



WikiSHARK

An Online Tool for Analyzing
Wikipedia Traffic and Trends

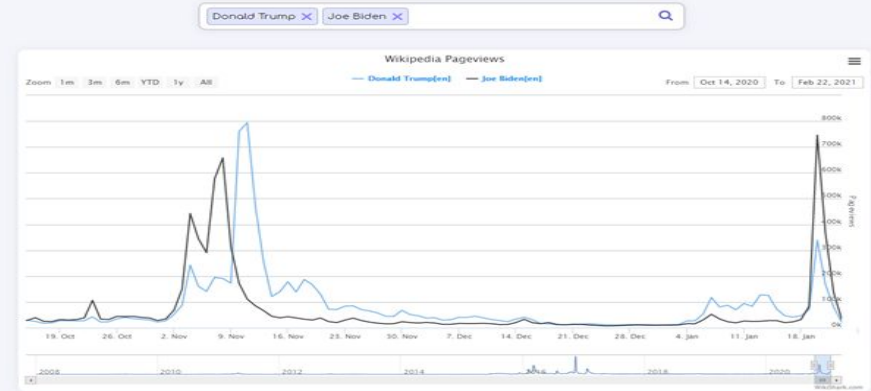
Elad Vardi
Lev Muchnik
Alex Conway
Micha Y. Breakstone



Wiki Workshop 2021

WikiShark - An Online Tool for Analyzing Wikipedia Traffic

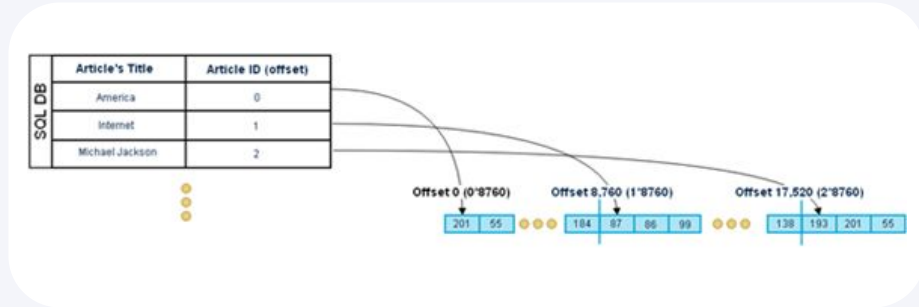
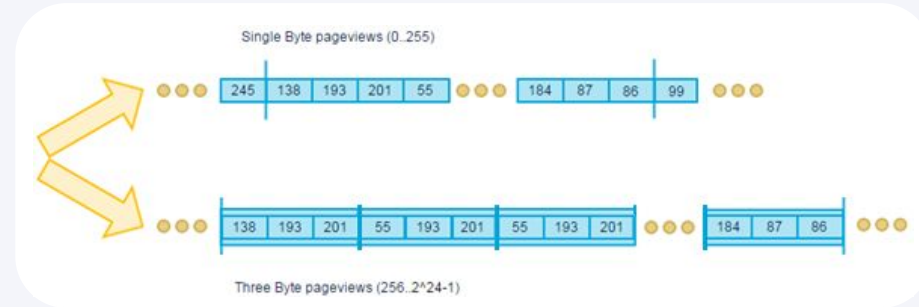
- Available at www.wikishark.com
- All data since 2008 is query-able
- Data is updated hourly
- Features in include:
 - Traffic querying (pageviews) for any wikipedia title
 - Full trends support
 - Comparing Traffic of 2 or more Wikipedia Titles
 - Data Can be downloaded as a data file or as an image
 - Chrome Extension for in-page data plotting



How is the data stored?

- Transposed (by title and not by time)
- In a minimal size binary file
- Sequential
- We hold a key-value dictionary for $O(1)$ title data offset extraction

This allows efficient data extraction and performing big data queries



if you're interested in collaborating with us on either data or analysis,
we're happy to explore such opportunities - elad@wikishark.com