New readership data

Some things we've been learning recently about how Wikipedia is read

Wikimedia Foundation Tech Talk, March 18, 2016
Tilman Bayer (Senior Analyst, Reading department)

About this talk

Various recent insights and new sources of data on how readers read Wikipedia (or: a bit of a grab bag of interesting stuff ;)

Not one research project, rather some byproducts of data analysis done by the WMF Reading team, with related findings by others at WMF, and academic researchers
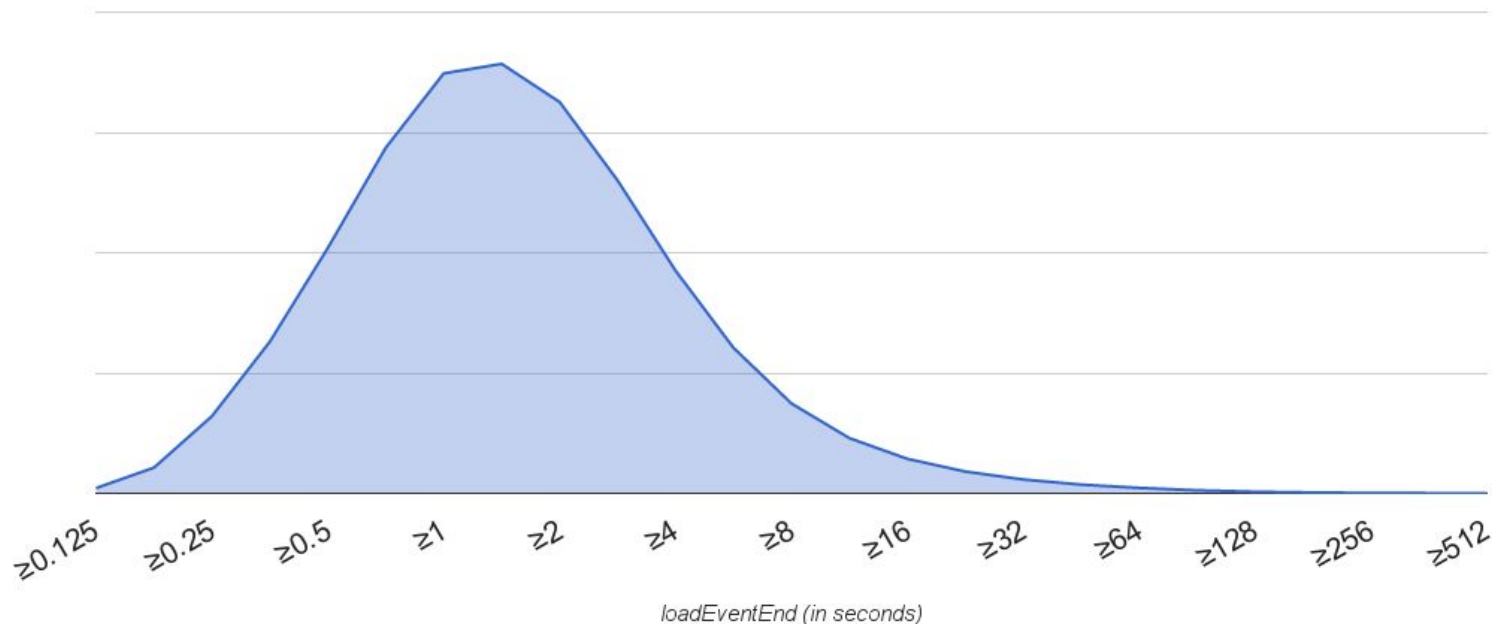
Not about:

- Pageview data (well studied already, and now available via API)
  - Big picture: Desktop still >50% but decreasing, majority of users already on mobile web (but with less pageviews per reader), apps ~1%
- New core metrics on uniques etc. (coming soon)

Structure of this presentation

- About this talk
- How long do readers have to wait?
- Which parts of an article do readers read? (geometrically)
- Which parts of an article do readers read? (by topic)
- Which parts of an article do readers read? (other aspects)
- Other current WMF research on readers
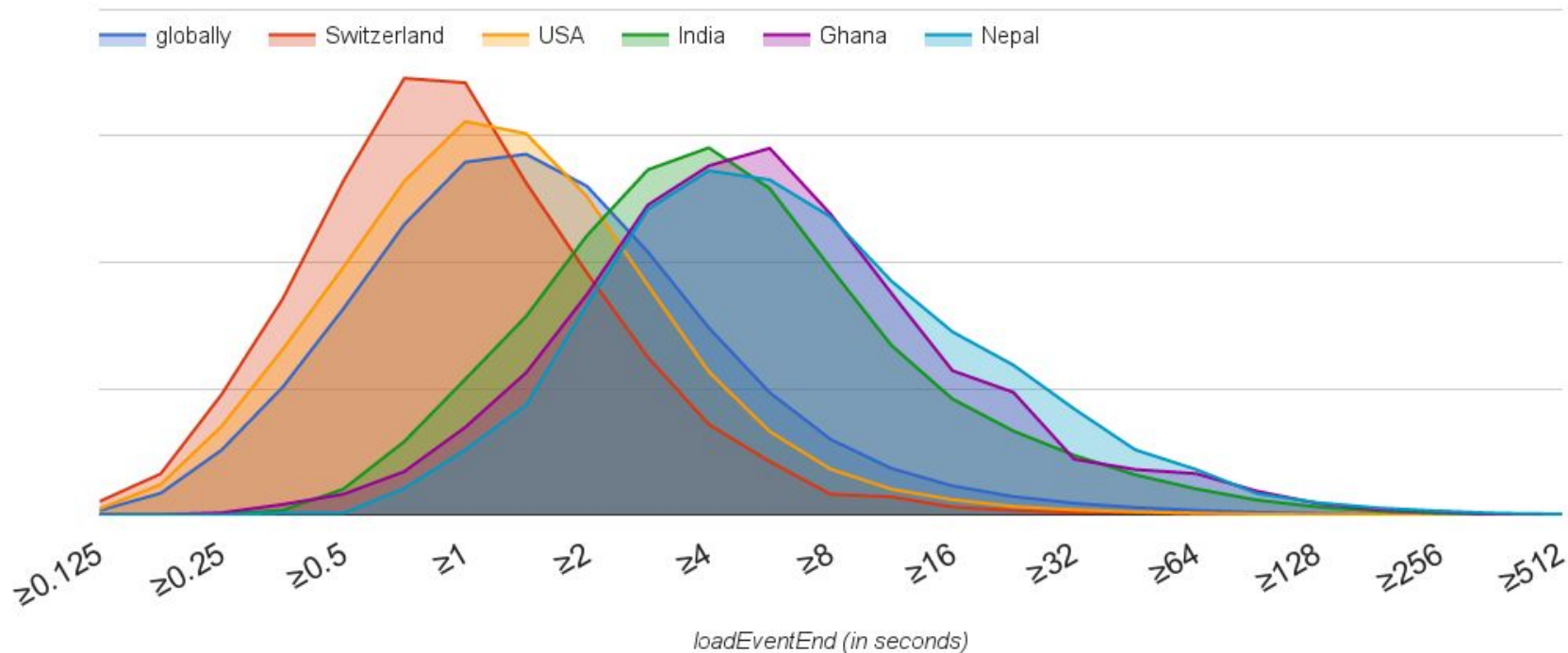
How long do readers have to wait?

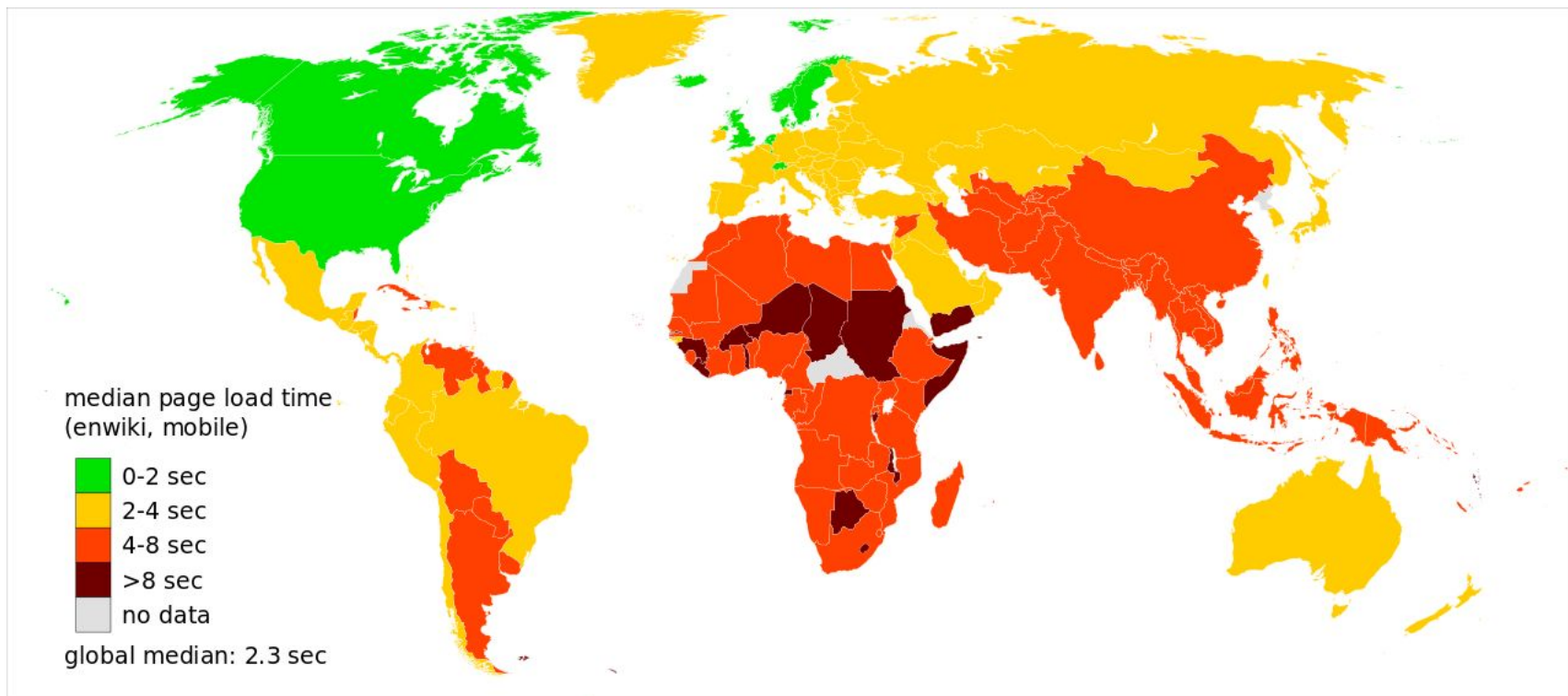**Load time distribution (desktop and mobile web, globally)**

loadEventEnd (in seconds)

Time until a page has completed loading (loadEventEnd, enwiki, mobile and desktop)
Depends on connection speed, page size, ...

How long do readers have to wait?

Load time distribution in various countries (desktop and mobile web)

Legend: globally | Switzerland | USA | India | Ghana | Nepal

x-axis: ≥0.125 ≥0.25 ≥0.5 ≥1 ≥2 ≥4 ≥8 ≥16 ≥32 ≥64 ≥128 ≥256 ≥512

loadEventEnd (in seconds)

How long do readers have to wait?

median page load time
(enwiki, mobile)

- 0-2 sec
- 2-4 sec
- 4-8 sec
- >8 sec
- no data

global median: 2.3 sec

Median time until a page has completed loading (loadEventEnd, enwiki, mobile only)

How long do readers have to wait?

Caveat: Data does not cover readers whose connection times out or who leave the page before it finishes loading.

A lot of recent and ongoing work by the Foundation's Performance and Reading teams addresses readers on slow connections, see list:

- T113066 "Make Wikipedia more accessible to 2G connections"

Which parts of an article do readers read?

We don't know much about this yet

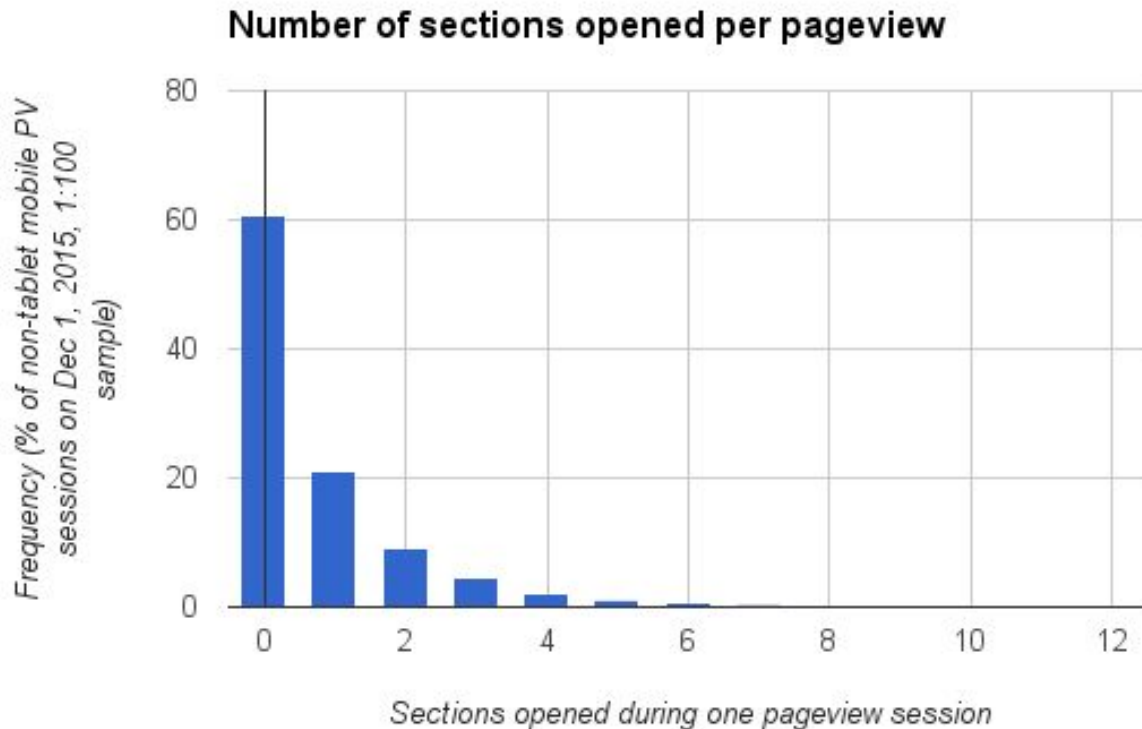But there are now some new data sources.

Data source 1: section expansion actions on mobile Wikipedia

- By default, all top-level section below the lead section are shown collapsed
- Users can tap on section heading to expand

How far do readers read into an article?

On ca. 60% of mobile pageviews, only the lead section is looked at.

Number of sections opened per pageview

Frequency (% of non-tablet mobile PV sessions on Dec 1, 2015, 1:100 sample)

Sections opened during one pageview session

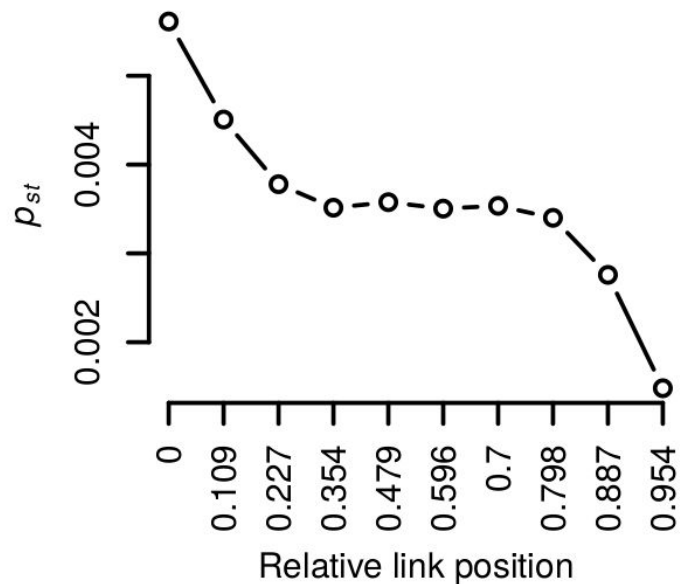# How far do readers read into an article?

Data sources from other instrumentations done for Product development:

- During one week in 2014, **25% of app users** (devices) scrolled to the end of a page at least once.
- In 2016, the end of a page was seen on **40% of pageviews on desktop** (Vector) and **32% of pageviews on mobile web beta** (caveat: within an opt-in group of readers).

Which parts do readers read? Click actions

Data source: [Clickstream dataset](#) (published by WMF in 2015 - English WP, desktop)

- Click rate decreases from top to bottom (see chart)
- Wikilinks in the lead section receive between 26% and 43% of clicks .
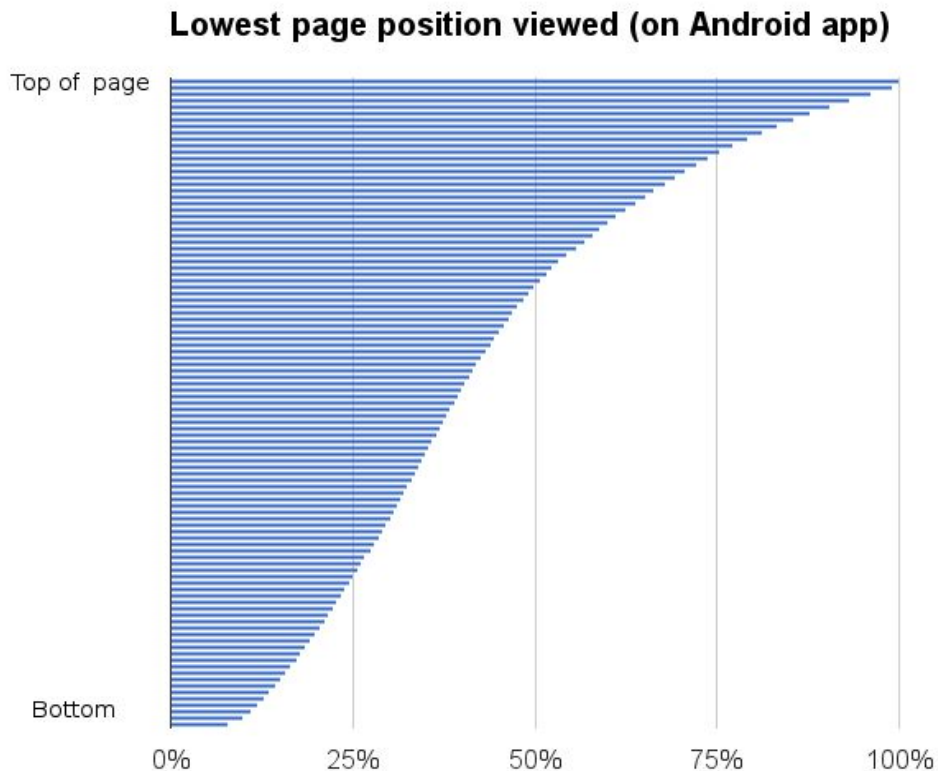- Wikilinks on the left side of the screen are more likely to be clicked.

Ashwin Paranjape, Bob West, Jure Leskovec, Leila Zia: Improving Website Hyperlink Structure Using Server Logs. WSDM'16, February 22–25, 2016, San Francisco, CA, USA. PDF
Lamprecht, Daniel; Helic, Denis; Strohmaier, Markus (2015-04-22). "Quo Vadis? On the Effects of Wikipedia's Policies on Navigation". Ninth International AAAI Conference on Web and Social Media. Ninth International AAAI Conference on Web and Social Media.
Dimitrov, Dimitar; Singer, Philipp; Lemmerich, Florian; Strohmaier, Markus (2016-04-11). Visual Positions of Links and Clicks on Wikipedia (PDF). 25TH INTERNATIONAL WORLD WIDE WEB CONFERENCE. Montréal, Québec, Canada. p. 2.

Clickthrough rate of wikilinks by their position in a page (left = top, right = bottom)

Which parts do readers read? Scroll actions

Android app instrumentation

(records lowest page position viewed)

Lowest page position viewed (on Android app)
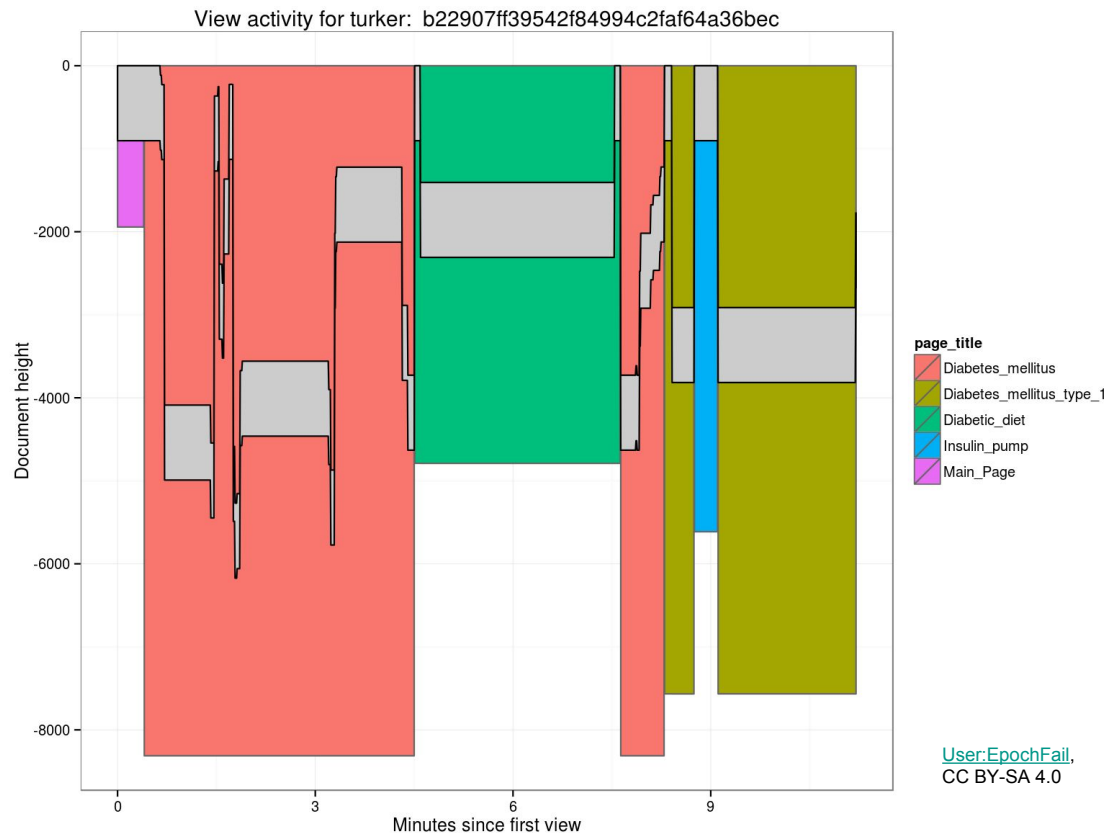
Top of page

Bottom

0%   25%   50%   75%   100%

Which parts do readers read? Scroll actions
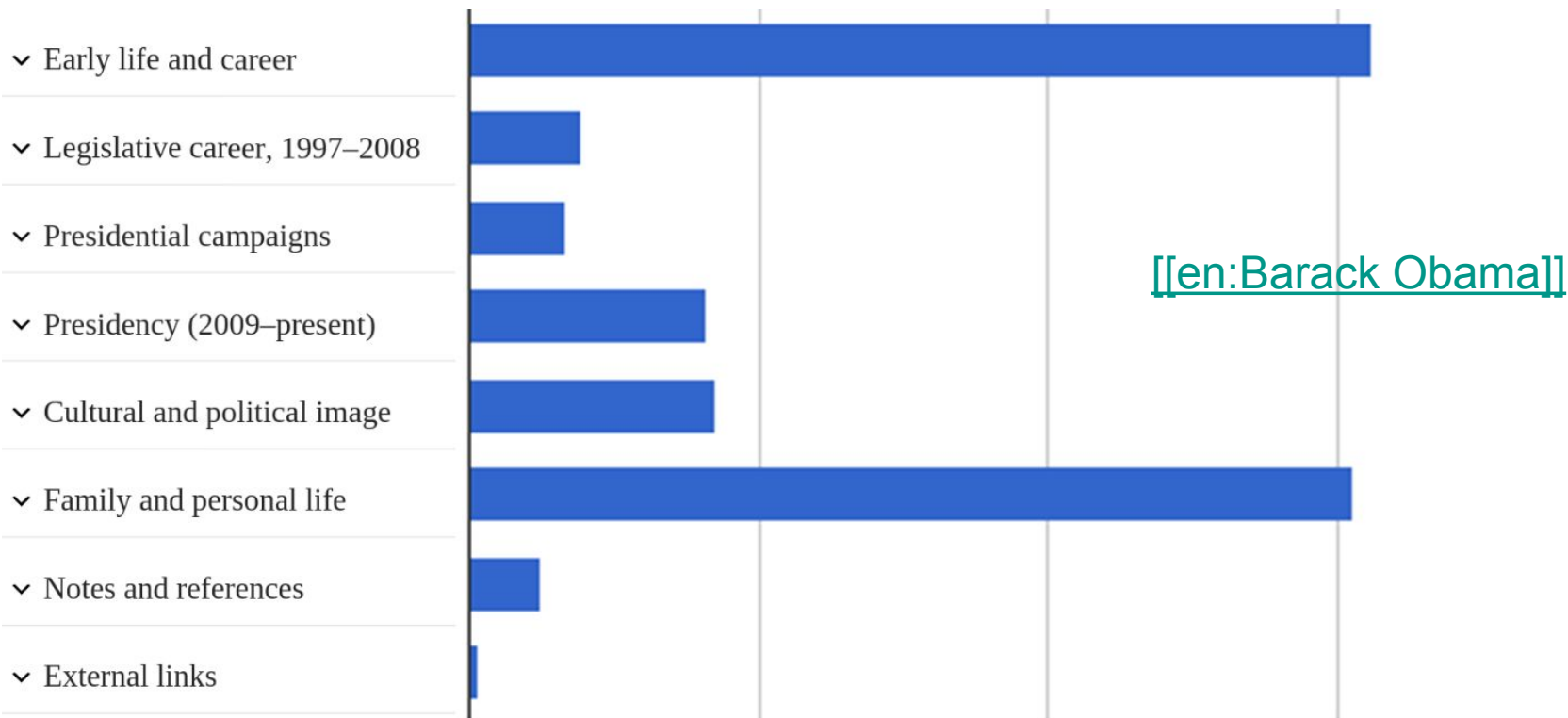
Small study with Amazon Mechanical Turk users

Figure: The movement of the view pane (grey) over time in different articles (colored) while a reader is researching questions about diabetes.

Aaron Halfaker, Kenneth Shores, Michael Zarro: Research:Directed diabetes info-seeking behavior in Wikipedia (Meta-wiki)

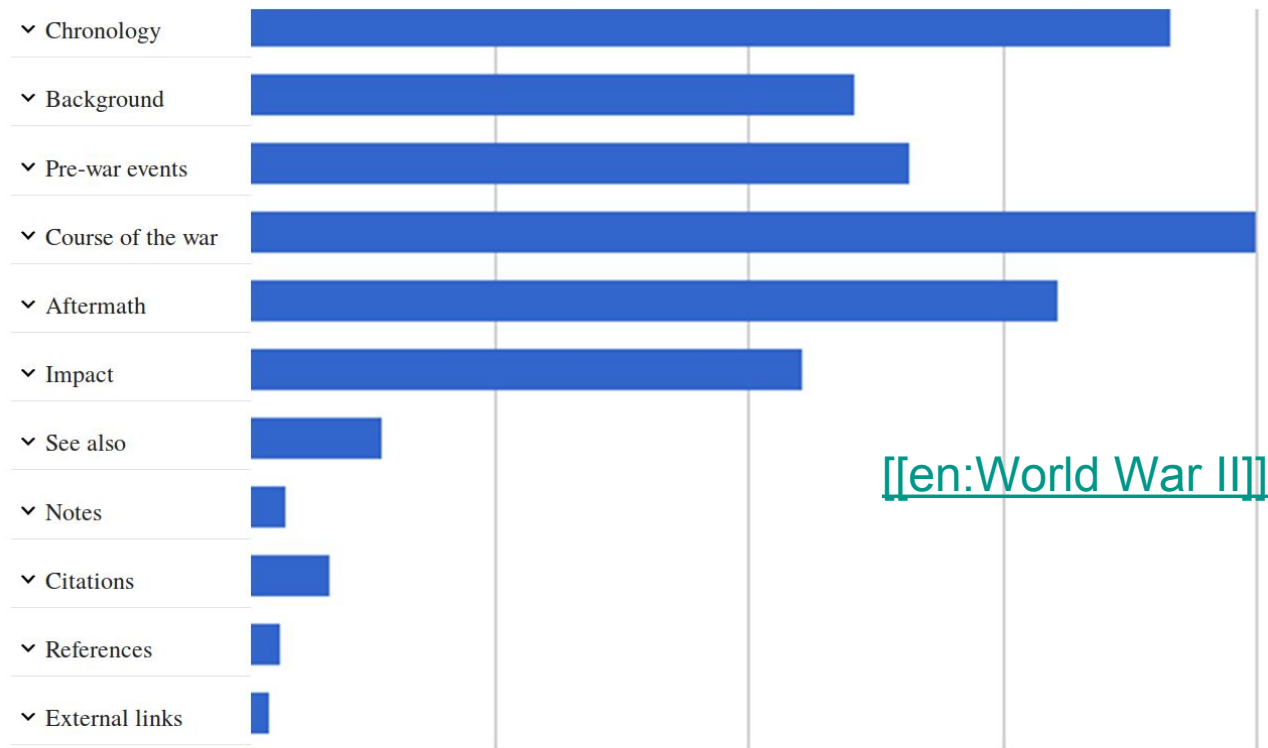View activity for turker: b22907ff39542f84994c2faf64a36bec

page_title
- Diabetes_mellitus
- Diabetes_mellitus_type_1
- Diabetic_diet
- Insulin_pump
- Main_Page

Which sections are readers interested in?

- Early life and career
- Legislative career, 1997–2008
- Presidential campaigns
- Presidency (2009–present)
- Cultural and political image
- Family and personal life
- Notes and references
- External links

[[en:Barack Obama]]

From section expansions data

Which sections are readers interested in?

[[en:World War II]]

Which parts do readers look at? Eyetracking

Standard technique, but expensive equipment and training required. Not used by WMF so far.

Research at Universität Regensburg:

- E.g. in "lookup" tasks, readers spend >45% of time on scanning TOC and lists, in "learn" tasks it's <10%
- Examined eyes movement (saccades) between images and text features: "39% of all contact points lead from image to image, in mutual directions (previous or next). All text contact points [e.g. footnotes] sum up to a total of 37%. In 5% of all cases, an introduction triggers a saccade to an image. "

Knäusl, Hanna (2014). "Situationsabhängige Rezeption von Information bei Verwendung der Wikipedia" (Thesis at the University of Regensburg). p. 202

Rösch, Barbara (2014). "Investigation of Information Behavior in Wikipedia Articles". Proceedings of the 5th Information Interaction in Context Symposium. IIiX '14. New York, NY, USA: ACM. pp. 351–353. doi:10.1145/2637002.2637062. ISBN 978-1-4503-2976-7. (paywalled)
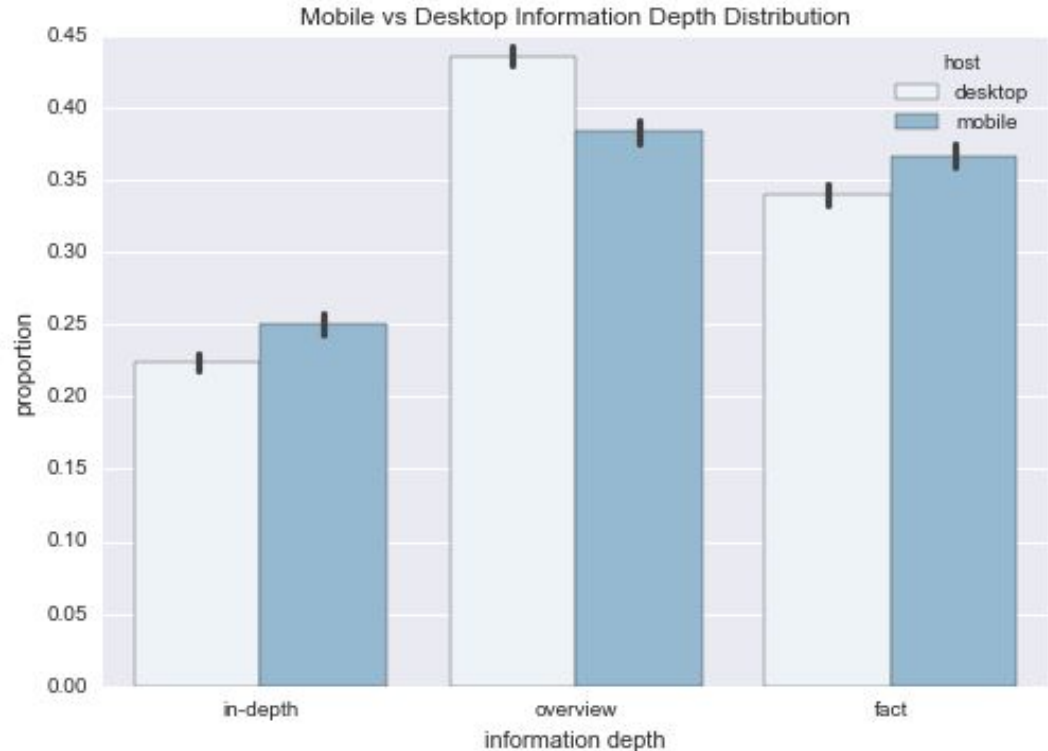
Other current WMF research

- Qualitative research: In-depth interviews with individual readers in the Global South (Mexico)

  - See [[mw:Reading/Mexico Readers Research]]

- Ongoing series of reader surveys to build a taxonomy of Wikipedia reader and of articles (based on usage by different kinds of readers)

Other current WMF research: Reader surveys

- Descriptive analysis
  - What depth of knowledge are readers seeking:
    - In-depth
    - overview / "dictionary definition"
    - isolated fact lookup
  - Conditional on other survey responses, e.g., given that a reader is doing a quick look-up, how likely he is to be in Wikipedia for a school project versus for intrinsic learning

Ellery Wulczyn, CC BY-SA 4.0

Other current WMF research: Reader surveys

(cont.)

- Predictive analysis
  - Predict the depth of information sought from pageview session features?
  - Use pageview session data to provide better read recommendations?

See [[m:Characterizing Wikipedia Reader Behaviour]]

Questions?

- User:Tbayer (WMF)
- tbayer at wikimedia dot org
- IRC (Freenode): HaeB

See also:

- [[m:Research:Newsletter]] / @wikiresearch
- [[mw:Reading]]
- [[m:Research:Which_parts_of_an_article_do_readers_read]]