

## 有道购物搜索合作商家提交商品数据案例

“合作商家提交商品数据”的服务目前已经被许多商家所采用，在实际应用中，极大地提高了商品数据收录的准确性和及时性。本文档从这些商家中提取典型的案例，介绍成功的设计和实现，帮助更多的商家更快更好的实现商品数据的自动提交。

### 目录

|                                  |       |
|----------------------------------|-------|
| 有道购物搜索合作商家提交商品数据案例.....          | - 1 - |
| 案例一：某家中小规模时尚类网上商城（HTTP+CSV）..... | - 1 - |
| 案例二：某家中小规模B2C商城（HTTP+XML）.....   | - 2 - |
| 案例三：某家中等规模B2C商城（HTTP+CSV）.....   | - 2 - |
| 案例四：国内知名大型网上书城（FTP+CSV）.....     | - 3 - |

### 案例一：某家中小规模时尚类网上商城（HTTP+CSV）

- 商品数量：<1 万
- 商品数据大小：<1MB
- 主要问题：
  1. 该商城商品数据更新频率较快，要求至少每天更新一次数据，并且希望每天只提交更新数据；
  2. 该商城获取商品数据信息比较方便，可以在网页端直接进行查询；
  3. 该商城具有较强的技术实力。
- 解决方案：
  1. 采用方式：http，http 接口：
    - a) st: 更新数据的起始时间，例如：2009-08-10
    - b) et: 更新数据的起始时间，例如：2009-08-19
    - c) token: md5 加密，例如：md5("2009-08-102009-08-19\*\*\*\*")，\*\*\*\*为约定密钥
  2. 数据格式：CSV
  3. 更新方式：有道购物搜索每天凌晨 1 点访问商家提供的 url，url 根据每天的日期自动生成。Md5 加密密钥用于安全控制。
- 点评：这种方式对于商家和有道都是开发成本较少、安全较高、维护较容易的一种方式，推荐商家采用这种方式。

## 案例二：某家中小规模 B2C 商城（HTTP+XML）

- 商品数量：<5 万
- 商品数据大小：<5MB
- 主要问题：
  1. 该商城商品数据更新频率较快，要求至少每天更新一次数据；
  2. 该商城获取商品数据信息比较方便，可以在网页端直接进行查询；
  3. 该商城具有较强的技术实力，并愿意投入时间和精力。
- 解决方案：
  1. 采用方式：http
  2. 数据格式：xml，商家提供了比较完整的数据接口，包括数据的全量和增量模式，每次查询返回结果限制以及翻页的功能。
  3. 更新方式：有道购物搜索每天凌晨 1 点访问该数据接口，每天做增量的更新，每两周下载全量的数据。
- 点评：http+xml 这种模式适合数据量不是特别大、服务器效率较高、有较强研发实力的商家采用。相对而言，xml 格式数据的扩展性更好。

## 案例三：某家中等规模 B2C 商城（HTTP+CSV）

- 商品数量：<10 万
- 商品数据大小：<10MB（gzip 压缩之后）
- 主要问题：
  1. 该商城商品数据更新频率较快，要求至少每天更新一次数据；
  2. 该商城上传 ftp 速度较慢，基本不能上传成功。
- 解决方案：
  1. 采用方式：http
  2. 数据格式：CSV 格式，使用 gzip 压缩；因为 csv 格式数据冗余较少。
  3. 更新方式：商城每天凌晨 0 点制作出一份\*.csv.gzip 数据文件，放在指定 url 下面。有道每天凌晨 1 点对该 url 进行抓取，由商家控制访问 IP 以保证数据安全。
- 点评：http 是比较灵活的方式，如果不想投入太多的时间和精力开发 web 接口，可以采用这种文件下载的方式。

## 案例四：国内知名大型网上书城（FTP+CSV）

- 商品数量：近百万
- 商品数据大小：>100MB（zip 压缩之后）
- 主要问题：商品数量比较多，数据文件比较大；该商家上传 ftp 速度较慢。
- 解决方案：
  1. 采用方式：FTP，因为 FTP 支持断点续传。
  2. 数据格式：CSV 格式，使用 zip 压缩；使用 CSV 格式是因为 CSV 格式数据冗余较少。
  3. 更新方式：每周三，商家从数据库导出商品数据，压缩后手工上传 FTP。
- 点评：对于大型的商城，使用 FTP 是比较稳定可靠地方式。