Quarry Queries

# Query 1

**Query Subject Matter**

Top ANI editors since 2015-08-30

**Username of Query Author**

MusikAnimal

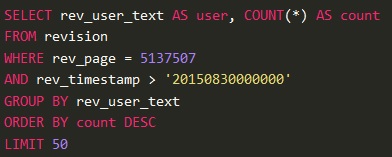
**Database Used**

The queries do not mention the use of any particular database. Hence, the English Wikipedia database has been used by default (enwiki).

**First Query URL**

MusikAnimal. Top ANI editors since 2015-08-30. (n.d.). Retrieved October 20, 2016, from Quarry βeta, <https://quarry.wmflabs.org/query/12186>

**Screenshot**

****

**Description**

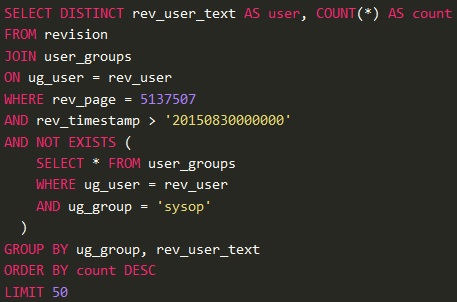
In the query displayed above, the author is displaying the users (user id) and the number of edits made by them to Administrator Noticeboard Incidents on Wikipedia and has grouped the output by user id. He has used the revision table for this purpose which stores the data for every edit done to a page within a wiki. The author has limited the number of output records to 50. Two important checks have been added to this query.

1. All records having a timestamp greater than 30th August 2015 have been considered.
2. The author has added a value for rev\_page and locked the scope of the search to a single article. On further analysis the article of interest was found to be ‘Wikipedia: Administrators’ noticeboard/Incidents/Wiki lobby campaign’. Further reasons as to why the author has worked on this particular article is quite interesting and has been explained in the next two query modifications.

**Second Query URL**

MusikAnimal. Top ANI editors since 2015-08-30. (n.d.). Retrieved October 20, 2016, from Quarry βeta, https://quarry.wmflabs.org/query/12176

**Screenshot**

****

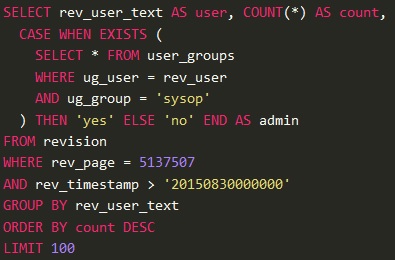
**Description**

On delving deeper into the article ‘Wikipedia: Administrators’ noticeboard/Incidents/Wiki lobby campaign’, I was able to glean that it majorly dealt with an incident wherein a group named ‘Committee for Accuracy in Middle East Reporting in America (CAMERA) ‘ was orchestrating a secret long term campaign to infiltrate Wikipedia and re-write Palestinian history on important Wikipedia pages. Hence a discussion was held by administrators in Wikipedia to talk about the measures to prevent such incidents in the future along with the suitable actions to be taken against the offenders. The author of this query has thus modified the earlier query to find out the contributions of users who are ‘non-admins’ to this discussion. The new modification includes the use of a new table i.e. user\_groups which maps the users to the level of rights assigned to them. Hence, the author has included a subquery which does not select all users who have made contributions to the article and who are admins. The author has then displayed the user id and count of the number of edits made by the user in the output and grouped the same by the user group that the user belongs to and their user id.

**Third Query URL**

MusikAnimal. Top ANI editors since 2015-08-30. (n.d.). Retrieved October 20, 2016, from Quarry βeta, https://quarry.wmflabs.org/query/12187

**Screenshot**

****

**Description**

In the query shown above, the author has made further modifications by shifting the subquery to the ‘SELECT’ clause. The subquery has been included within a ‘CASE’ clause which enables the author to tag a label to the user depending on whether the user is an ‘admin; or not. An additional output column named ‘admin’ has been displayed to include this data. For example, if a user who is an ‘admin’ has made certain contributions to the article, then such users are tagged with a ‘yes’ in the output and vice versa.

# Query 2

**Query Subject Matter**

Articles created by IPs and bots

**Username of Query Author**

ASammour

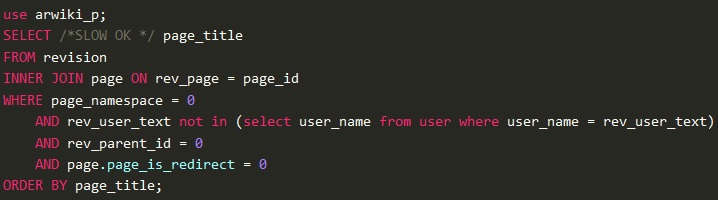
**Database Used**

arwiki\_p (Armenian Wikipedia)

**First Query URL**

ASammour. Articles created by IPs. (n.d.). Retrieved October 20, 2016, from Quarry βeta, https://quarry.wmflabs.org/query/5446

**Screenshot**

****

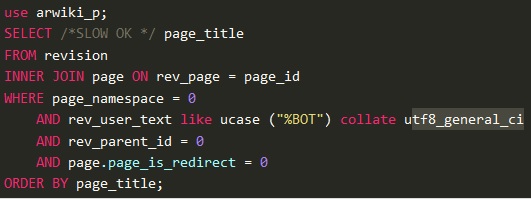
**Description**

In the query displayed above, the author is displaying titles of articles which were edited by users who do not have a registered username on Wikipedia. The data for such users are stored in Wikipedia in the form of their IP addresses. The author has used the revision table for this purpose which stores the data for every edit done to a page within a wiki. In the ‘WHERE’ clause the author has used the user\_name column which belongs to the user table. This column can only store values of user names which are unique and not in the form of IP addresses. Thus by including the user\_name = rev\_user\_text condition, the author has eliminated the possibility of selecting such user names. The author has also included the rev\_parent\_id column and equated it to 0 which means we will only select rows which have been edited for the first time, thereby avoiding duplicates. Using this query the author can find out the articles which have been edited by contributors who haven’t registered with Wikipedia.

**Second Query URL**

ASammour. Articles created by bots. (n.d.). Retrieved October 20, 2016, from Quarry βeta, https://quarry.wmflabs.org/query/5448

**Screenshot**

****

**Description**

The author has revised the earlier query by displaying articles that have been edited by non-users i.e. bots. The author has used the revision and page tables in a similar context as the previous query. The important change made by the author is the inclusion of the condition rev\_user\_text like ucase (“%BOT”) which identifies edits made by non-users. The articles fetched will then be collated by the UTF-8 standard. The results of this query can help in recognizing the different articles which have been edited by non-humans or programs. In a similar vein Wikipedia admins can use this query to view users or admins who are contributing to various articles.

# Query 3

**Query Subject Matter**

Types of files

**Username of Query Author**

Stefan2

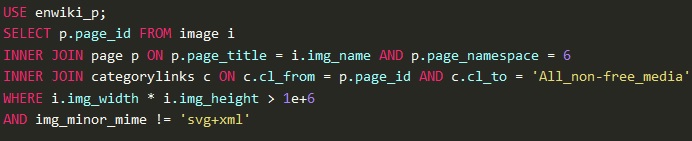
**Database Used**

Enwiki\_p (English Wikipedia)

**First Query URL**

Stefan2. Very big non-free files. (n.d.). Retrieved October 20, 2016, from Quarry βeta, https://quarry.wmflabs.org/query/2888

**Screenshot**

****

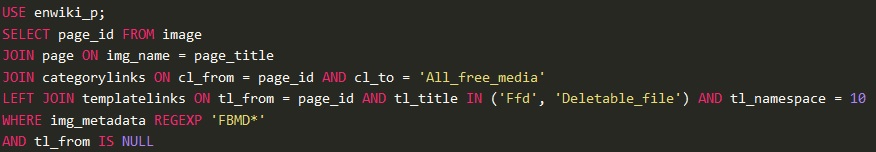
**Description**

In the query displayed above, the author is displaying the page id’s of media description pages (denoted by a page\_namespace value of 6) containing links to non-free media files with big sizes. To determine which files can be considered as big, the author has specified dimensions for the same by using the img\_width and img\_height columns present in the image table. The image table has details for all images and uploaded files. Furthermore, to ensure that the media being referenced is ‘Non Free’, the author has included the categorylinks table. The author has also specified the type of files being queried in the img\_minor\_mime column. The results of this query can identify page id’s of articles which have huge files embedded in them.

**Second Query URL**

Stefan2. Facebook files. (n.d.). Retrieved October 20, 2016, from Quarry βeta, https://quarry.wmflabs.org/query/9630

**Screenshot**

****

**Description**

The query shown above is an extension of the previous query functionality. In this query, the author is displaying the page id’s of pages containing Facebook links. As opposed to the previous query, the author has ensured that the content being referenced falls under the ‘All\_free\_media’ category (which makes sense since Facebook is free). The author has also ensured that the media being referenced is not a template and the key to the target Facebook page cannot be deleted. The referenced file’s property has also been made equivalent to ‘Facebook metadata’ which is the indicator that we are referencing Facebook files or media. Word Count: 1172