INFM 600 - Quarry Queries

**User 1 - Kenrick95**

Kenrick is user from Indonesia and is found to be active on Indonesia Wikipedia, (Kenrick, 2015). Following are a set of queries by Kenrick:

1. **Query 1 - https://quarry.wmflabs.org/query/11200  
   Title - List of Articles created as part of Punjab Editathon - 2016**

This query is based on the WikiConference India 2016, (“WikiConference India 2016/Punjab Edit-a-thon,” 2016), which targeted on adding or improving articles related to Punjab, a state in India. This particular query lists out all the articles created as a part of the Punjab Edit-a-thon.

The list has been extracted from the *page* and *categorylinks table* in the Wikimedia database. The *page table*, (“Manual: Page table,” 2006), is considered as the “core of the Wiki”. This table contains a page id for each page, the page title, a page counter to keep a track of how many times the page has been viewed and many other such attributes that define a page. However, the entire content of the page is stored in the *text table.* The *categorylinks table*, (“Manual: Categorylinks table,” 2005),stores entries into a category with the help of [[Category:*Title*]] or [[Category:*Title*|*sortkey*]]. *cl\_from* attribute stores the page and *cl\_to* attribute stores the name of the desired category.

What the user is trying to do here is to find out all the articles in which the edits were made as part of the Punjab Edit-a-thon. By extracting this, one can see the total number of articles that were edited during the competition. This is done by sorting the articles using the category name “Articles created as part of Punjab Editathon - 2016”. The query results can help in the post-competition documentation and analytics.

1. **Query 2 - https://quarry.wmflabs.org/query/11197  
   Title - Revisions on Articles created as part of Punjab Editathon - 2016**

This query is related to the above query and is also based on the WikiConference in India 2016, (“WikiConference India 2016/Punjab Edit-a-thon,” 2016).

It extracts all the edits made in the articles under the category of “Articles created as part of Punjab Editathon - 2016”. This has been achieved using the *categorylinks table* and the *revision table.* The *revision table*, (“Manual: Revision table,” 2006), contains the metadata for every edit done on a wiki page. With the help of this metadata we can extract the revisions made under a category. This can help the organizers of the event understand which all edits made are genuine and which all are just repetitive and unnecessary.

**User 2 - James Hare (NIOSH)**

James Hare is a paid contractor for National Institute for Occupational Safety and Health, (Hare, 2015). He is a popular personality on Wikimedia projects and is known as Harej. WikiProject Occupational Safety and Health is a merger of WikiProject NIOSH and WikiProject Occupational Safety, (“Wikipedia: WikiProject occupational safety and health,” 2015). Following are the queries he has worked on regarding the articles that link NIOSH to the WikiProject OSH.

1. **Query 1 - https://quarry.wmflabs.org/query/9773  
   Title - Articles tagged by WikiProject OSH that do not link to NIOSH**

This query extracts all the articles in Wiki that were tagged by Occupational Safety and Health project that do not link to National Institute for Occupational Safety and Health. This basically includes a list of occupational hazards that have Wiki links but its scope is still potential. Since they are yet to be explored by this WikiProjects, they have not yet been ‘tagged’.

This result has been achieved by joining the *page table* with the *categorylinks table* and joining the results with the *externallinks table.* This extracts all the articles that were under the category “WikiProject Occupational Safety and Health articles” but have no external links in the NIOSH website. This query can help extract the potential pages that the WikiProject needs to work on that have been tagged.

1. **Query 2 - https://quarry.wmflabs.org/query/9774  
   Title - Articles that link to NIOSH \*not\* tagged by WikiProject OSH**

This query does somewhat the exact opposite of the previous query. In this query the user is trying to extract all those articles that have external links to the NIOSH website but have not been categorized under “WikiProject Occupational Safety and Health articles”, ie., not tagged by WikiProject OSH.

This is achieved by joining the *externallinks table* with the *page table* and joining the resulting table with the *categorylinks table.* The *externallinks table*, as the name suggests, stores all the external links to a page in it, (“Manual: Externallinks table,” 2006).

This query can help the WikiProject members see if the records revealed are actually not related to Occupational Safety and Health or have been not been tagged by mistake.

**User 3 - Blahma**

Marek Blahuš is a user who is from Czech and often works with cswiki, (Blahuš, 2006). Following are the set of queries which deals with links to Masaryk University and its staff and students.

1. **Query 1 - https://quarry.wmflabs.org/query/12756  
   Title - Biographies on cswiki linking Masaryk University but in no related category**

This query extracts all the pages that are related to Masaryk University but do not have a category link to any of the Masaryk University content. This is done by using the *pagelinks table, page table* and *categorylinks table.* The *pagelinks table* tracks all the internal links in the Wiki, (“Manual: Pagelinks table,” 2006).

So the user tracks all the pages that have page links related to Masaryk University but do not have a category link to Masaryk University. This query can help identifying such pages and including them in the category so as to increase the visibility of those Wiki pages.

1. **Query 2 - https://quarry.wmflabs.org/query/12759  
   Title - Masaryk University staff and alumni with no link to university or faculty in article text**

This query is also part of the series queries done by the user which tries to find out if particular wiki pages related to Masaryk University are linked to it completely or not. This query extracts all the pages of Masaryk University staff and alumni who have category links to Masaryk University Teachers but do not have page links to Masaryk University or faculty in the article at all.

**Word Count - 1023 words**

**References**

1. Kenrick. (2015). User: Kenrick95. Retrieved October 17, 2016, from https://meta.wikimedia.org/wiki/User:Kenrick95
2. WikiConference India 2016/Punjab Edit-a-thon. (2016). Retrieved October 17, 2016, from https://meta.wikimedia.org/wiki/WikiConference\_India\_2016/Punjab\_Edit-a-thon
3. Manual: Page table. (2006, January 26). Retrieved October 17, 2016, from https://www.mediawiki.org/wiki/Manual:Page\_table#page\_latest
4. Manual: Categorylinks table. (2005, April 13). Retrieved October 17, 2016, from https://www.mediawiki.org/wiki/Manual:Categorylinks\_table
5. Manual: Revision table. (2006, January 26). Retrieved October 17, 2016, from https://www.mediawiki.org/wiki/Manual:Revision\_table
6. Hare, J. (2015, August 03). Retrieved October 18, 2016, from https://meta.wikimedia.org/wiki/User:James\_Hare\_(NIOSH)
7. Wikipedia: WikiProject occupational safety and health (2015). . In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Wikipedia:WikiProject\_Occupational\_Safety\_and\_Health
8. Manual: Externallinks table. (2006, May 22). Retrieved October 18, 2016, from https://www.mediawiki.org/wiki/Manual:Externallinks\_table
9. Blahuš, M. (2006). Usuario: Blahma. In *Wikipedia*. Retrieved from https://es.wikipedia.org/wiki/Usuario:Blahma
10. Manual: Pagelinks table. (2006, January 26). Retrieved October 18, 2016, from https://www.mediawiki.org/wiki/Manual:Pagelinks\_table