1. 数据库
2. mozilla.sql 文件包含了数据库表结构和目前抽取的数据。
3. 数据库中的表结构：

security\_id：从<https://www.mozilla.org/en-US/security/advisories/>上找到的所有bug id。

security\_bug：从<https://www.mozilla.org/en-US/security/advisories/>上找到的所有bug 的report数据。其中cc是指report中讨论人数。History则是从report修改历史中抽取出bug状态变化的历史。Fixed\_time等一些时间是对history进行处理得到。但是由于bug状态变动缺少完整的周期，所有有很多缺失时间。

bugs：在security bug的id区间内随机生成bug id，然后抓取相应的bug report，去除其中状态为duplicated的report。数据处理同上。

bug\_cvs\_id：从<https://github.com/ehsan/mozilla-cvs-history.git>这个地址中下载的mozilla-cvs-history。按照倒序访问git commit，如果commit中含有id信息则记录当前步骤数为cvs\_id，同时记录下bug\_id。

bug\_comments和security\_bug\_comments：根据所需要的bug\_id取出对应的cvs\_id集合，按照上述顺序访问git commit，取出相应commit的数据。diff\_file表示修改的文件路径，用“;”分隔；diff\_line表示修改代码的位置，格式为常见的git diff结果格式，没有提取具体代码，每个文件的代码用“;”分隔。

1. Git history
2. 目前找到两个可用的mozilla history数据包。<https://github.com/ehsan/mozilla-cvs-history.git>中的时间跨度稍微大一点，但是并不完整，其中的分支变更信息简单。<https://github.com/ehsan/mozilla-history.git>中的时间跨度小一点，但是数据更加完整，分支变更复杂。两个都可以直接利用git下载。