WHERE CITATIONS COME FROM:

STEP #1:
Through a convoluted process, a user's brain generates facts. These are typed into Wikipedia.

The "scroll lock" key was designed by future Energy Secretary Steven Chu in a college project.

STEP #2
A rushed writer checks Wikipedia for a summary of their subject.

US Energy Secretary Steven Chu, Nobel Prize Winner and creator of the ubiquitous "scroll lock" key, testified before Congress today...

STEP #3
Surprised readers check Wikipedia, see the claim, and flag it for review. A passing editor finds the piece and adds it as a citation.

Google is your friend, people.

STEP #4
Now that other writers have a real source, they repeat the fact.

Wikipedia

CITED FACTS

References proliferate, completing the Citogenesis process.
Basic anatomy of a wikipedia page

Definition? [edit]

This does seem like a definition, but there should still be a Wikipedia article on a page. It is a commonly used term. If this does get moved, is there a way that this article? --22:16, 26 September 2005 Twilight Realm

Move discussion in progress [edit]

There is a move discussion in progress on Talk:Page which affects this page. Please participate on that page and not in this talk page section. Thank you. —RMC

References [edit]


• http://www.loc.gov/marc/bibliographic/ecbdhome.html  
History of Wikipedia

- Launched in 2001, under non-profit WikiMedia
- An encyclopedia using the wiki system:
  - “A wiki invites all users—not just experts—to edit any page or to create new pages within the wiki Web site.
  - Wiki promotes meaningful topic associations between different pages by making page link creation intuitively easy and showing whether an intended target page exists or not.”
- In 2009 there was a significant drop of editors (due to stricter “rules” & less low-hanging fruits), and it’s been stable since 2012
How is a Wikipedia page made?

- Theory: Anyone can write new articles!! (true in 2007)
- Reality: Peer-reviewed journal style

1. Submit a draft
2. Article is reviewed or auto-approved if experienced
3. New articles are patrolled
4. Published
How are Wikipedia pages maintained?

- Anyone can edit! But new edits are also heavily patrolled (& bot checked)
- That said, a 2015 experiment had 19/30 well-formatted vandalisms undetected for 2 months - these were not on pop culture articles
How are Wikipedia pages maintained?

- Articles are also assessed and rated based on their completeness, usefulness to reader, style and edits (i.e. infrequently edited articles can be demoted)
- Edits should not be implemented/reverted more than 3 times, or they go to an Arbitration Committee.
- Edit wars can have a Request for Comment, which means 3rd parties can vote/comment on the war.

<table>
<thead>
<tr>
<th>Quality</th>
<th>Top</th>
<th>High</th>
<th>Mid</th>
<th>Low</th>
<th>???</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>✪ FA</td>
<td>1,373</td>
<td>2,172</td>
<td>2,119</td>
<td>1,481</td>
<td>169</td>
<td>7,314</td>
</tr>
<tr>
<td>✪ FL</td>
<td>161</td>
<td>596</td>
<td>661</td>
<td>603</td>
<td>104</td>
<td>2,125</td>
</tr>
<tr>
<td>☣ A</td>
<td>293</td>
<td>603</td>
<td>749</td>
<td>485</td>
<td>87</td>
<td>2,217</td>
</tr>
<tr>
<td>☣ GA</td>
<td>2,670</td>
<td>6,136</td>
<td>12,304</td>
<td>14,807</td>
<td>1,711</td>
<td>37,628</td>
</tr>
<tr>
<td>B</td>
<td>14,170</td>
<td>27,368</td>
<td>43,719</td>
<td>43,118</td>
<td>17,324</td>
<td>145,699</td>
</tr>
<tr>
<td>C</td>
<td>13,956</td>
<td>43,111</td>
<td>103,445</td>
<td>181,656</td>
<td>67,300</td>
<td>409,468</td>
</tr>
<tr>
<td>Start</td>
<td>18,475</td>
<td>86,934</td>
<td>369,069</td>
<td>1,182,365</td>
<td>374,357</td>
<td>2,031,200</td>
</tr>
<tr>
<td>Stub</td>
<td>4,471</td>
<td>32,392</td>
<td>274,862</td>
<td>2,434,480</td>
<td>868,275</td>
<td>3,614,480</td>
</tr>
<tr>
<td>List</td>
<td>4,054</td>
<td>14,611</td>
<td>44,520</td>
<td>139,789</td>
<td>76,578</td>
<td>279,552</td>
</tr>
<tr>
<td>Assessed</td>
<td>59,623</td>
<td>213,923</td>
<td>851,448</td>
<td>3,998,784</td>
<td>1,405,905</td>
<td>6,529,683</td>
</tr>
<tr>
<td>Unassessed</td>
<td>118</td>
<td>510</td>
<td>1,967</td>
<td>16,501</td>
<td>441,309</td>
<td>460,405</td>
</tr>
<tr>
<td>Total</td>
<td>59,741</td>
<td>214,433</td>
<td>853,415</td>
<td>4,015,285</td>
<td>1,847,214</td>
<td>6,990,088</td>
</tr>
</tbody>
</table>

About this table
How are Wikipedia pages maintained?

- People wage wars on edits, typically not on accuracy but on presentation/representation of info, including (from Lamest Edit Wars):
  - Was Chopin Polish, French, Polish–French, or French–Polish?
  - (On Caesar salad) Is it spelled Caesar, Cesar, César, or Cesare? - 11 years & going
  - In the name of the programming language C#, is that # thing after the C a number sign or the musical sharp symbol? (Microsoft said it’s an octothorpe)
  - A revert war on whether the tiger can properly be described as the "most powerful living cat" (complete with accusations that people were "tiger fanboys") gradually led to arguments about how tigers would match up vs. bears and crocodiles (oh my!), complete with another revert war about the inclusion of a YouTube video showing a tiger fighting a crocodile, eventually leading to the article being semi-protected. The debates about bears and crocodiles continue on the talk page.
Meme break

Wikipedia: Please donate, we would be so thankful
Me: Donates
Refreshes page:

Respectfully ask it’s users to donate money in order to keep the site running without ads

I am once again asking for your financial support.
How trustworthy is the info on Wikipedia?

- A few Israeli researchers trained an ML classifier to identify writings & edit patterns by professionals and found that 10-30% of Wikipedia contributors have subject-matter expertise.

- Articles that are extremely important, commonly viewed or commonly vandalised gets a semi-protected (blocks editing to new users) or protected status (requires edits to be requested and reviewed).
How trustworthy is the info on Wikipedia?

- Since vandalism typically involves deleting or modifying content that doesn’t follow the style guidelines/has citation... intentional vandalism, sockpuppeting, meatpuppeting, canvassing & conflicts of interests are extremely hard to detect.

- Systemic bias oh boy

- Wikipedia: “The online encyclopedia does not consider itself to be a reliable source and discourages readers from using it in academic or research settings. Researchers, teachers, journalists, and public officials do not regard Wikipedia as a reliable source.”
Concluding remarks

You now have basic tools for applying linear models to practical settings.

Regularizers can be mixed and matched with different losses (e.g., Logistic regression with $L_1$ regularization).

The overarching framework for linear models is Generalized Linear Model (GLM); sometimes covered in COS424. Or see Wikipedia.

Next time: Applying linear models to problems in other disciplines.
Criticisms of Wikipedia

Systematic bias (especially when it comes to biographies)

- ‘POV pushers’ can derail the editing process
- Wikipedia contributors are 82% male
- Almost half live in Europe and a fifth in North America (9.7% and 4.8% of global population respectively)
- Contributors that join in the last two years feel less empowered to succeed if they are women, live in Eastern Asia, or are not fluent in English

Wikipedia articles for a person has to meet the criteria for notability, which, of course, is heavily affected by systemic sexism & racism

My friend Linh wrote this wikipedia article for her class last semester:
https://en.wikipedia.org/wiki/Soraya_Santiago_Solla
Other crowdsourced encyclopedia attempts

- Uncyclopedia (2005): It’s a potato
Other crowdsourced encyclopedia attempts

Other crowdsourced encyclopedia attempts

- Citizendium (2007): Contributors have to be verified with their real identity & have at least a bachelor’s degree. Enforce gentle guidance & strict oversight.
  - It’s practically dead
Other crowdsourced encyclopedia attempts

  - It’s dead
Other crowdsourced encyclopedia attempts

  - No real quality control i.e. most articles are in pop culture or inaccurate breaking news
Other crowdsourced encyclopedia attempts

- Encyclosphere (202?): A network of format-standardized encyclopedias where anyone can write articles, and articles are rated by readers & experts.
- Allows competition against Wikipedia
- Pioneered by Larry Sanger, who:
  - Co-founded Wikipedia, quit Wikipedia in 2002
  - Founded Citizendium, quit as Citizendium Editor-in-Chief in 2010 & entirely in 2020
  - COO of Everipedia, quit Everipedia in 2019
  - Honestly he’s still trying to fix the encyclopedia system from all the issues Wikipedia has
Wikipedia’s connection to ~big tech~

Google
- uses snippets from Wikipedia directly in search results (Knowledge Graphs)

Youtube
- Links to Wikipedia pages for its information panels (short blurbs under potential misleading videos)

Alexa/ Siri/ etc.
- Recite Wikipedia passages when people ask questions

Implies that what Wikipedia says is “the truth” on the matter

Often these tech services rely on Wikipedia to stay successful

Wikimedia Enterprise
- Service designed to sell Wikipedia’s content to companies (aka going to start charging companies money for using Wikipedia’s content)
Some more memes for the road

@WikiOOC is a Twitter that modifies wikipedia pages using the good ol’ inspect element.