文泉学堂埋点分析

# 概述

COUNTER规范是国外厂商普遍遵循的数字资源使用量统计标准，在国内外图书馆的数字资源采购绩效评估方面发挥着重要的作用。数字资源的统计数据通过COUNTER报告向用户展示。COUNTER标准规定了报告的格式以及需要包含的信息，有一些信息如数字资源的标题、作者、ISBN等是可以从现有数据库获取的，但数字资源的使用量信息（COUNTER报告的核心信息）现在是缺少的，需要在网站中接入埋点以采集使用量数据。经调研发现文泉学堂的数字资源以书籍为主，COUNTER标准规定了16种报告，并非所有COUNTER报告都是需要导出的。本文将针对文泉学堂的实际情况分析需要支持哪些报告，以及如何接入埋点以符合COUNTER统计规范。

# COUNTER报告分析

COUNTER Release 5（以下简称R5）的报告由四个主报告组成，用户可以通过设置过滤器和其他配置信息来创建自定义的视图。R5还为每种主报告定义了不同类型的标准视图，即预设了一些配置信息，用户通过报告的ID即可获取标准视图。COUNTER标准规定内容提供商必须提供**适用于他们的Host\_Type的主报告和标准视图**，始终为空的标准视图除外。所以对于文泉学堂而言，由于绝大部分资源的类型为电子书，所以只要提供适用于电子书的COUNTER即可，其他的报告由于不适用而始终为空，不需要提供。需要提供的报告以及标准视图，还有标准视图是通过哪些配置创建的见表1。

表1 COUNTER报告及标准视图

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主报告 | | 标准视图 | | |
| 报告名称 | 报告内容 | 报告名称 | 报告内容 | 预设的配置 |
| 平台主报告  Platform Master Report, PR | 平台的访问活动 | 平台用量报告  Platform Usage, PR\_P1 | 平台的用量 | **Metric\_Types=**  Searches\_Platform; Total\_Item\_Requests; Unique\_Item\_Requests; Unique\_Title\_Requests  **Data\_Types=**  未设置，采用默认值all  Access\_Method=  Regular  **Attributes\_To\_Show=**  采用默认值none  **granularity=**  month |
| 标题主报告  Title Master Report, TR | 标题访问活动 | 图书请求量报告  Book Requests, TR\_B1 | 非OA\_Gold图书用量 | **Item\_ID=**  未设置，采用默认值all  **Metric\_Type=**  Total\_Item\_Requests; Unique\_Title\_Requests  **Data\_Type=**  Book  **Section\_Type=**  未设置，采用默认值all  **YOP=**  未设置，采用默认值all  **Access\_Type=**  Controlled  **Access\_Method=**  Regular  **Attributes\_To\_Show=**  采用默认值none  **Granularity=**  month |
| 图书拒访量报告  Book Access Denied, TR\_B2 | 图书（题名/种）的拒绝访问量 | **Item\_ID=**  未设置，采用默认值all  **Metric\_Type=**  Limit\_Exceeded; No\_License  **Data\_Type=**  Book  **Section\_Type=**  未设置，采用默认值all  **YOP=**  未设置，采用默认值all  **Access\_Type=**  未设置，采用默认值all  **Access\_Method=**  Regular  **Attributes\_To\_Show=**  采用默认值none  **Granularity=**  month |
| 不同访问类型的图书用量报告  Book Usage by Access Type | 不同访问类型的图书用量 | **Item\_ID=**  未设置，采用默认值all  **Metric\_Type=**  Total\_Item\_Investigations; Total\_Item\_Requests; Unique\_Item\_Investigations; Unique\_Item\_Requests; Unique\_Title\_Investigations; Unique\_Title\_Requests  **Data\_Type=**  Book  **Section\_Type=**  未设置，采用默认值all  **YOP=**  未设置，采用默认值all  **Access\_Type=**  未设置，采用默认值all  **Access\_Method=**  Regular  **Attributes\_To\_Show=**  采用默认值none  **Granularity=**  month |

COUNTER报告主要分为头部信息和正文两个部分，头部主要是报告的基本信息，包括报告的创建时间、报告名称、应用了哪些过滤器、需要展示哪些属性以及异常信息等。报告正文则是包含一个报告项数据的列表，正文的内容根据报告类型和配置的不同而不同。表2.1和表2.2对适用于文泉学堂的JSON报告的各字段值做简要说明。表3列出了适用于文泉学堂的各报告的JSON格式的示例。

表2.1 COUNTER报告头部字段说明

报告头部所需的字段以及值的格式对于所有报告都是一样的

|  |  |
| --- | --- |
| 字段 | 说明 |
| Report\_Name | 报告的名称 |
| Report\_ID | 报告标识符 |
| Release | 报告的版本 |
| Institution\_Name | 报告使用所属机构的名称 |
| Institution\_ID | 报告使用所属机构的ID，以namespace + value的方式确定，namespace允许的值为ISIL，ISNI，OCLC，对于由内容提供商分配的本地标识符，则为内容提供商的平台ID。 |
| Report\_Filters | 应用于报告的过滤器，支持的过滤器有：  Access\_Method、Access\_Type、Begin\_Date、End\_Date、Database、Data\_Type、Item\_Contributor、Item\_ID、Metric\_Type、Platform、Section\_Type和YOP  不同的过滤器适用于不同类型的报告，具体详情可参考COUNTER官网3.3.8节的表3.y |
| Report\_Attributes | 报告的属性，支持的属性有：  Attributes\_To\_Show、Exclude\_Monthly\_Details、Granularity、Inlude\_Component\_Details和Include\_Parent\_Details  不同属性能否使用视报告的类别以及报告是表格形式还是JSON形式而不同，具体详情可参考COUNTER官网3.3.8节的表3.x |
| Exceptions | 生成报告中发生的异常 |
| Created | 使用的日期和时间，格式为RFC3339日期时间格式（yyyy-mm-ddThh：mm：ssZ）。 |
| Created\_By | 创建COUNTER报告的组织或系统的名称。 |
| Customer\_ID | 客户的ID |

表2.2 COUNTER报告正文字段说明

|  |  |  |
| --- | --- | --- |
| 字段 | 说明 | 适用报告 |
| Platform | 标识活动发生的平台/内容宿主 | 所有报告 |
| Data\_Type | 使用的内容的类型,，具体取值见COUNTER规范3.3.2 | PR、DR、TR、IR |
| Access\_Method | 获取方式，具体取值见COUNTER规范4.1.3, 4.2.3, 4.3.3 | 所有报告 |
| Performance | 具体指标情况，其中Period记录起始和结束时间，Instance是一个数组，每个元素是一个指标类型和它的计数 | 所有报告 |
| Title | 报告其使用情况的书籍或期刊的名称。仅适用于标题报告。 | TR, TR\_B1, TR\_B2, TR\_B3, TR\_J1, TR\_J2, TR\_J3, TR\_J4 |
| Item\_ID | 项的标识符 | 除了PR, PR\_P1的报告 |
| Publisher | 内容项的发行商的名称。 | 除了PR, PR\_P1的报告 |
| Publisher\_ID | 发行商的唯一标识符 | 除了PR, PR\_P1的报告 |
| Section\_Type | 当按块或节访问内容时，此属性描述内容单元的性质。有关更多详细信息，参见COUNTER标准3.3.3节 | TR |
| YOP | 报告的项的发行年份。 有关更多详细信息，请参见3.3.7。 | TR, IR, TR\_B1, TR\_B2, TR\_B3, TR\_J4 |
| Access\_Type | 访问类型，有关更多详细信息，请参见3.3.5。 | TR, IR, TR\_B3, TR\_J3, IR\_A1 |

这里只列出了报告中需要用到的字段信息，其他信息暂时不会用到，后续补充进来。注意表格形式的报告和JSON格式的报告所包含的字段会有些不同，如，表格形式的报告Metric\_Type和Begin\_Date/End\_Date要单独拿出来，而JSON格式的报告都归纳到Report\_Filters中去了，如果发现属性与COUNTER报告第4章的描述有出入，可能是JSON格式与表格格式的区别。

表3 COUNTER报告示例

|  |  |
| --- | --- |
| 报告名称 | 报告示例 |
| PR | **{**  **"Report\_Header": {**  **"Created": "2016-09-08T22:47:31Z",**  **"Created\_By": "EBSCO Informtion Services",**  **"Customer\_ID": "12345",**  **"Report\_ID": "TR\_J1",**  **"Release": "5",**  **"Report\_Name": "Journal Requests (Excluding \"OA\_Gold\")",**  **"Institution\_Name": "Mt. Laurel University",**  **"Institution\_ID": [**  **{**  **"Type": "ISNI",**  **"Value": "1234 1234 1234 1234"**  **}**  **],**  **"Report\_Filters": [**  **{**  **"Name": "Begin\_Date",**  **"Value": "2015-01-01"**  **}**  **],**  **"Report\_Attributes": [**  **{**  **"Name": "Attributes\_To\_Show",**  **"Value": "Data\_Type|Access\_Method"**  **}**  **],**  **"Exceptions": [**  **{**  **"Code": 3031,**  **"Severity": "Warning",**  **"Message": "Usage Not Ready for Requested Dates",**  **"Help\_URL": "string",**  **"Data": "request was for 2016-01-01 to 2016-12-31; however, usage is only available to 2016-08-31"**  **}**  **]**  **},**  **"Report\_Items": [**  **{**  **"Platform": "EBSCOhost",**  **"Data\_Type": "Platform",**  **"Access\_Method": "Journal",**  **"Performance": [**  **{**  **"Period": {**  **"Begin\_Date": "2015-01-01",**  **"End\_Date": "2015-01-31"**  **},**  **"Instance": [**  **{**  **"Metric\_Type": "Total\_Item\_Requests",**  **"Count": 21**  **}**  **]**  **}**  **]**  **}**  **]**  **}** |
| PR\_P1 | **{**  **"Report\_Header": {**  **"Created": "2016-09-08T22:47:31Z",**  **"Created\_By": "EBSCO Information Services",**  **"Customer\_ID": "12345",**  **"Report\_ID": "PR\_P1",**  **"Release": "5",**  **"Report\_Name": "Platform Usage",**  **"Institution\_Name": "Mt. Laurel University",**  **"Institution\_ID": [**  **{**  **"Type": "ISNI",**  **"Value": "1234 1234 1234 1234"**  **}**  **],**  **"Report\_Filters": [**  **{**  **"Name": "Begin\_Date",**  **"Value": "2019-01-01"**  **},**  **{**  **"Name": "End\_Date",**  **"Value": "2019-02-28"**  **},**  **{**  **"Name": "Metric\_Type",**  **"Value": "Searches\_Platform|Total\_Item\_Requests|Unique\_Item\_Requests|Unique\_Title\_Requests"**  **},**  **{**  **"Name": "Access\_Method",**  **"Value": "Regular"**  **}**  **],**  **"Exceptions": [**  **{**  **"Code": 3031,**  **"Severity": "Warning",**  **"Message": "Usage Not Ready for Requested Dates",**  **"Help\_URL": "",**  **"Data": "Service has not yet processed the usage for February 2019"**  **}**  **]**  **},**  **"Report\_Items": [**  **{**  **"Platform": "EBSCOhost",**  **"Performance": [**  **{**  **"Period": {**  **"Begin\_Date": "2019-01-01",**  **"End\_Date": "2019-01-31"**  **},**  **"Instance": [**  **{**  **"Metric\_Type": "Searches\_Platform",**  **"Value": 1000**  **},**  **{**  **"Metric\_Type": "Total\_Item\_Requests",**  **"Value": 400**  **},**  **{**  **"Metric\_Type": "Unique\_Item\_Requests",**  **"Value": 300**  **},**  **{**  **"Metric\_Type": "Unique\_Title\_Requests",**  **"Value": 200**  **}**  **]**  **}**  **]**  **}**  **]**  **}** |
| TR, TR\_B1, TR\_B2, TR\_B3 | **{**  **"Report\_Header": {**  **"Created": "2016-09-08T22:47:31Z",**  **"Created\_By": "EBSCO Informtion Services",**  **"Customer\_ID": "12345",**  **"Report\_ID": "TR\_J1",**  **"Release": "5",**  **"Report\_Name": "Journal Requests (Excluding \"OA\_Gold\")",**  **"Institution\_Name": "Mt. Laurel University",**  **"Institution\_ID": [**  **{**  **"Type": "ISNI",**  **"Value": "1234 1234 1234 1234"**  **}**  **],**  **"Report\_Filters": [**  **{**  **"Name": "Begin\_Date",**  **"Value": "2015-01-01"**  **}**  **],**  **"Report\_Attributes": [**  **{**  **"Name": "Attributes\_To\_Show",**  **"Value": "Data\_Type|Access\_Method"**  **}**  **],**  **"Exceptions": [**  **{**  **"Code": 3031,**  **"Severity": "Warning",**  **"Message": "Usage Not Ready for Requested Dates",**  **"Help\_URL": "string",**  **"Data": "request was for 2016-01-01 to 2016-12-31; however, usage is only available to 2016-08-31"**  **}**  **]**  **},**  **"Report\_Items": [**  **{**  **"Title": "Journal of Economics",**  **"Item\_ID": [**  **{**  **"Type": "Print\_ISSN",**  **"Value": "0931-865"**  **}**  **],**  **"Platform": "EBSCOhost",**  **"Publisher": "Springer",**  **"Publisher\_ID": [**  **{**  **"Type": "ISNI",**  **"Value": "1234 1234 1234 1234"**  **}**  **],**  **"Data\_Type": "Journal",**  **"Section\_Type": "Article",**  **"YOP": "2010",**  **"Access\_Type": "Controlled",**  **"Access\_Method": "Regular",**  **"Performance": [**  **{**  **"Period": {**  **"Begin\_Date": "2015-01-01",**  **"End\_Date": "2015-01-31"**  **},**  **"Instance": [**  **{**  **"Metric\_Type": "Total\_Item\_Requests",**  **"Count": 21**  **}**  **]**  **}**  **]**  **}**  **]**  **}** |

# 埋点分析

COUNTER报告中的数据大部分是现有数据库里可以获取到的，但由于使用量（Metric\_Type的各项指标）和用户在网站上的操作有关，是需要采集的，现在的网站还不具备这样的能力，此节将结合文泉学堂的实际情况和COUNTER定义的数据获取规则分析何时采集数据。表4是Metric\_Type包含的各项。

表4 请求和访问的Metric\_Type列表

|  |  |  |
| --- | --- | --- |
| Metric\_Type | 内容 | 适用的报告 |
| 总访问量  Total\_Item\_Investigations | 记录用户对于图书章节的所有**访问**活动（应用了双击过滤）。当用户对于某一章节多次重复访问时，访问量将累加。 | PR, DR, TR, IR, DR\_D1, TR\_B3, TR\_J3 |
| 总项目请求量  Total\_Item\_Requests | 记录用户对于图书章节的所有**请求**活动（应用了双击过滤）。当用户对于某一章节多次重复访问时，访问量将累加 | PR, DR, TR, IR, PR\_P1, DR\_D1, TR\_B1, TR\_B3, TR\_J1, TR\_J3, TR\_J4, IR\_A1, IR\_M1 |
| 唯一项目访问量  Unique\_Item\_Investigations | 记录用户对于图书章节的不重复的**访问**活动。当用户对于某一章节多次重复访问时，访问量不累加。 | PR, DR, TR, IR, TR\_B3, TR\_J3 |
| 唯一项目请求量  Unique\_Item\_Requests | 记录用户对于图书章节的不重复的请求活动。当用户对于某一章节多次重复**获取**时，请求量不累加。如，用户同时获取HTML和PDF版本，请求量只记录为1。 | PR, DR, TR, IR PR\_P1, TR\_B3, TR\_J1, TR\_J3, TR\_J4, IR\_A1 |
| 唯一标题访问量  Unique\_Title\_Investigations | 记录用户对于某种图书的不重复的访问活动。当用户对于某一种图书多次重复访问时，访问量不累加。该指标为电子图书平台提供横向可比的电子书用量统计数据。 | PR, DR, TR, TR\_B3 |
| 唯一标题请求量  Unique\_Title\_Requests | 记录用户对于某种图书的不重复的请求活动。当用户对于某一种图书多次重复获取时，请求量不累加。该指标为电子图书平台提供横向可比的电子书用量统计数据。 | PR, DR, TR, PR\_P1, TR\_B1, TR\_B3 |

除指标类型之外，还有一些细节需要理解

## 访问(Investigation)和请求(Request)之间的区别

访问(Investigation)和请求(Request)之间的区别见图1。由此可见访问包含了请求，所以在用户执行请求操作时，我们需要同时累加访问量和请求量。

图1 请求与访问的关系



## 标题(Title)和项(Item)的区别

项是指图书的某一部分，如摘要，某一章节，参考文献等，标题则是指项所属的图书。很自然的当我们访问了图书的某项，就可以认为是访问了这本书（标题），项和标题的访问量都需要累加。不过要注意，当同一个用户访问同一本书的不同项时，不会对标题访问量重复计数，即只有第一次访问了图书的某项时，才会累加标题访问量。

## 总(Total)和唯一（Unique）的区别

总访问量记录重复的活动，而唯一访问量记录不重复的活动。比如用户重复访问图书的某一章节时，每次总访问量都要累加，而唯一访问量只累加一次，每当访问未访问过的章节时，唯一访问量会累加。

下面将结合文泉学堂的网站页面分析何时采集数据。

当在检索结果页面点击图书标题时（图2）会进入图书详情页（图3），默认会展示图书的简介，可认为是访问了图书的摘要，此时Total\_Item\_Investigation累加, Unique\_Item\_Investigation若未访问过此书摘要则累加，Unique\_Title\_Investigation若未访问过此图书则累加。

图2 图书列表页



图3 图书详情页



若点击了试读

图5 图书目录



图6 试读界面

