

Using Technical **Tools** to Research **Pain Points** of **Wikimedia** Projects

Nivas, Navya & Vasanth

Indic MediaWiki Developers User Group

Introduction



1 The WHY?

2 The HOW?

3 Discussion

The Why?

The How?



Leveraging Existing **Tools**...



Quarry

<https://quarry.wmcloud.org/>



Stats @ Wikimedia

<https://stats.wikimedia.org/>



PageViews

<https://pageviews.wmcloud.org/>

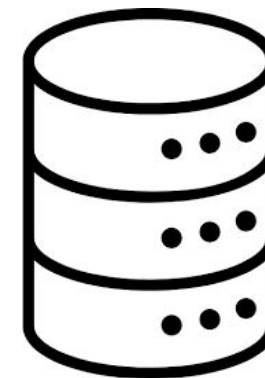


Misc

<https://translatetagger.toolforge.org/>

How to retrieve **insights** from Wikimedia projects with existing tools?

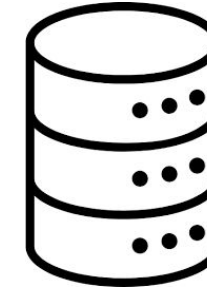
Tools: Quarry, SQL



https://www.mediawiki.org/wiki/Manual:Database_layout

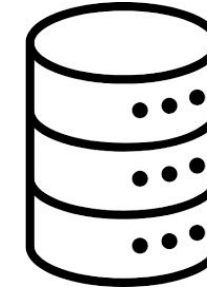
<https://quarry.wmcloud.org/>

Edit Contributions of రచ్చబండ Participants Over Time: 3, 6, 12, and 24 Months



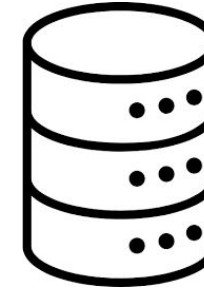
```
SELECT
  COUNT(rev_id) AS edit_count,
  CASE
    WHEN rev_timestamp >= DATE_FORMAT(NOW() - INTERVAL 3 MONTH, '%Y%m%d%H%i%s') THEN 'Last 3 Months'
    WHEN rev_timestamp >= DATE_FORMAT(NOW() - INTERVAL 6 MONTH, '%Y%m%d%H%i%s') THEN 'Last 6 Months'
    WHEN rev_timestamp >= DATE_FORMAT(NOW() - INTERVAL 12 MONTH, '%Y%m%d%H%i%s') THEN 'Last 12 Months'
    WHEN rev_timestamp >= DATE_FORMAT(NOW() - INTERVAL 24 MONTH, '%Y%m%d%H%i%s') THEN 'Last 24 Months'
    ELSE 'Older'
  END AS period
FROM
  revision
  JOIN page ON page_id = rev_page
WHERE
  page_title = 'రచ్చబండ' -- Adjust this to the correct page_title according to language.
  AND page_namespace = 4
  AND rev_timestamp >= DATE_FORMAT(NOW() - INTERVAL 24 MONTH, '%Y%m%d%H%i%s')
GROUP BY
  period;
```

User Engagement on రచ్చబండ Page Over Time: 3, 6, 12, and 24 Months



```
SELECT
  COUNT(DISTINCT actor_name) AS user_count,
  CASE
    WHEN rev_timestamp >= DATE_FORMAT(NOW() - INTERVAL 3 MONTH, '%Y%m%d%H%i%s') THEN 'Last 3 Months'
    WHEN rev_timestamp >= DATE_FORMAT(NOW() - INTERVAL 6 MONTH, '%Y%m%d%H%i%s') THEN 'Last 6 Months'
    WHEN rev_timestamp >= DATE_FORMAT(NOW() - INTERVAL 12 MONTH, '%Y%m%d%H%i%s') THEN 'Last 12 Months'
    WHEN rev_timestamp >= DATE_FORMAT(NOW() - INTERVAL 24 MONTH, '%Y%m%d%H%i%s') THEN 'Last 24 Months'
    ELSE 'Older'
  END AS period
FROM
  revision
  JOIN page ON page_id = rev_page
  JOIN actor ON actor_id = rev_actor
WHERE
  page_title = 'రచ్చబండ'
  AND page_namespace = 4 -- Adjust this to the correct namespace if needed.
  AND rev_timestamp >= DATE_FORMAT(NOW() - INTERVAL 24 MONTH, '%Y%m%d%H%i%s')
GROUP BY
  period;
```


Top 10 **thanked** users



```
SELECT
  actor_name AS thanked_user,
  COUNT(*) AS thanks_count
FROM
  logging
JOIN actor ON actor_id = log_actor
WHERE
  log_type = 'thanks'
  AND log_timestamp BETWEEN '20240901000000' AND '20240930235959'
GROUP BY
  actor_name
ORDER BY
  thanks_count DESC;
```

thanked_user	thanks_count
Pranayraj1985	37
Oirattas	11
Chaduari	6
యర్ల రామారావు	5
Doc James	5
Mothiram 123	3
Saiphani02	3
డా. గన్నవరపు నరసింహమూర్తి	3
K.Venkataramana	2
Arjunaraoc	2
వైజాసత్య	1
Rajasekhar1961	1
MGA73	1
స్వరలాసిక	1
Muralikrishna m	1

Showing 1 to 15 of 15 entries

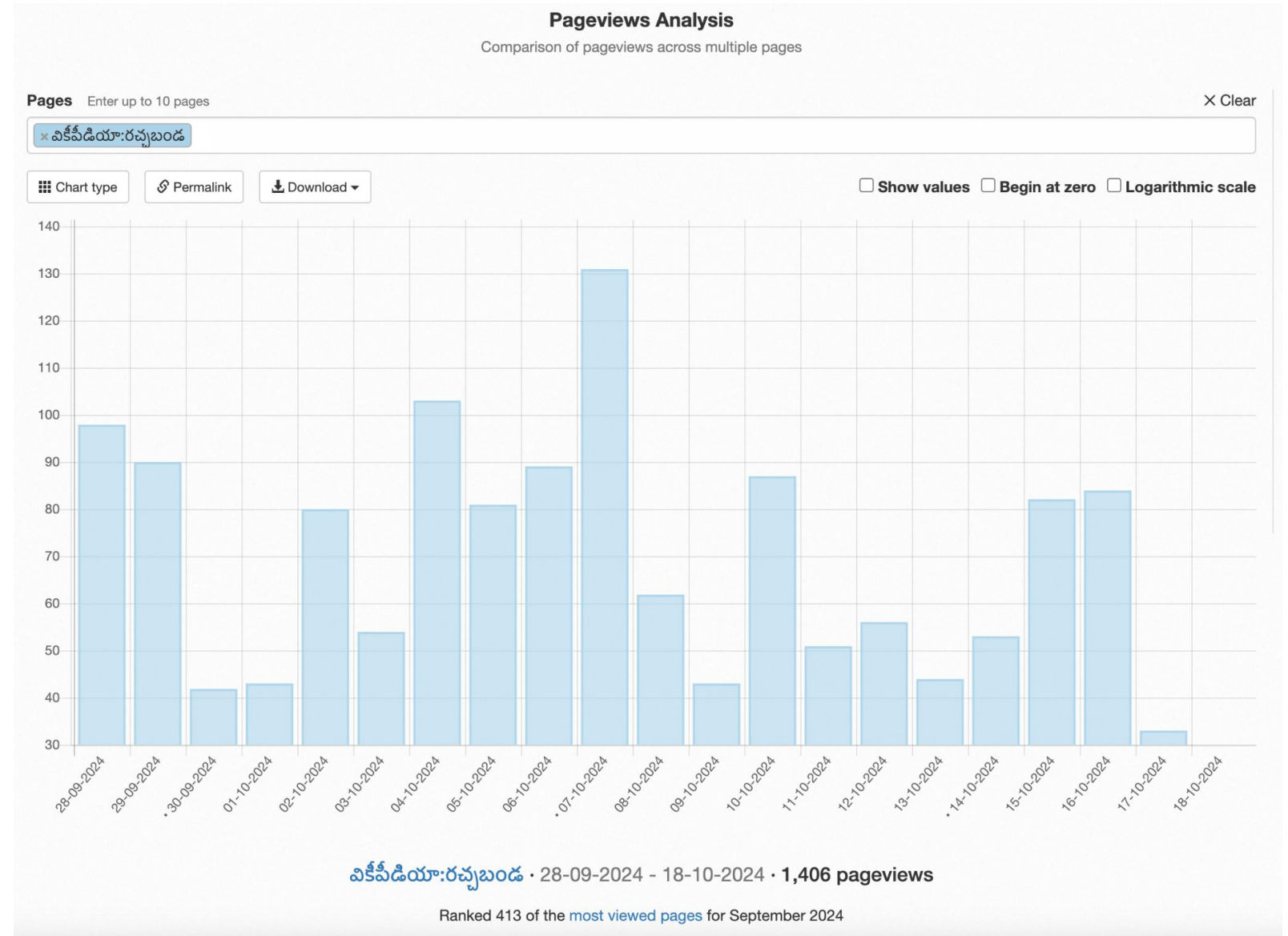
<https://quarry.wmcloud.org/query/87261>

User Engagement Metrics for రచనబండ Page

Rolling over 3, 6, 12,
and 24 months

[Page view analysis](#)

<https://pageviews.wmcloud.org/>



Something **more**...

Wiki Translate Tagger

Input Wikitext

```
== Program Schedule ==
{| class="wikitable"
!SNo
!Program
!Time
!Speakers
|-
| colspan="3" |'"Day 1 - October 19, 2024, Saturday'"
|
|-
|1
|Welcome & UCoC
|9:30 am - 9:45 am
|Nitesh, Chinmayee, and Tanveer
|-
```

Convert to Translatable Wikitext

Translatable Wikitext Output

```
== <translate>Program Schedule</translate> ==
{| class="wikitable"
!<translate>SNo</translate>
!<translate>Program</translate>
!<translate>Time</translate>
!<translate>Speakers</translate>
|-
| colspan="3" |<translate>''"Day 1 - October 19, 2024, Saturday'"</translate>
|
|-
|<translate>1</translate>
|<translate>Welcome & UCoC</translate>
|9:30 am - 9:45 am
|<translate>Nitesh, Chinmayee, and Tanveer</translate>
|-
```

Copy to Clipboard



Prompt Engineering for Query Generation



Prompts for queries



Can you help me write an SQL query that counts the number of edits made on a specific page in the last 3, 6, 12, and 24 months? The query should:

“ “ “

- Use the revision and page tables from the MediaWiki database.
- Filter edits for a specific page with `page_title = 'ಶೃಂಗೇರಿ'` and `page_namespace = 4`.
- Use a CASE statement to categorize the edits into time periods (3 months, 6 months, 12 months, 24 months).
- Group the results by each time period and return the count of revisions (`edit_count`) for each.
- Please include conditions to ensure that only revisions within the last 24 months are considered.

“ “ “



Good Practices for prompting...

1. Be **specific** and detailed about desired output along with format
2. Seperate the **question** and **instructions**
3. Follow an **iterative** approach
4. more...

Reference for examples: [Open AI Help](#)

Thank you!



program@cis-india.org /
contact@indicmediawikidev.org

[Indic MediaWiki Developers User Group - Meta](#)