

Wikimania 2011

**Categorization system
strengths and pitfalls**

Disclaimer & who's speaker

- This presentation came to existence out of my consternation over ways some users are able to misuse system of categorization on Wikipedia
- I'm SpeedyGonsales from hr (Croatian) Wikipedia, which made 100,000 articles last month (number of articles does not matter, yeah right :D)

History

Categorization system on Wikipedia arose from two needs (if I recall it correctly, it was long ago):

- to eliminate the need for manual making of lists of articles
- to facilitate easy finding of related articles, to facilitate better structure/organization of articles

History and questions

- On big projects (like en Wikipedia) categorisation system is relatively easy to implement - it is pretty obvious which categories are needed, but what to do on small projects? To go with big system like Universal Decimal Classification, to make custom system or somewhere in between?
- To allow empty categories at all, or to disallow them completely? Number of articles in categories? To limit them to two hundreds (200) as that number of articles is shown by default on one page, or to leave it unlimited?

Analysis

From this two needs categorization system in some articles on at least two wikis (base analysis and conclusions originate from authors experience on his home wiki and en wikipedia, but examples of both strengths and pitfalls spans some ten wikipedia projects with whose languages the author is familiar) tend to become live organism on its own, and in that process lot of different dangers are noted, ranging from usage of categorisation system for unnecessary discrimination (separation of singers by color and race) or applying categories to live or dead people to push them in some inappropriate groups, with weak or no substantiation for that in article references, to giving articles of 100 or 200 hundred words with 20 or more categories.

Current solution

Although experienced users (administrators) of Wikipedia can treat problems which arise with categories same as all other problems which arise every day in busy schedule of active administrator :), sharing of "best practices" can be of use because categorisation is actually not small thing, as if need arise to change categories for 1000 or more articles, that means as much edits on database, be it manual or bot executed, and overall structure of categories have to be well thought.

What I didn't wrote

... on wikimania website, and what also **exist**

- Categories actually can be used in some really dirty ways, like either POV pushing or unnecessary or more or less slight violation of both **Wikipedia:Biographies of living persons** and "In the case of deceased individuals, material must still comply with all Wikipedia policies and prompt removal of questionable material is proper. The burden of evidence for any edit on Wikipedia rests firmly on the shoulders of the person who adds or restores the material, applying not just to verifiability of sources, but to all Wikipedia content policies and guidelines."

Example

- Joseph Stalin page on en wiki, categories:

[[Category:Anti-fascists]] [[Category:Atheism activists]] [[Category:Former Eastern Orthodox Christians]] [[Category:Cold War leaders]] [[Category:Comintern people]] [[Category:Deaths from stroke]] [[Category:Double Heroes of the Soviet Union]] [[Category:Generalissimos]] [[Category:Communists from Georgia (country)]] [[Category:People of World War II from Georgia (country)]] [[Category:Russian people of Georgian descent]] [[Category:Great Purge perpetrators]] [[Category:Heads of government of the Soviet Union]] [[Category:Heads of the Communist Party of the Soviet Union]] [[Category:Heroes of Socialist Labour]] [[Category:Holodomor perpetrators]] [[Category:Joseph Stalin|]] [[Category:Marshals of the Soviet Union]] [[Category:Marxist theorists]] [[Category:Old Bolsheviks]] [[Category:People buried in the Kremlin Wall Necropolis]] [[Category:People from Gori]] [[Category:People of the Russian Civil War]] [[Category:Poets from Georgia (country)]] [[Category:Politburo of the Central Committee of the Communist Party of the Soviet Union members]] [[Category:Recipients of the Order of Victory]] [[Category:Russian atheists]] [[Category:Smallpox survivors]] [[Category:Soviet Ministers of Defence]] [[Category:Soviet people of World War II]] [[Category:Soviet politicians]] [[Category:World War II political leaders]] [[Category:1878 births]] [[Category:1953 deaths]]

- 32 categories (is it maybe too many? Or, is there a set number which we should call "too many"?)
- Sense of category "Category:Smallpox survivors". Does that define him in encyclopedically significant way?
- Just this 2 (remarks). Or, should he have category Mass murderers? Current solution is: Holodomor perpetrators is subcategory of category Mass murderers, with total 6 articles

More of that

bs:Josif Visarionovič Džugašvili Staljin

- 1 category, Vođe Sovjetskog Saveza (Leaders of Soviet Union)
- Better than en wiki?

hr:Josif Staljin

- [[Kategorija:Vojni zapovjednici u Ruskom građanskom ratu]] [[Kategorija:Predsjednici vlada]] [[Kategorija:Sovjetski političari]] [[Kategorija:Komunisti do 1945.]] [[Kategorija:SSSR u Drugom svjetskom ratu]] [[Kategorija:Političari u Drugom svjetskom ratu]] [[Kategorija:Komunisti 1945. - 1989.]] [[Kategorija:Sovjetski vojni zapovjednici u Drugom svjetskom ratu]] [[Kategorija:Biografije, Gruzija]]
- 9 categories, no mention of Mass murders (I am not writing this because I am proud of hr wiki, I'm writing because I'm aware of problem, and I would like that others also be aware)
- hr wiki does not have category Mass murderers, but have category Genocidi, in which there is article Holodomor

Something to think

- Category:People convicted of war crimes (en wiki)
- Ratni zločinci (War criminals – bs wiki)
- Probably most would agree that Stalin is guilty for war crime (Katyn wood massacre), but he's not convicted, so he's not war criminal for en wiki, just Holodomor perpetrator (meaning Mass murderer)
- bs wiki has Ratni zločinci u Drugom svjetskom ratu (War criminals in WWII, but there is no Stalin there)

You think somewhere is better?

- `[[Kategorie:Regierungschef der Sowjetunion]] [[Kategorie:Parteichef der Sowjetunion]] [[Kategorie:Politiker (20. Jahrhundert)]] [[Kategorie:Person im Russischen Bürgerkrieg]] [[Kategorie:Marxismus-Leninismus]] [[Kategorie:Träger des Leninordens]] [[Kategorie:Held der Sowjetunion]] [[Kategorie:Marschall der Sowjetunion]] [[Kategorie:Held der sozialistischen Arbeit]] [[Kategorie:Täter des Großen Terrors (Sowjetunion)]] [[Kategorie:Stalinismus]] [[Kategorie:Georgier]] [[Kategorie:Sowjetbürger]] [[Kategorie:Vertreter des Atheismus]] [[Kategorie:Pseudonym]] [[Kategorie:Massaker von Katyn]] [[Kategorie:Geboren 1878]] [[Kategorie:Gestorben 1953]] [[Kategorie:Mann]] [[Kategorie:Josef Stalin|]] [[Kategorie:Russe]]`
- de wiki, 21 categories, more than hr & bs, less than overkill at en wiki. Massaker von Katyn is mentioned, which is subcategory of **Sowjetische Besetzung Ostpolens**, Deutschland im Zweiten Weltkrieg and **Oblast Smolensk**. So he's evidently no war criminal.
- He's in cat Täter des Großen Terrors (Sowjetunion), which is in Sowjetische Geschichte | Kommunistische Partei der Sowjetunion | Rechtsgeschichte (Russland) | Stalinismus | Zwischenkriegszeit | 1937 | 1938, so no Mass murdered.

Should we go further?

My first point would be:

- Categorization system on "big" Wikipedias is elaborate, but is either funny (read **ridiculous**) in part – like en wiki example of Stalin, or not adequate enough (all wikis presented here)
- On small wikis categorization system is sometimes too small (hr wiki is "shining" example), but what to do? To have 2 categories like War criminals and Convicted war criminals? Convicted ... is exact category, but War criminals *per se* can be tricky. Is it better to have it or not?

Where to go, how to solve...

the conundrum?

- To make global (meta) rules about categories? (I do not like it)
- To allow `Category:Smallpox survivors` for Stalin on en wiki, or to try to limit number of categories in articles, at least project by project? Or to leave idea of systemic solution, and expect that in time all non necessary categories will be deleted, and necessary categories will be added?

Important question again

- Once again, category War criminals – to have it or not?

Possible solution (in small letters)

- Maybe not to limit number of categories, but to minimize the system to meaningful ones (Leaders of Soviet Union) and not evil enough (Born that year, Deceased that year), and to leave all other possibly discriminating (either missing or hard to define) categories to be written in articles and properly sourced. To make that meta policy?
- And even if that become meta policy, question of decision which categories are meaningful (every sound project should be able to do it for itself, but is it best solution? As we are already talking about that :))

Almost end

If somebody congratulated hr wiki for 100,000 articles and I forgot to respond to it:

”Thank you!”

Remark:

- If article about deceased man (over 50 years) and well documented (sourced) can raise any questions, what could anybody wrote about some living person categories across 10 or 50 bigger Wikipedias? Messy, or perfect? :)

End (*da capo*)

Meta policy?

Meta solution similar to idea of omegawiki applied to persons data (birth, death)?

Every project own problem?

If so, how to best solve tricky (not exact) categories?

Thanks (and strenghts)

Thank you on your attention. If this presentation changes something or solve some problem (whenever that may be), I am already happy.

- I wrote about consternation (pitfalls), I never mentioned strenghts. Well, you all know strenghts by heart if you are using Wikipedia at all, sometimes with slight help from web search engines you can find needed data at least 10 times faster than 10 years ago, and more if there were no categories either/or something like Altavista :)

Disclaimer again (and thanks)

2nd disclaimer:

- I chose bs wiki because I understand the language and point is evident, and current situation depicted is most possibly usual for a project in evolving phase.
- I want to congratulate the Wikipedians, for having depth of 230, which is 2nd only to en wiki. I would like to know is it result of no bot/mass adding articles, or because of it? :)

End

Really?

Yup.

More is reserved for members of Wiki(m)pedia kabal, be afraid! You were warned. **Read this fast like pills/medical disclaimer, and make font so small that nobody can read it without lot of effort.**