

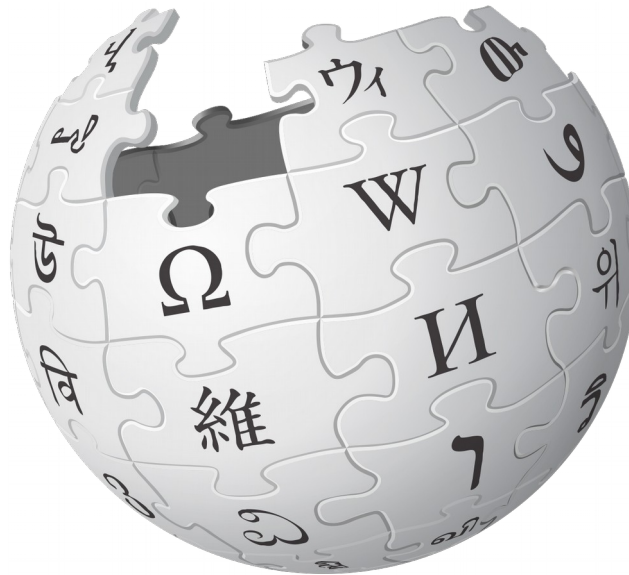
Research on Wikipedia

Seminar Kick Off



Why Bother?

- Wikipedia is the fifth most visited Website globally¹
 - preceded only by Google, Youtube, Facebook, Twitter
- It is an important source of information^[citation needed]
- It is a huge collaborative effort
 - i.e., we can also observe processes of collaboration



¹ <https://www.visualcapitalist.com/the-50-most-visited-websites-in-the-world/>

Stimulus Question

- If you were to operate Wikipedia, how much would it cost?



Stimulus Question

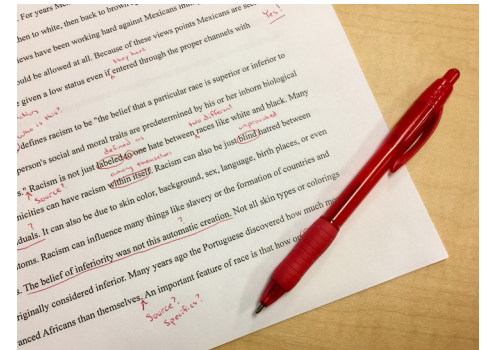
- If you were to operate Wikipedia, how much would it cost?
- Study from 2013¹:
 - Administrative costs are about \$25M per year
 - Freelance writers would charge ~\$300 for an average Wikipedia article
 - \$1.8B for 6M articles in the English Wikipedia
 - \$16.8B for all 56M articles in all languages
 - Keeping Wikipedia up to date
 - Study from 2011²: French Wikipedia would cost \$32M/year
 - French Wikipedia then was ~1M articles
 - total cost of 1.8B/year of keeping all of Wikipedia up to date

¹ <http://infojustice.org/wp-content/uploads/2013/10/band-gerafi10032013.pdf>

² <http://www.amplyd.com/these/Valorisation%20du%20b%C3%A9n%C3%A9volat%20sur%20Wikip%C3%A9dia%20-%20FEB2012%20-%20Vincent%20Juhel.pdf>

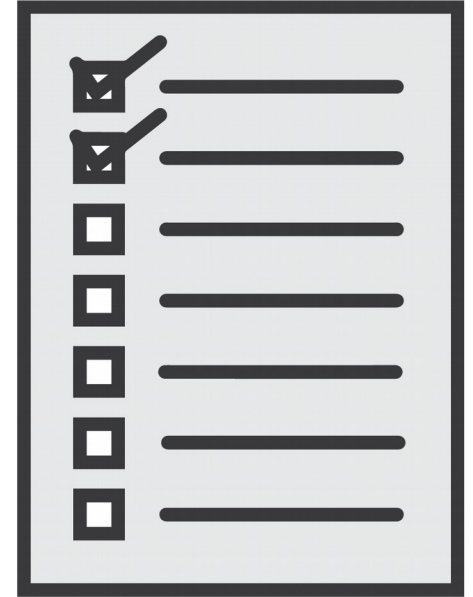
Organization

- Requirements
 - Familiarize yourself with a particular aspect of research on Wikipedia
 - Present it in the seminar
 - Write a seminar paper
 - Review others' seminar papers
 - it is a good idea to also read the main papers for the topics you review
- First step
 - Pick topics of interest
 - Send a ranked list to Ms. Bianca Lerner by the end of the day



Organization

- We will use a process called “peer review”
 - widely used (and discussed) in science
 - you will review your fellow students’ seminar papers
- Timeline
 - Prepare an initial reading list (min. 3 scientific papers) until March 15th
 - Prepare a draft until April 18th
 - You will get two seminar papers to review
 - Submit your reviews until May 2nd
- Seminar (i.e. ,presentations, discussions)
 - May 10th, 17th, 31st, June 14th (we might skip one of those dates)
 - Most likely online (but who knows?)
- Final seminar paper submission: July 4th



Preparing Your Seminar Paper

- Conduct initial literature research
 - get your hands on some core papers on the topic
- Read follow up papers about the topic
 - which follow-up studies have been conducted?
- Find papers citing the original paper
 - who says what about it?
 - who uses it, and for which purposes?
- Anything else you want to discuss
- Yet another hint
 - Most studies look at the English Wikipedia
 - ...but there are also studies on other languages!



Preparing a Review

- 1st rule: be constructive!
- What you should point at
 - can you follow easily? is there information missing at any point?
 - are all claims well supported?
 - do you have any questions not answered?
 - aspects underrepresented
- What you should not do
 - provide general criticism (“don’t like the paper”)
 - correct every spelling mistake
 - rewrite the seminar paper



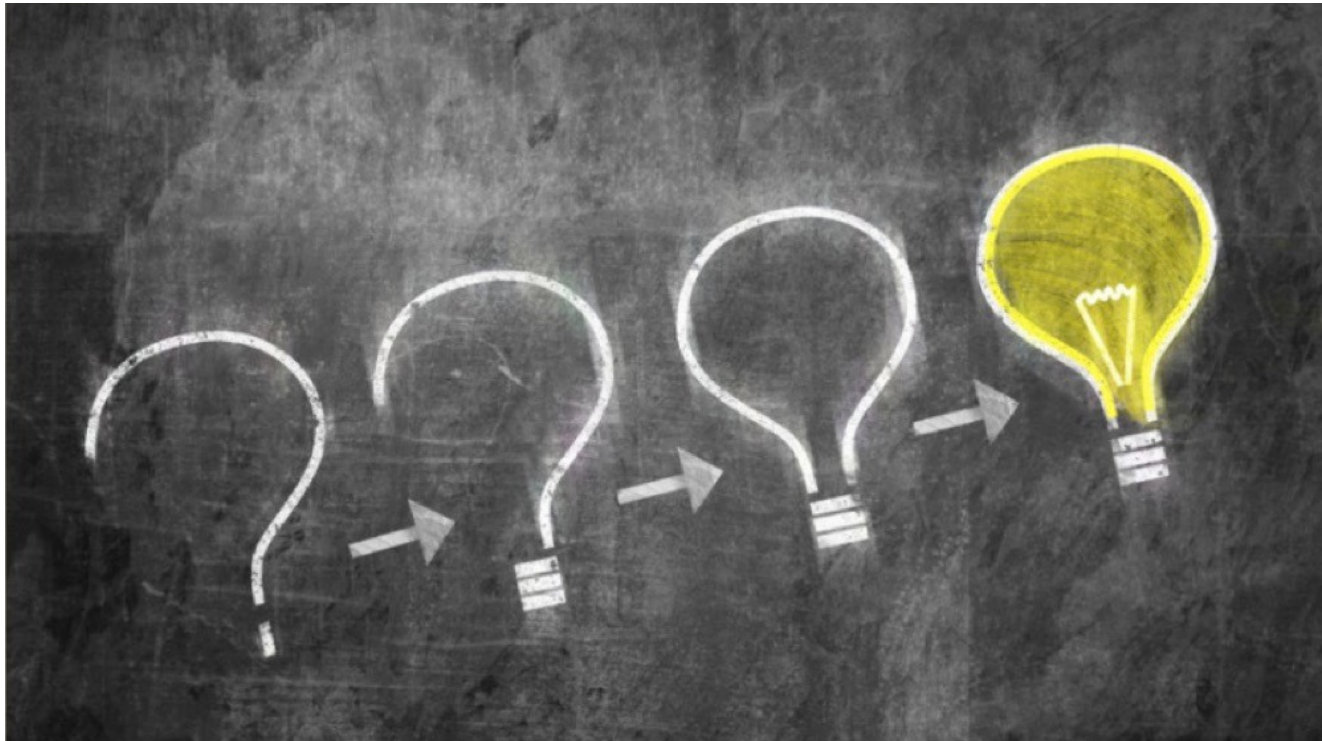
Preparing the Presentation

- You don't have to start at zero
 - we're all doing research on Wikipedia
- Focus on key aspects
 - which research questions do the studies you present address?
 - what are the major findings?
 - which implications do the authors (and you) see?
- Be illustrative
 - use examples
 - show examples of actual data, if available
- Be entertaining
 - it's (partly) up to you whether we are having fun here ;-)



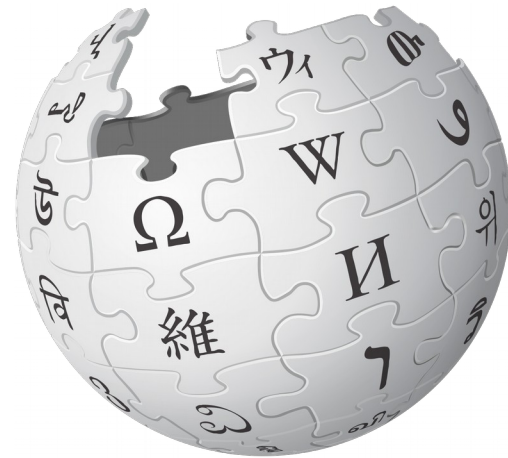
Let's Talk about Topics

- Note: these are suggestions
 - you are free to set your own focus
 - and formulate questions which you find interesting



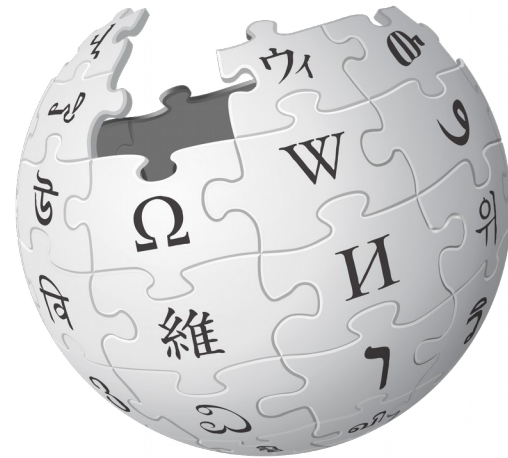
Topic Overview

- Size, growth, and extent of Wikipedia
 - How large is Wikipedia?
 - How does it grow over time?
 - e.g., no. of articles, article length, topical coverage
- Article quality & trustworthiness
 - How accurate is the information in Wikipedia?
 - How does its quality compare to commercial encyclopedias?
- Article coverage
 - Which topics are well covered?
 - Which areas are underrepresented?



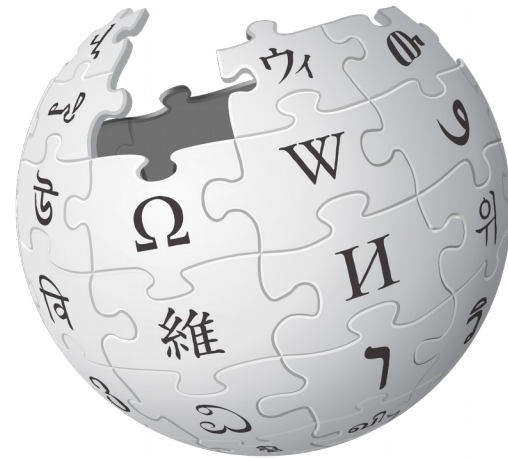
Topic Overview

- Usage of sources and citations
 - How well is Wikipedia backed by sources?
 - What sources are dominant (newspapers, scientific papers, ...)
- Usage of images
 - How many images are there on Wikipedia?
 - Which topics are most backed by images?
 - What role do they play
 - e.g., portrait of a person, illustrative diagram, ...
- Biases
 - Which biases exist in Wikipedia?
 - e.g., Gender biases



Topic Overview

- User Demographics
 - Who edits/uses Wikipedia?
 - Statistics about countries, languages, gender, age, ...
- User Motivation
 - Why do people edit Wikipedia?
 - What are the most common motivations for edits?
- Governance and Policies
 - Which rules exist in Wikipedia?
 - How are they negotiated?
 - How good is the compliance to those rules?



Topic Overview

- Vandalism
 - How frequent is vandalism in Wikipedia?
 - How quickly is it detected/fixed?
 - Are there different kinds of vandalism?
- Collaboration Processes
 - How are articles created?
 - Are there common patterns of collaboration?
 - Are there different user roles in the creation of articles?
- Edit wars and controversies
 - How frequent are edit wars?
 - How are controversial articles handled?



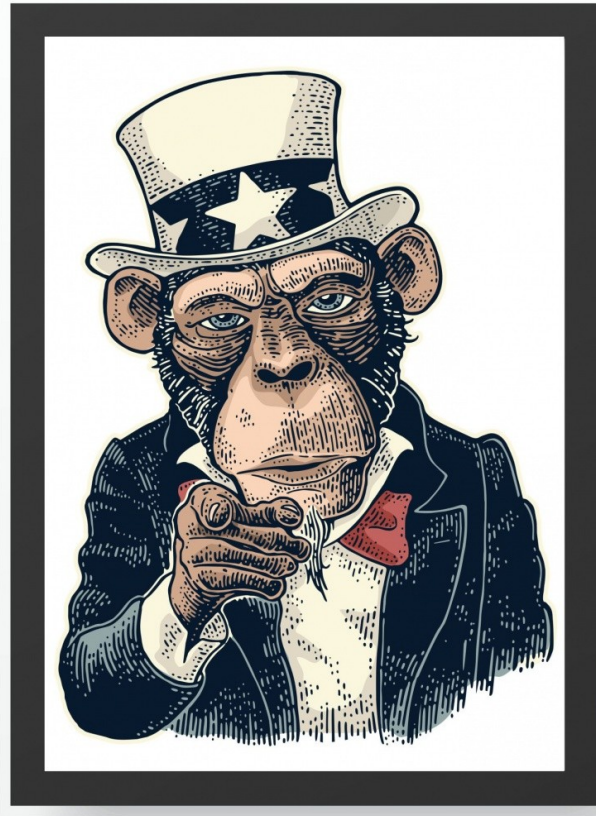
Topic Overview

- Bots on Wikipedia
 - How prevalent are bots in Wikipedia?
 - What do they do?
 - Which rules exist for bot deployment?
- Usage of Wikipedia
 - e.g., navigational patterns
 - article popularity
 - outside references to Wikipedia
 - e.g., how often is Wikipedia cited in scientific publications or news articles?



Now it's up to you...

- ...pick your favorites
- Maybe you have other ideas? They are welcome!



Questions?

