**需求概要**

Step1：在指定网站上抓数据，输出固定格式excel文档A。

Step2：将step1输出的excel文档A与现有的其他文档对比，输出新的excel文档B。

交付：文档A、文档B（可多个excel文件）、游戏封面。

**具体步骤及需求如下：**

1. **在指定网站上爬数据。**
   1. **“gamelist20210830”**。以该list里的游戏列表为基础，在指定网站上搜索对应的游戏（以第2列里的名称搜索）。

“Gamelist20210830“截图如下：

Graphical user interface, text, application

Description automatically generated

在vgtime.com上先搜索” .hack//G.U. Last Recode”，在游戏详情页抓到对应的数据; 接着在网站上搜索“幻想高尔夫“，在游戏详情页里爬到对应的数据; 接着轮到”428被封锁的涩谷”，……以此类推。

1.2、**什么网站，抓什么？**

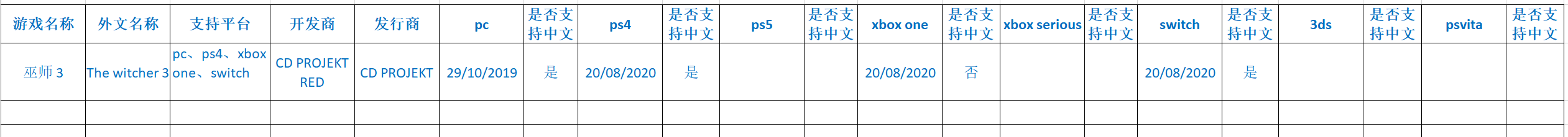
**Vgtime.com；**

**抓如下数据**：

提取游戏名称、外文名称、支持平台、是否支持中文、开发商、发行商、各平台发售时间, 以及游戏封面。

注意，仅关注8个平台：PC, PS4, PS5, XBOX ONE, XBOX SERIOUS, Switch, 3DS, PSVITA

格式和例子如下：



以上可看出，有的平台不支持巫师3，可留空；有的平台支持中文，填写“是”，反之填写“否”。

另外，所有游戏封面单独放一个文件夹，以中文游戏名称命名。

**1.3、步骤一**

* 打开vgtime.com，搜索”gamelist20210830”里的“.hack//G.U.Last Recode”为例：

Graphical user interface, website

Description automatically generated

* 搜索结果页面下，默认是“资讯”，须点击“游戏”，变成如下，先抓取游戏封面：

Graphical user interface, application

Description automatically generated

* 然后进入游戏详情页，抓取游戏中文名与外文名，游戏支持平台、各平台发售时间

Graphical user interface

Description automatically generated

* PS4与PC右边未显示支持中文，所以该游戏各平台均不支持中文，填充为“否”。

Graphical user interface, application

Description automatically generated

**在正常情况下，以巫师3为例，中文/外文名称、支持中文均有，并且有的平台发售时间不一样（指针移到平台上，发售时间会变化），如下**：

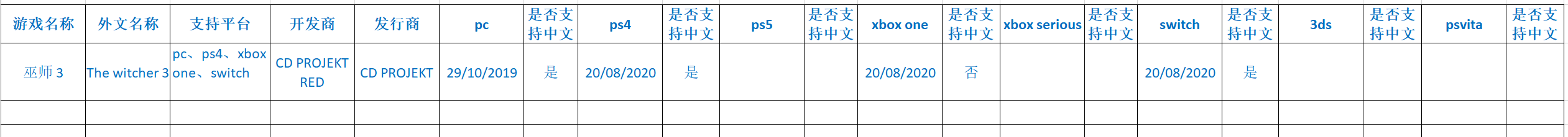
Graphical user interface, application, website

Description automatically generated

以上可看出，发售时间有2行，以“最早发售”为准。

平台右边均显示“中”，表示支持中文。如未显示“中”，则表示该平台不支持中文。

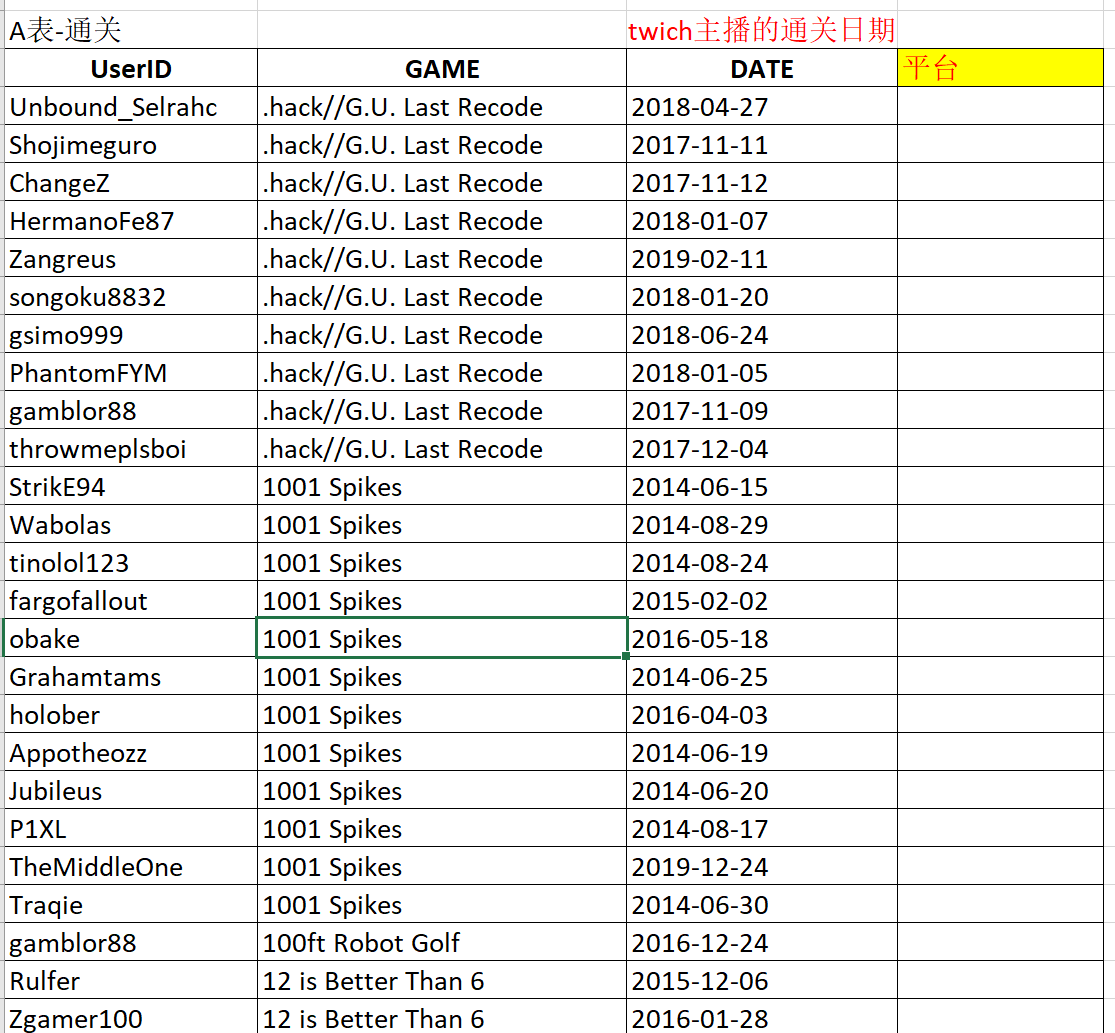
**最后，“gamelist20210830”里的游戏全部抓完后，输出如下格式表（文档A）：**



**如果在vgtime里未搜索到的游戏，或其他异常，请log下来。**

1. **对比文档A与主播表的数据，输出文档B。**

* 目前有一个现成的主播表，如下，求对应平台。



* 将文档A与主播表对比，

Scatter chart

Description automatically generated with medium confidence

**如果未搜索到的游戏，或其他异常，请log下来。**

输出文档B。

1. **交付物为文档A和文档B, 与游戏封面。**