量、客户评论等数据还能为采访人员提供更为理想的选择和评价依据。那么通过 Chrome 图书采访插件对网上书店书目信息高效自动化的信息抽取技术，可以弥补传统手工获取的不足，节省大量人力和时间，以最便捷的方式组织成图书馆的采访订单，提高采访的效率就是本课题研究的重要现实意义。

2 研究内容及方法（思路、方法、具体内容）

2.1 Chrome 浏览器及扩展

Chrome 是一个由 Google（谷歌）公司开发的网页浏览器。该浏览器是基于其他开源软件所撰写，包括 WebKit，目标是提升稳定性、速度和安全性，并创造出简单且有效率的使用者界面。截止 2017 年 3 月 chrome 浏览器全球市场份额达到 58.64%，而且在国内 360 浏览器，Sogou 浏览器等众多浏览器都是基于 Chrome 的内核进行开发的，原则上讲只要是 Chrome 扩展都能在这些浏览器上使用。这也大大的提高开发 Chrome 扩展的可用性。

Chrome 扩展是用于扩充 Chrome 浏览器功能的程序，主要用于对浏览器功能的增强，它更强调与浏览器相结合。通过扩展可以对用户当前浏览的页面进行操作，实际上就是对用户当前浏览页面的 DOM 进行操作，当用户访问这些页面后，相应脚本即可自动运行，从而实现对页面 DOM 进行任意增、删、改等复杂操作。而它的界面则使用 HTML 和 CSS 进行描述，这样的好处是可以用很短的时间构建出赏心悦目的 UI。同时由于编写扩展的语言 JavaScript 是一门解释语言，无需对其配置编译器，调试代码时只要刷新一下浏览器就可以看到修改后的结果，这使得开发周期大大缩短。

2.2 扩展的前端开发

2.2.1 基本知识及准备

一个 Chrome 扩展是由 HTML、CSS、JavaScript、图片等文件压缩而成。每个扩展由下列文件组成：

- 一个 manifest 文件(主文件, json 格式, 必备)
- 至少一个 HTML 文件(主题可以没有 HTML 文件)
- JavaScript 文件 (可选, 非必须)
- 任何其他需要的文件(比如图片, CSS 文件等)

根据实现需要本项目研究扩展, 具体的文档结构如图1所示:

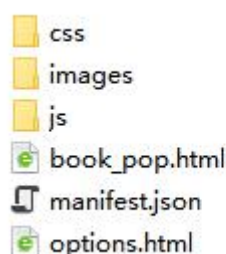


图1 采访扩展文档结构图

css 目录存放 css 文件, 用于控制设置页面, 注入内容的界面样式; images 目录存放扩展用到的所有图片文件, 包括扩展按钮图标等; js 目录存放扩展用到的所有 js 文件, 包括有背景页、设置页、注入内容页、通用类等; Book_pop.html 是 Brower action 的弹出页面, 用于展示后端链接入口和扩展的设置入口; Options.html 是扩展设置页面, 用于设置后端处理的 URL 地址; Mainfest.json 是图书采购扩展的主文件, 也是 Chrome 扩展唯一要求必须存在的文件, 文件的内容是以 JSON 格式存储, 用来描述扩展的扩展名字、版本、调用文件、可用域等内容。文件源代码如下:

```
{
  "manifest_version": 2,
  "name": "图书采购之星",
  "version": "1.0.0",
  "description": "该插件能够实现当当, 京东, 豆瓣, 亚马逊等图书类网站的图书信息的直接数据采集, 并能够与汇文系统实现采集数据的对接, 生成采购订单。",
  "permissions": [ "http://**/*", "https://**/*"],
  "web_accessible_resources": [ "images/add_cart1.png", "images/add_cart2.png", "images/recom.png", "js/jquery.min.js", "js/book_action.js"],
```

```

"browser_action": {"default_icon": "images/book-48.png", "default_title": "图书采购之星", "default_popup": "book_pop.html" },
"background": {
  "scripts": ["js/jquery.min.js", "js/book_background.js"],
  "persistent": true
},
"options_page": "options.html",
"icons": {"48": "images/book-48.png", "128": "images/book-128.png"},
"content_scripts": [
  {
    "matches": ["http://*/*", "https://*/*"],
    "css": ["css/icon.css", "css/content.css"],
    "js": ["js/jquery.min.js", "js/book_action.js", "js/book_contentscript.js"]
  }
]
}

```

以上文件 `name` 定义了扩展的名称，`version` 定义了扩展的版本，`description` 定义了扩展的描述，`icons` 定义了扩展相关图标文件的位置，`web_accessible_resources` 定义了扩展在注入的目标页面上所需使用的资源的路径（相对于扩展的安装根目录），`permissions` 属性定义了需要跨域的权限。`browser_action` 指定扩展的图标放在 Chrome 的工具栏中，`browser_action` 中的 `default_icon` 属性定义了相应图标文件的位置，`default_title` 定义了当用户鼠标悬停于扩展图标上所显示的文字，`default_popup` 则定义了当用户单击扩展图标时所显示页面的文件位置。`content_scripts` 指定脚本注入域及注入页面用到的 `css` 和 `js` 文件，由于拥有图书信息的网站很多，划词采购在所有页面都需要用到，随着各网站对信息安全的重视，`https` 的站点也越来越多，并将会成为未来的主流，因此本扩展选用匹配所有 `http/https` 站点，当用户访问这些页面后，相应的采购注入脚本即可自动运行，从而对页面 `DOM` 进行操作，并提取图书信息。

2.2.2 前端核心功能开发

通过研究项目确定了以当当网、京东、豆瓣、亚马逊、微博读书等主流综合性图书网站及图灵社区、Chin-pub 等专业技术类图书网站为对象进行研究，通过 Chrome 扩展精确实现对网页内题名、ISBN、价格、图片等重要图书信息的抽取；并且通过正则表达式获取所有含 13 位 ISBN 号网页的图书信息抽取；实现所有网站划词选定图书题名的图书信息抽取。这些核心功能是通过 `content_scripts`（内容脚本）来实现的，`content_scripts` 在 Chrome 扩展编写中扮演着非常重要的角色，其

内容就是 JavaScript 代码，虽然其物理位置是在扩展目录中，但其内容代码是根据匹配规则被注入到对应的访问网页中。在扩展的开发中，笔者将注入的类型分为以下几种：

2.2.2.1 所有网站中宽度和高度都大于90像素的图片

通过对上述网站的调查分析，笔者发现绝大部分图书的封面图片的宽度和高度都会大于 90 像素，对于这一通用特征可以大大的提高注入代码的可用性。具体的实现上是，在网页加载完成后，通过 `jQuery("img")` 来获取所有的图片，然后再对图片的宽度和高度做相应的判断，如果符合要求，则为该图片添加一个鼠标经过事件，也就是当鼠标放在图片上时会弹出一个“采购”的按钮，点击“采购”按钮后将图片的链接发送给后端程序进行处理，提取相应的图书信息数据，具体效果如图 2 所示。其关键代码如下：

```
//给图片加上采购按钮
jQuery("img").each(function() {
    if (this.offsetWidth > 90 && this.offsetHeight > 90) {
        jQuery(this).mouseover(function() {
            thisDom = jQuery(this).parents('a').attr('href');
            .....
            showFloatBtn(this, "#mySideBarFloatBtn");
        });
        jQuery(this).mouseout(function() {
            hideFloatBtn("#mySideBarFloatBtn");
        });
    }
});
//处理采购点击事件
jQuery("#mySideBarFloatBtn").click(function() {
    ... ..
    var item=thisDom;
    jQuery.ajax({
        type: "get",
        url: localStorage.SysRoot + "getvalue.php?act=list&t=tips&url=" + item + "&f=" + from,
        dataType: "text",
        success: function(data) {
            //处理成功后的操作
        }
    });
    ... ..
}
```



图2 图片弹出方式注入页面效果

2.2.2.2 网站的列表页

当当网、京东、豆瓣、亚马逊、微博读书、图灵社区、Chin-pub 等网站的图书列表页，通过正则匹配的方式将“加入采购”，“继续采购”两个按钮放置在不影响网站正常使用，而且又能使采访人员能够便利的进行图书的采购操作，“断续采购”按钮用于对已有馆藏数据的再次采购操作，通过“继续采购”采购的数据在后端数据库里将进行标识，以便与其它查重后数据进行区分。



图3 列表页注入页面效果

2.2.2.3 划词获取图书信息

该功能主要是方便采购人员在浏览非图书类网站时，如在看新闻、微信、微博等有关于图书的推荐内容时，能够通过划词选中书名的方式获取到图书作者、ISBN、内容简介、价格、馆藏信息等内容，并且可以通过“加入采购”，“继续采购”两个按钮直接将图书加入到后台的采购数据库中，可以极大的减少采购人员查找和组织生成采购信息的时间和精力。如图4所示，在浏览新闻时看到《人民的名义》一书，采访人员可以通过划词的方式获取到该书的详细情况及馆藏和借阅信息，可以通过采购按钮进行数据的采集。



图4 划词方式注入页面效果

2.2.3 前端其它内容开发

2.2.3.1 扩展设置页面

为了使该扩展更加具有可用性及其可扩展性，本扩展通过增加设置页面来填写后端处理程序的链接地址，这样可以让同样使用汇文系统的其它兄弟图书馆，在不修改扩展代码的情况下就可以方便的部署。由于设置页面中的链接是利用浏览自带的基于域名的 localStorage 来存储，而 content_scripts 是注入到采访人员浏览页面中

的，如果 `content_scripts` 直接读取 `localStorage` 存储的设置链接，将会获取空值，本扩展的解决办是 `content_scripts` 通过 `runtime.sendMessage` 和 `background` 通信，由 `background` 读写扩展所在域的 `localStorage` 存储的设置链接，然后再传递给 `content_scripts` 实际数据的交互。具体的架构图如图 5 所示：

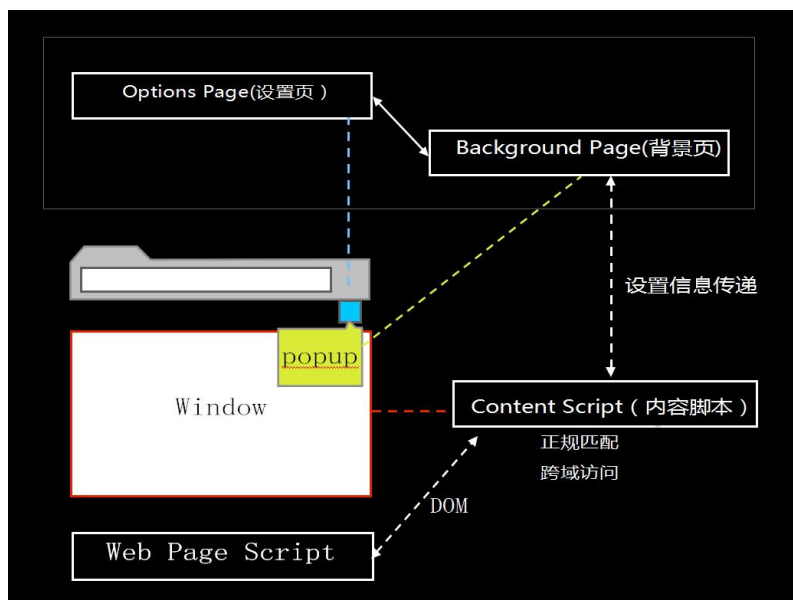


图 5 扩展框架图

2.2.3.2 弹出窗口页

开发以本页面的作用主要是用于沟通前端扩展与后端数据处理页面的联系，可以减少采购人员操作步骤，点击未处理或已处理订购，可直接进入后端页面。同时提供扩展设置的便捷链接。应用效果如图 6 所示：



图 6 扩展安装及弹出页效果

2.3 后端处理平台的开发

图书采访扩展的前端主要是提供获取数据的入口，方便采购人员进行图书信息数据的选择，而当采购人员进行点击操作后的处理就必须交由后端程序进行数据查重、抽取、采购清单的生成及导入汇文系统等操作。

在技术实现上该项目将采用最流行的编程语言之一的 PHP 语言来实现，同时以开源数据库软件 MYSQL 来作为项目的后台数据库。PHP 起源于自由软件，即开放源代码软件，使用 PHP 进行 Web 应用程序的开发具有安全性高；跨平台性好，PHP 几乎支持所有的操作系统平台，并且支持 Apache、IIS、Nginx 等多种 Web 服务器；支持广泛的数据库，可操纵多种主流与非主流的数据库，如 MySQL、Oracle 等，其中 PHP 与 MySQL 是现在最佳的组合，它们的组合可以跨平台运行。同时项目采用了 THINKPHP 开源 PHP 框架，框架的应用可以大大减少底层功能的代码编写，有助于创建更为稳定的应用。ThinkPHP 是一个快速、兼容而且简单的轻量级国产 PHP 开发框架。ThinkPHP 设计简洁实用，拥有出色的性能和至简的代码，易用性高，并且拥有众多的原创功能和特性，在易用性、扩展性和性能方面都有不错的表现，已经成长为国内最领先和最具影响力的 WEB 应用开发框架，众多的典型案例确保可以稳定用于商业以及门户级的开发。

2.3.1 后端运行环境的搭建

2.3.1.1 配置 ORACLE 扩展

“汇文文献信息服务系统”（以下简称汇文系统）是由江苏汇文软件有限公司研发。从 1999 年在全国范围内推广至今，有近 700 余家知名高校馆和公共馆使用，其中包括“985”学校中的 15 所，“211”学校中的 53 所大学。汇文对系统目前已经是图书馆界占据份额最大图书管理系统，到了全国广大用户的一致肯定和推崇，文软件在很短的时间内迅速成长为全国图书馆界公认的著名品牌，文系统显示了良好的适应性和市场竞争力。汇文系统采用 Client/Server、Browse/Server 双重体系结构，运用中间件技术，创建开放的、创新的、可扩展的、基于图书馆文献资源共享和文献服务共享的分布式应用软件系统。采用大型关系型数据库 Oracle 作为数据库服务平台。

为了使本项目的数据能够和汇文系统的数据实现直接的交互，必须安装 ORACLE

E 客户端。只有先完成 ORACLE 客户端才能使后端程序运行环境里的 PHP 通过扩展与 Oracle 数据库建立联系。ORACLE 客户端可以直接在官方网站下载, 安装完成后需要配置 tnsnames.ora, 也可以直接把安装有汇文系统的对应目录的 tnsnames.ora 拷贝到对应目录完成配置。

2.3.1.2 程序运行环境的安装

本项目采用是 Windows 的系统平台, 需要进行 Nginx Web 服务器, php 及 MySQL 数据库的安装和配置来进行后端程序的开发和运行。除了可以分别进行这些服务程序包的安装和配置之外, 现在网上也有很多成熟的集成套件可以完成这些程序的打包整体安装, 可以大大节省分别安装和配置需要花费的时间。因此本项目采用了 phpStudy 程序包, 该程序包集成最新的 Apache+Nginx+LightTPD+PHP+MySQL+phpMyAdmin+Zend Optimizer+Zend Loader, 一次性安装, 无须配置即可使用, 是非常方便、好用的 PHP 运行环境, 同时有专门的控制面板, 而且 phpStudy 全面适合多个版本的 Windows 操作系统 集成了 Apache、IIS、Nginx 和 LightTPD 等多种 Web 服务器。

完成 phpstudy 的安装后需要编辑 php.ini 文件, 打开 php_oci8_11g.dll 扩展, 用于连接 Oracle 数据库; 打开 php_openssl.dll, 用于让 PHP 可以获取 https 的内容。

2.3.1.3 HTTPS 自签名证书的生成和配置

由于浏览器默认阻止混合内容 (mixed content) 的加载及请求, HTTPS 页面里动态的引入 HTTP 资源, 比如引入一个 js 文件, 会被直接禁止. 而且在 HTTPS 页面里通过 AJAX 的方式请求 HTTP 资源, 也会被直接禁止的。因此如果采购扩展的注入网站的 Https 页面的话, 就必须让后端处理程序也要运行在有 Https 服务的服务器, 才能完成请示, 否则扩展的功能将失效。因此让 Web 服务加入 Https 的功能就变得相当必要了。采用 https 的服务器必须从 CA (Certificate Authority) 申请一个用于证明服务器用途类型的证书。该证书只有用于对应的服务器的时候, 客户端才信任此主机, 而 CA 证书一般免费证书很少, 需要交费。由于本项目开启 https 的目的主要是为了防止浏览器禁止掉后端处理程序的访问, 而且非真的需要加密高安全的应用, 所以采用自签名证书的方式开启 https 的服务。具体的实现过程如

下:

(1) 生成证书

a) 使用 openssl 生成密钥 privkey.pem: `openssl genrsa -out privkey.pem 1024`

b) 使用密钥生成证书 server.pem: `openssl req -new -x509 -key privkey.pem -out server.pem -days 365` 证书信息可以随便填或者留空, 只有 Common Name 要根据你的域名填写当然填写 IP 地址也是可行的。以本项目为例 Common Name (e.g. server FQDN or YOUR name) []: 192.168.89.10

(2) 配置 nginx

需要打开 phpstudy 安装目录下的 nginx/conf/目录, 编辑 nginx.conf 文件将在里面增加如果代码:

```
server {  
    listen 443;  
    server_name youdomain.com;  
    ssl on;  
    ssl_certificate /path/to/server.pem;  
    ssl_certificate_key /path/to/privkey.pem;  
    ...  
}
```

(3) 重启 nginx 服务就即可完成配置并生效。

2.3.2 后端核心功能数据库创建及汇文系统数据表分析

2.3.2.1 扩展自建数据表:

为了能够存储前采购扩展获取的数据和程序获取的数据, 需要用到 Mysql 数据库, 项目创建 recom_book 数据表, 其表结构如表 1 所示:

表 1 Mysql 中间库数据表

列名	数据类型	含义
id	int(12)	行号
isbn	varchar(20)	ISBN
title	varchar(100)	题名
f_title	varchar(100)	标准题名
press	varchar(50)	出版社

author	varchar(80)	作者
img	varchar(100)	封面图版
price	decimal(10,2)	价格
add_time	datetime	采购时间
url	varchar(100)	数据来源链接
cat_num	varchar(30)	分类号
order_num	int(5)	采购数量
title_repeat	int(1)	是否题名重复
isbn_repeat	int(1)	是否 ISBN 重复
title_booklist	varchar(255)	重名题名详情
is_force	int(1)	是否强制采购
is_export	tinyint(1)	是否导出
export_time	datetime	导出时间
ip	varchar(15)	IP

2.3.2.2 需要使用到的汇文系统表

为了能够实现图书数据的查重，馆藏数量的获取，订购数据的导入及订购目录的生成，扩展后端需要使用到的汇文系统数据表主要有以下几个：

表 2 扩展使用的汇文系统数据表列表

表名	表说明
item	馆藏图书表
order_record	订购记录表
marc_idx	MARC 索引表
marc_data	MARC 文件表
marc	MARC 数据表
get_mode_code	图书获得方式
fund_src_code	经费列表
doc_type_code	文献类型表
deposit_account_lst	收款列表
book_rank	图书借阅评价表
location_lst	馆藏地点列表

publish_busi	供应商例表
publish_sort_wkr	工作人员与供应商例表
marc_transform	Excel 数据与 Marc 对应关系表
lib_worker	图书馆工作人员表

2.3.3 后端核心功能的实现

2.3.3.1 前端图书数据的处理和图书信息的获取

按前端扩展获取图书信息的来源分别为列表页，详情页，及划词弹出页三种方式，只有在详情页内获取的图书数据最为全面包括题名、图片、价格、ISBN 等，其它页面，采购扩展只获取详情页链接。这些数据提交到端也需要分这两种情况进行处理。对于有详情信息的直接插入到 MYSQL 的采购中间库，而对于只提供详情页链接的情况，则让程序再次通过详情页来获取题名、图片、价格、ISBN 等信息，然后再插入到 MYSQL 的采购中间库。当然为了保证数据获取完整性，对于只能获取到 ISBN 项的情况下，本项目还通过程序主动去搜索 CIP 数据库、图家图书馆及标准书目网等网站数据的方式完善图书书目的其它数据项。以便于导入汇文系统后的数据处理及采购人员更好的进行二次数据的筛选。

2.3.3.2 图书数据的查重

数据查重是采访工作必不可少的步骤，为了减少手动查重的繁琐过程，在数据进入 MYSQL 采访中间库的时就进行查重处理，对已经存在订购记录或者已经加入采购的图书将会被过滤掉。而对于采访人员非常确认需要再次采购的已重复图书也可以通过“继续采购”的方式让数据跳过查重的步骤，当然对于这种操作在存入中间库时会做标出标识，以便采访人员进行二次判断。

数据查重的实现过程是通过 ISBN 为关键字，首先查询中间库 (recom_book) 表，如果数据已经存在就取消数据插入；然后再查询汇文系统 (item) 表确认是否馆藏重复，查询 (order_record) 表确认是否订购记录重复。通过题名为关键字查询题名是否重复，对于题名重复的数据仍然会将数据插入中间库，但会在 title_booklist 字段记录将题名重复的出版年、馆藏数量、借阅量等信息。实际功能的关键代码如下：

```

... ..
$sql="select id from recom_book where isbn='".str_replace('-', '', $isbn[1])."'";
$id=$db->getone($sql);
if($id>0&&$is_force==0){//已经推荐
    continue;
}else if($is_force==1){//强制采购
    $book->query("update recom_book set is_force=1 where id='$id'");
}
$marc_rec_no=$odb->getOne("select count(*) as num from item where MARC_REC_NO=(select max(MARC_REC_NO) from marc where M_isbn like '%$isbn%' or M_isbn='".$book_info['isbn']."'");
$is_order=$odb->getOne("select order_set_qty from order_record where MARC_REC_NO=(select max(MARC_REC_NO) from marc where M_isbn like '%$isbn%' or M_isbn='".$book_info['isbn']."'");
if($marc_rec_no or $is_order){//ISBN 重复
    $book->query("update recom_book set isbn_repeat=1 where id='$id'");
}
if($book_info['f_title']{
    $f_title=$book_info['f_title'];
}else{
    $f_title=trim($book_info['title']);
}
$sql="select sum(i.TOTAL_CIRC_TIMES)as num, count(i.BAR_CODE)as book_num,m.M_isbn,m.pub_year,i.MARC_REC_NO from item i left join MARC m on i.MARC_REC_NO=m.MARC_REC_NO where i.MARC_REC_NO in(select MARC_REC_NO from marc_idx where marc_idx_cont='".iconv("utf-8", "GBK", $f_title)."' and MARC_DATA_CODE='200a') group by m.M_isbn,m.pub_year,i.MARC_REC_NO order by m.pub_year desc,i.marc_rec_no desc";
$title_list=$odb->getAll($sql);
if(count($title_list)>0){
    $tempstr="";
    foreach($title_list as $item){
        $tempstr=$item['PUB_YEAR'].", 复本数:".$item['BOOK_NUM'].", 借阅量: ".$item['NUM']."\n";
    }
    $book->query("update recom_book set title_repeat=1 ,title_booklist='".$tempstr.'" where id='$id'");
}
... ..

```

2.3.3.3 馆藏情况的获取

为了能够让采访人员在浏览图书网页时，看到馆藏的情况采购扩展通过 AJAX 请求的方式将 ISBN 号传递给后端处理程序获取馆藏情况，查询完成后在前端展示。馆藏情况获取主要是通过查询汇文系统的 item 表和 marc 表，通过 Marc 查询到 Ma

rc_rec_no, 再通过 Marc_rec_no 在 Item 表中计算出已有馆藏的数量, 通统计出 book_lend_flag 为 1 的数量查询出已经借出数量。具体实现效果如图 7 所示, 实现功能的关键代码如下:

```
... ..
$isbn=format_isbn($_GET['isbn13']);
$sql="select count(*) as num,MARC_REC_NO from item where MARC_REC_NO=(select MARC_REC_NO from marc where M_isbn like'%$isbn%') group by MARC_REC_NO";
$result=$odb->getRow($sql);
$res['have']=intval($result["NUM"]);
$link=$result['MARC_REC_NO'];
$sql="select count(*) as lend_nums from item where MARC_REC_NO=(select MARC_REC_NO from marc where M_isbn like'%$isbn%') and book_lend_flag=1 ";
$result=$odb->getRow($sql);
$res['lend_nums']=$result["lend_nums"];
echo '<a href="http://202.192.89.4/opac/item.php?marc_no='. $link.'" target="_blank">馆藏册数: ' . ($res['have']).', 借出册数: ' . $res['lend_nums'] . '</a>';
... ..
```



图7 馆藏数量显示

2.3.3.4 采购数据的汇总与处理

该功能主要实现对采购中间库数据的统一展示, 实现数据的排序展示、查重

状态的筛选，导出到 EXCEL 表，数据的删除以及对采购数量的确认。主要是在采购数据导入到汇文系统之后让采购人员能够有一个整体的控制，可以对误操作数据进行删除，可以自定义采购数量，可以导出采购清单发送给书商，可以再次进行整体数量的查重等。具体实现效果如图 8 所示，实现功能的关键代码如下：

```

... ..
$Books=$Book->field('id,title,is_force,isbn_repeat,title_repeat,export_time,title_booklist,f_title,is_ex
port,url,isbn,img ,price,cat_num')->where($where)->order($order)->select();
foreach($Books as $key=>$val){
    $Books[$key]['flag']=repeat_flag($val['isbn'],$val['title']);
    $Books[$key]['order_num']=cal_num($val['cat_num'],$val['price']);
    if($val['f_title']){
        $Books[$key]['f_title']=$val['f_title'];
    }else{
        $Books[$key]['f_title']=trim($val['title']);
    }
}
}
... ..

```

书名	查重结果	强制采购	分类号	价格	订数	已导出
麦肯锡大数据指南	查重结果: 过滤自 ISBN重复 题名重复 任意重复 不重数据		F279.712-62	45.00	2	●
高性能iOS应用开发			TN929.53	89.00	2	●
首饰艺术设计(第2版)			TS934.3	58.00	2	●
食品病原微生物学	2015, 复本数:2, 借阅量: 0; 2007, 复本数:6, 借阅量: 6; 题名重		TS201.3	59.80	2	●
阿里生态圈:马云在互联网的修炼						

图 8 采访数据的汇总与处理

2.3.3.5 与汇文系统的采购对接

该功能主要是实现在采访人员选定需要加入采购的数据后，将选定的数据导入

到汇文系统中，直接生成采购的清单，并且清单的状态是已发订状态。该功能需要采访人员进行登录的验证，查询是否具有处理采访数据的权限。主要是通过查询汇文系统数据的 publish_busi、publish_sort_wkr、lib_worker 这三张表来获取供应商列表，如果列表为空则判定期不具体相应权限。采购数据对接的其具体的实现过程是，首先采访人员选择需要导入的数据，可以单独勾选也可以全选，然后点击导入按钮，在弹出的窗口选择供应商及填写订购批次，点击提交之后程序会自动处理实现订购数据的导入，采访人员可以通过实时进度条查询导入的进度。

导入程序的处理流程是根据选定数据的 ID 查询出 ISBN、价格、题名、作者、出版社、分类号等信息，再通过查询 marc_number 表获取汇文系统的最大记录号，然后生成 MARC 文件插入到 Marc_data 表，分别将 ISBN、题名、作者等关键信息插入到 Marc_idx 索引表及 Marc 图书基本信息表中，最后是在 order_record 订购记录表中加入订购批次、订购数量，在 location_fund_det1 中加入订户信息即采购的数量、经费来源等。具体操作过程及结果展示如图 9 所示。

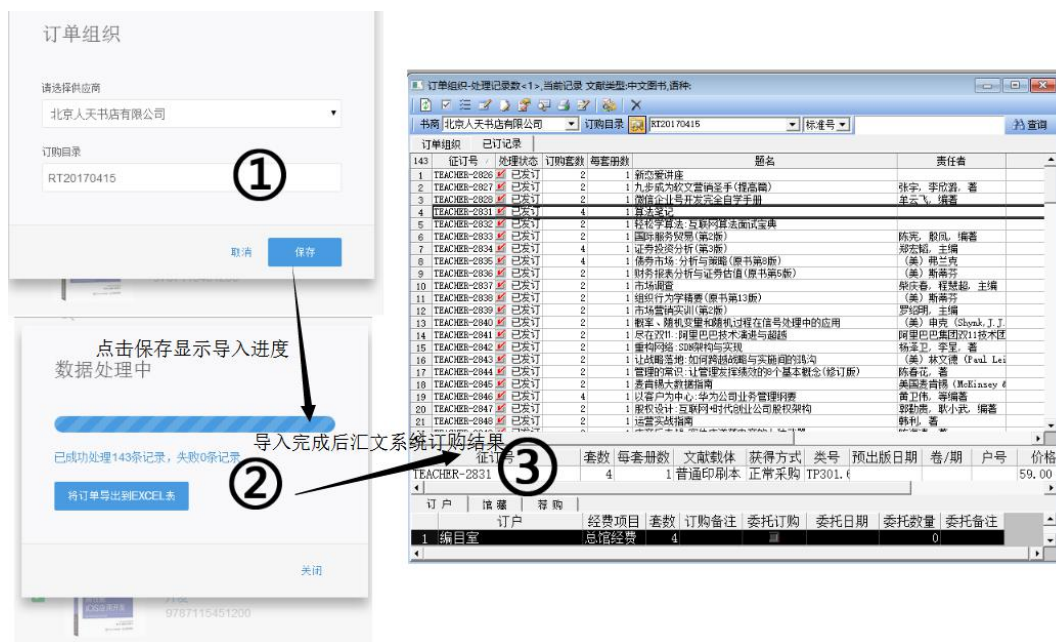


图 9 采访数据导入流程

2.3.3.6 采购订单的导出

该功能主要是实现已经发订数据的导出，方便将订购数据发送给供应商进行采购。项目使用了 PHPExcel 的类进行操作，可以对导出的数据格式进行优化。实现原效果如图 10 所示：

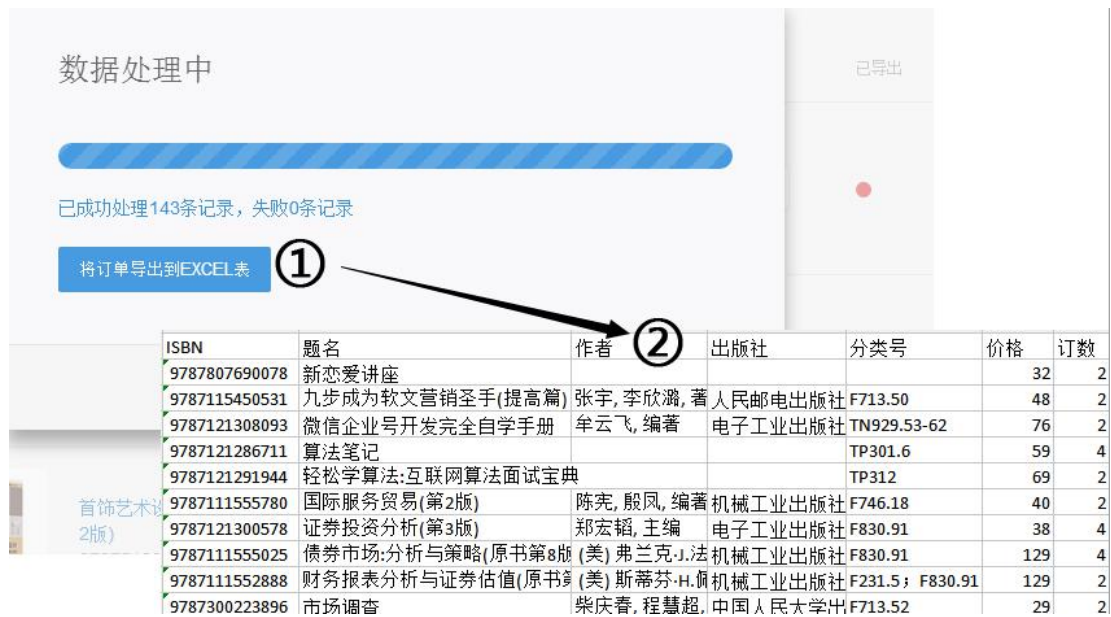


图 10 采访数据导出流程

3 结论与建议

通过对基于 Chrome 扩展采访插件的开发, 证明了该项目的可行性和可用性。该扩展的开发可以大大的节省采访人员获取图书信息的时间, 能够提供更多样化的信息供采访人员进行采购的决策。该扩展的应用大大的方便了对小说类、心理学类等热门、畅销图书采购, 通过访问网站上用户的评论信息、购买数量、排行榜等信息, 可以减少这些类型图书采购时的盲目性, 保障该类图书的及时到馆, 提供给读者借阅, 提高图书的借阅率。该扩展在对于多卷书、专题图书数据的获取拥有较强的优势, 采访人员可以很方便的完成这些图书的补全采购。该项目的成果可以在很少量修改的情况下供使用汇文系统的其它院校使用, 拥有一定的推广性。

本项目目前的功能只能是相对够用, 有很多功能可以在以后的工作和应用当中不断的改进, 比如可以增加批量采集当前页内所有图书信息的功能, 用于对专题页图书, 排行榜图书的采访, 更大限度的提高效率; 同时应该加强扩展程序的不断更新, 由于现有很多网站都在不断的进行着功能和样式的更新, 很可以造成原本能抽取到的信息由于页面构造或样式的调整而无法获取, 为了保证扩展的可用性, 应该不断对程序进行修改; 再次可以增强数据管理的能力, 如增加 Marc 格式数据的导入及导出, 增加订到信息的导入, 增加统计功能等。

4 项目成果（发表的文章、开发的软件、取得的实践效果等）

开发了基于 Chrome 浏览器的采访扩展，并且实现与汇文系统的直接对接，已经用于图书馆的实际采访业务。

5 参考文献

- [1]张云丽. 高校合作开发中文图书采访平台的思考——以武汉大学和厦门大学图书馆为例[J]. 图书馆学研究, 2014, (7): 50-53.
- [2]刘桂杰. 网络平台构建合理图书采购渠道[J]. 辽宁广播电视大学学报, 2009, (1): 111-112.
- [3]魏少鹏,夏小玲. 基于 Chrome 扩展的爬虫系统设计与实现[J]. 软件导刊, 2016, 15(3):76-80.
- [4]沈洪洲. 一种基于 Chrome 扩展程序的网络数据采集方法[J]. 计算机应用与软件, 2016, 33(7): 10-13, 55.
- [5]田秋颐. 高校图书馆网络采访工作的实现与对策[J]. 江西图书馆学刊, 2012, (2): 60-62.
- [6]黎邦群. OPAC 划词翻译功能的设计与实现[J]. 现代图书情报技术, 2011, (11): 92-96.
- [7]乔子健,陈德健,孙艳春. 一种面向 Chrome 浏览器的课堂云笔记插件[J]. 计算机科学, 2015, 42(12): 46-48, 52.
- [8]宋旅黄,赵冉. “互联网”+背景下中文纸质图书采购模式的变革——以三新书业为例[J]. 图书情报知识, 2016, (2): 56-61.
- [9]罗琳. “互联网”+时代高校图书馆图书采购模式的创新与发展[J]. 图书馆学刊, 2016, 38(9): 36-38.
- [10]http://open.se.360.cn/open/extension_dev/overview.html [EB/OL]. 2016-04-06
- [11]<http://www.oracle.com/technetwork/database/features/instant-client/index-097480.html> [EB/OL]. 2016-08-06
- [12]http://www.pc6.com/softview/SoftView_104670.html [EB/OL]. 2016-08-06
- [13]<http://slproweb.com/products/Win32OpenSSL.html> [EB/OL]. 2017-01-06
- [14]<http://www.netmarketshare.com/> [EB/OL]. 2017-04-06